Identification of the RGG Box Motif in Shadoo: RNA-Binding and Signaling Roles?

Susan M. Corley and Jill E. Gready

Computational Proteomics and Therapy Design Group, Division of Molecular Bioscience, John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra ACT 2601 Australia.

Abstract: Using comparative genomics and *in-silico* analyses, we previously identified a new member of the prion-protein (PrP) family, the gene *SPRN*, encoding the protein Shadoo (Sho), and suggested its functions might overlap with those of PrP. Extended bioinformatics and conceptual biology studies to elucidate Sho's functions now reveal Sho has a conserved RGG-box motif, a well-known RNA-binding motif characterized in proteins such as FragileX Mental Retardation Protein. We report a systematic comparative analysis of RGG-box containing proteins which highlights the motif's functional versatility and supports the suggestion that Sho plays a dual role in cell signaling and RNA binding in brain. These findings provide a further link to PrP, which has well-characterized RNA-binding properties.

Keywords: prion protein, RGG motif, RNA-binding protein, Shadoo, comparative genomics, conceptual biology, methylation, phosphorylation

Introduction

In 2003 we discovered a new gene, *SPRN*, which codes for a 151-residue protein (including N- and C-terminal signal sequences) with topographical similarities unique to prion protein (PrP), and which is highly conserved between fish and mammals (for an analysis of the similarities between PrP and Sho see Premzl et al. 2003). We called this new protein Shadoo (Sho; shadow of prion protein). Like PrP, Sho is most abundant in brain (Premzl et al. 2003; Uboldi et al. 2006; Watts et al. 2007). Although the functions of Sho are as yet little characterized, it has been shown by gain and loss of function experiments (RNAi and overexpression) to be essential for CNS development in zebrafish (L. Sangiorgio, University of Milan, pers. comm.) and may have a neuroprotective effect similar to PrP (Watts et al. 2007). While PrP is notorious for its association with the transmissible spongiform encephalopathies such as Creutzfeldt Jacob Disease and Bovine Spongiform Encephelopathy (Mad Cow Disease) it has become clear in recent years that PrP has a range of normal functions, including in neurogenesis and neural plasticity (Kanaani et al. 2005; Moya et al. 2005; Santuccione et al. 2005; Steele et al. 2006). In defining the natural functions of Sho we are investigating links of the protein's properties to those of other proteins, including PrP.

Here we report our finding that Sho has a conserved 'RGG-box' motif (Kiledjian and Dreyfuss, 1992) defined as a sequence of closely spaced Arg-Gly-Gly (RGG) repeats interspersed with other, often aromatic, amino acids. The RGG box proteins are one class of RNA-binding proteins (RBPs) involved in various aspects of RNA processing, including splicing, stabilizing, transport and translation of mRNAs (Burd and Dreyfuss, 1994). In addition to being an RNA-binding motif, the RGG box of some proteins is known to mediate interactions with other proteins; for a recent detailed example see Lukasiewicz et al. (2007).

The capacity to bind RNA constitutes another point of similarity with PrP which is known to bind RNA and DNA (Grossman et al. 2003). While it has been established that PrP is competent to bind nucleic acid, it is also known to bind many other ligands including polyanionic glycosaminoglycans ('GAGs'). Given its propensity to bind polyanions, it is currently unclear whether the binding of nucleic acids is biologically relevant, that is, whether a normal function of PrP involves this type of interaction or whether binding observed experimentally may be a non-specific interaction. However, others have observed that PrP modifies DNA structure in a manner similar to proteins involved in transcriptional regulation (Bera et al. 2007) and have queried whether PrP may be involved in the biogenesis or transport of nucleic acid (Lima et al. 2006).

Correspondence: Jill E.Gready, Tel: +61 2 6125 8304; Email: jill.gready@anu.edu.au

The approach we have used here is underpinned by a novel combination of comparative genomics (Hedges and Kumar, 2002) and conceptual biology (Blagosklonny and Pardee, 2002). By comparing Sho sequences from species ranging from fish to human and integrating these results with those of a comprehensive analysis of published sequence data and experimental findings, we have been able to put our observations into the broader context of RGG-box proteins. This has allowed us to formulate functional hypotheses for Sho.

Materials and Methods

The amino acid sequences of 12 Sho proteins ranging from fish to human were used in this study. Ten of these sequences are available from GenBank [*Homo sapiens* Np_001012526, *Canis lupus familiaris* CAJ43798, *Bos taurus* CAJ43799, *Mus musculus* NP_898970, *Monodelphis domestica* CAF43800, *Gallus gallus* CAJ43796, *Xenopus tropicalis* CAJ43801, *Danio rerio* CAD35503, *Takifugu rubripes* CAG34291, *Tetraodon nigroviridis* CAG30521]. The sequences for *Ornithorhynchus anatinus* (platypus), *M. domestica* (American opossum), *G. gallus* (chicken), and *X. tropicalis* were initially extracted from the genomic databases (*N. Chakka*, unpublished work of this group). The sequences for *X. tropicalis* and *X. laevis* were also verified experimentally (*T. Vassilieva* and *N. Chakka*, unpublished work of this group). The sequences were aligned using ClustalW (Chenna et al. 2003). Subsequent manual adjustments in the N-terminal region were made to the alignment.

The Swiss-Prot protein database was searched using the program Prosite (Hofmann et al. 1999) http://au.expasy.org/ for known motifs within the Sho sequences. We also searched Swiss-Prot for all proteins that have an RGG-box motif, which we defined as being a sequence of at least 3 RGG repeats with no more than 6 residues between the repeats. This search produced 10 archaeal, 229 bacterial, 14 viral and 1632 eukaryotic sequences, within which there are 607 fungal, 300 plant and 70 human sequences. Examination of the human sequences showed that some well-known RGG-box proteins had not been picked up by this search. The search was then broadened to include proteins with 2 RGG repeats separated by 9, 8, 7, 6 or 5 residues. The results were visually inspected and those proteins with at least one 'RG' between the RGG repeats were included in our list. All uncharacterized proteins or redundant sequences were excluded. The remaining human proteins are collected in Table S1. We have only recorded the sequence beginning and ending with an RGG repeat. It should be noted that the functional RGG box may extend beyond the sequence denoted in Table S1. The RGG sequences were subsequently aligned using ClustalW.

Results and Discussion

Sho—RGG box

A sequence alignment of the N-terminal segment from residue 25 to 42 (the mature protein starts at residue 24) of Shos from different species (Fig. 1) reveals a strictly conserved arginine methylation site (GGRGG) (Lee and Bedford, 2002) at the beginning of a cluster of RGG repeats.

Species	Start	End	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Human	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	S	A	**R**	**G**	**G**	V	**R**	**G**	**G**
Mouse	25	41	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	S	A	**R**	**G**		V	**R**	**G**	**G**
Dog	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	S	A	**R**	**G**	**G**	L	**R**	**G**	**G**
Bovine	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	S	A	**R**	**G**	**G**	**R**	**G**	A	A
Possum (Mdl)	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	A	A	**R**	**G**	**R**	S	**R**	S	S
Platypus	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	A	A	**R**	**G**	A	A	**R**	**G**	A
Chicken	25	42	K	**G**	**G**	**R**	**G**	**G**	S	**R**	**G**	A	A	**R**	**G**	M	A	**R**	**G**	A
Frog (Xl)	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	**G**	A	**R**	**G**	S	S	**R**	**G**	T
Frog (Xt)	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	**G**	A	**R**	**G**	A	S	**R**	**G**	A
Fish (Danio)	25	42	K	**G**	**G**	**R**	**G**	**G**	A	**R**	**G**	S	A	**R**	**G**	T	A	**R**	**G**	**G**
Fish (Fugu)	25	42	K	**G**	**G**	**R**	**G**	**G**	S	**R**	**G**	S	S	**R**	**G**	S	P	S	**R**	S
Fish (Tetraodon)	26	43	K	**G**	**G**	**R**	**G**	**G**	S	**R**	**G**	S	S	**R**	**G**	S	P	S	**R**	S

Figure 1. Alignment of the RGG-box sequence at the N-terminal end of Shos from fish to mammals. LHS are sequence numbers. Mdl, *Monodelphis domestica*; Xl, *Xenopus laevis*; Xt, *Xenopus tropicalis*; Danio, *Danio rerio*; Fugu, *Fugu rubripes*, Tetraodon, *Tetraodon nigroviridis*. Note that region starts with completely conserved KGG triplet. Complete RGG triplets are bolded.

In Shos from human and most other Eutherian mammals there are three RGG repeats, with the first and third separated by 9 residues (**RGG**ARGSA**RGG**V**RGG)**. Thus, the RGG box of human Sho consists of 15 residues: 4 positively charged Arg residues, 7 Gly residues—6 of them dipeptides 'GG' which give the sequence a large degree of flexibility–, and 4 small intervening residues. This pattern diverges slightly in other species; the first RGG repeat is conserved but the second and third RGG repeats are truncated to RG in some cases. In summary, although there is some variability in the number of Gly residues in the Sho RGG box, for species from fish to human there is conservation of Lys25 and the following 3 Arg residues which are regularly spaced with 3 intervening residues between each Arg. The increased prevalence of Gly-Gly dipeptides in the higher Eutherian mammals could suggest evolutionary pressure for increased flexibility in this domain.

Comparative analysis of RGG-box proteins—structure and composition

Proteins with an RGG-box motif, as defined for the purpose of this study (Methods), are presented in Supplementary Information Table S1. Most (#2–#34) are known to have an RNA-binding function. The subset of proteins highlighted in this paper is presented in Table 1. Analysis of all the proteins listed in Table S1 reveals that the RGG box is generally found at the end of the protein sequence, particularly at the C-terminus (Fig. 2A) and is mostly 10–19 residues in length (Fig. 2B). We found a slight preference for RGG repeats to be separated by 9 intervening residues (RGG-X9-RGG), as in Sho, but overall the spacing is variable (Fig. 2C).

The amino acid composition of the sequences was analysed by calculating the proportion of basic (Arg, Lys and His), acidic (Glu and Asp), aromatic (Phe, Trp and Tyr), polar (Ser, Thr, Asn, Gln and Cys), Gly and the other non-polar amino acids (Ala, Val, Leu, Ile, Met and Pro) which make up each sequence and then producing a frequency distribution for the entire set of proteins (Fig. 2D). As expected, a majority of sequences is Gly rich, with peaks in frequency at 50%–60% Gly composition while basic residues peak at 20%–30%. Although a significant number of sequences do not contain an aromatic acid between the RGG repeats, it is possible that there are aromatic residues in close sequence or spatial proximity to this domain. Very few sequences contain acidic residues.

The Sho RGG sequence conforms to these general structural and compositional parameters. It is found at the end of the protein (N-terminus), is 15 residues long and is comprised of 47% Gly, 27% basic, 20% non-polar and 7% polar residues, and has no acidic or aromatic residues. We aligned the RGG sequence of Sho against other sequences with RGG-X9-RGG spacing in order to identify those most similar to Sho (Fig. 3). Several sequences have 50% or more residues identical to those in the Sho RGG box. Experimental studies have demonstrated that the Fragile X Mental Retardation Protein (FMRP) (#32, Table S1) (Zanotti et al. 2006) and the Herpes Simplex protein ICP27 (Mears and Rice, 1996) bind RNA with their RGG boxes which, like Sho, consist of 2 RGG repeats separated by 9 residues.

Overall, our comparative analysis supports the prediction that the RGG box of Sho is competent to bind RNA.

Sho—predicted arginine methylation and phosphorylation sites

The Arg methylation site in Sho is completely conserved in all species from fish to human, suggesting functional importance. Arginine methylation is a common post-translational modification in RGG-box domains (Liu and Dreyfuss, 1995) which affects protein-protein interactions (Boisvert et al. 2005) and RNA binding (Dolzhanskaya et al. 2006). It influences diverse cellular processes, including cellular location of proteins (Passos et al. 2006) transcription, processing and transport of mRNAs (Yu et al. 2004) and signaling pathways (Boisvert et al. 2005).

Phosphorylation is another common post-translational modification found in RBPs. Methylation and phosphorylation mechanisms co-regulate a number of RGG-box proteins, possibly including Sho; again for a detailed example see Lukasiewicz et al. (2007). We identified 3 potential protein kinase C (PKC) phosphorylation sites (SAR (34–36 huSho), SLR (63–65 huSho) and SYR (119–121 huSho)) for Sho. One of these, SAR34–36, is within the RGG box and is found in all the Eutherian mammal sequences analysed (Fig. S1 in Supplementary Information). Phosphorylation of Ser34 would have a direct affect on the structure of the RGG box and most likely affect its function. Although the phosphorylation-site motifs are patterns with a high probability of random

Table 1. Subset of RGG-box proteins (see Table SI in Supplementary Information for full list).

No. (#)[a]	Name (ID)	RGG domain (residue numbers)	Other RNA-binding motifs[b]	Functions/Comments R[e]/P[f] (For details and references see Table SI)
1	Shadoo Q5BIV9	RGGARGSARGGVRGG (28–42)		PrP family member. Likely attached to cell membrane by GPI anchor.
12	hnRNP U Q00839	RGGGHRGRGGFNMRGGN-FRGGAPGNRGG (701–728)		RGG box first identified when a 26-residue sequence (M**RGG**NF**RGG**APGN**RGG**-YNRRGN) found to be sufficient for RNA binding. Expected involvement in splicing, pre-mRNA processing and stabilizes specific mRNAs. **R**
21	HABP4 Hyaluronan binding protein 4 Q5JVS0	RGGPRGGMRGRGRGG (185–199)		Also constitutes a hyaluronan binding motif, (R/K–X(7)-R/K) where X is not acidic. Binds strongly and specifically to hyaluronan and weakly to RNA. Involved in mRNA transport, chromatin remodeling, regulation of transcription. **R/P**
26	EWS Ewing sarcoma Q01844	RGGFDRGGMSRGGRG-GGRGGMGSAGERGG (304–332) and RGGPGGMRG-GRGGLMDRGGPGGMFRG-GRGGDRGGFRGGRGMDRG-GFGGGRRGG (565–617)	1 RRM[c]	Found on cell surface, nucleus and cyto-plasm. Is a transcriptional activator but this activity can be repressed by RGG box. May be involved in pre-mRNA splicing and transport. Suggested that EWS protein acts as a receptor or binding protein for ligands on cell surface, such as nucleic acids, and thus might mediate extracellular and nuclear events. **R/P**
32	FMRP Fragile X Mental Retardation Protein Q06787	RGGGGRGQGGRGRGG (534–548)	2 KH[d]	Binds many mRNA transcripts. Transports mRNA from nucleus to cytoplasm. Involved in neural plasticity through translational repression. **R/P**
33	Nucleolin P19338	RGGGRGGFGGRGGGRGGRG-GFGGRGRGGFGGRGGFRG-GRGG (656–696)	4 RRM	Found on cell surface, nucleus and cytoplasm. RGG box is necessary for efficient RNA binding but the RRMs are required for specific RNA recognition. Duplex DNA, ssDNA and RNA are all effective ligands. Acts as cell surface receptor—binds cytokine MK and HB-19 through its RGG box. **R/P**
34	G3BP1 Ras GTPase-activating protein-binding protein 1 Q13283	RGGLGGGMRGPPRGG (435–449)	1 RRM	Role in ras-signaling pathway affecting cell proliferation and survival as well as involved in RNA metabolism. Cleaves MYC mRNA and has helicase activity. Combining these functions, suggested to be member of novel sub-class of RBPs which act at level of RNA metabolism in response to cell signaling, thus allowing cell to rapidly control protein activity at a stage after transcription. **R/P**

[a]number of protein as appears in Table S1.
[b]RNA-binding motifs in addition to the RGG box.
[c]RRM = 80–90 amino acid sequence containing a RNP-1 (octapeptide) and RNP-2 (6 amino acid) consensus sequences.
[d]K homology region as in hnRNP K.
[e]RNA binding.
[f]Protein binding.

occurrence it is interesting to note that the presence of at least one phosphorylation site has been experimentally confirmed in 70% of the RGG-box proteins surveyed (Table S2 in Supplementary Information). This is a high proportion even taking into account the over-representation of nuclear proteins in the phosphoproteome (Olsen et al. 2006) and leads us to suggest that phosphorylation is particularly prevalent in RGG-box proteins. The finding of potential methylation and phosphorylation sites in Sho is another point of similarity with other RGG-box proteins. The existence of phosphorylation sites within Sho raises the possibility that Sho may be involved in a signaling pathway that is regulated by phosphorylation.

Functional significance

Sho differs from most of the other proteins surveyed in that it has no other RNA-binding motifs. This is unusual but not unique as hnRNP U (#12; Table S1) has no other RNA-binding motif apart from the RGG box. The RGG box is typically associated with binding to single-stranded nucleic acids, (Zhang and Grosse, 1997) whereas additional RNA-binding motifs may allow binding of a broader range of RNA targets as is the case for nucleolin (#33, Table S1) (Ghisolfi et al. 1992). The inherent flexibility of the RGG box (Ramos et al. 2003) can also enable binding to several RNA targets, as has been shown for FMRP (Darnell et al. 2004) which binds to many RNA targets but an affinity for RNA that forms a stable G-quartet structure (Menon and Mihailescu, 2007; Ramos et al. 2003). As Sho lacks other RNA-binding motifs, we expect it to bind single-stranded nucleic acid, and potentially a range of such targets, as for FMRP.

The RGG box is a positively charged domain known to interact electrostatically with other proteins and anionic molecules. A well-characterized

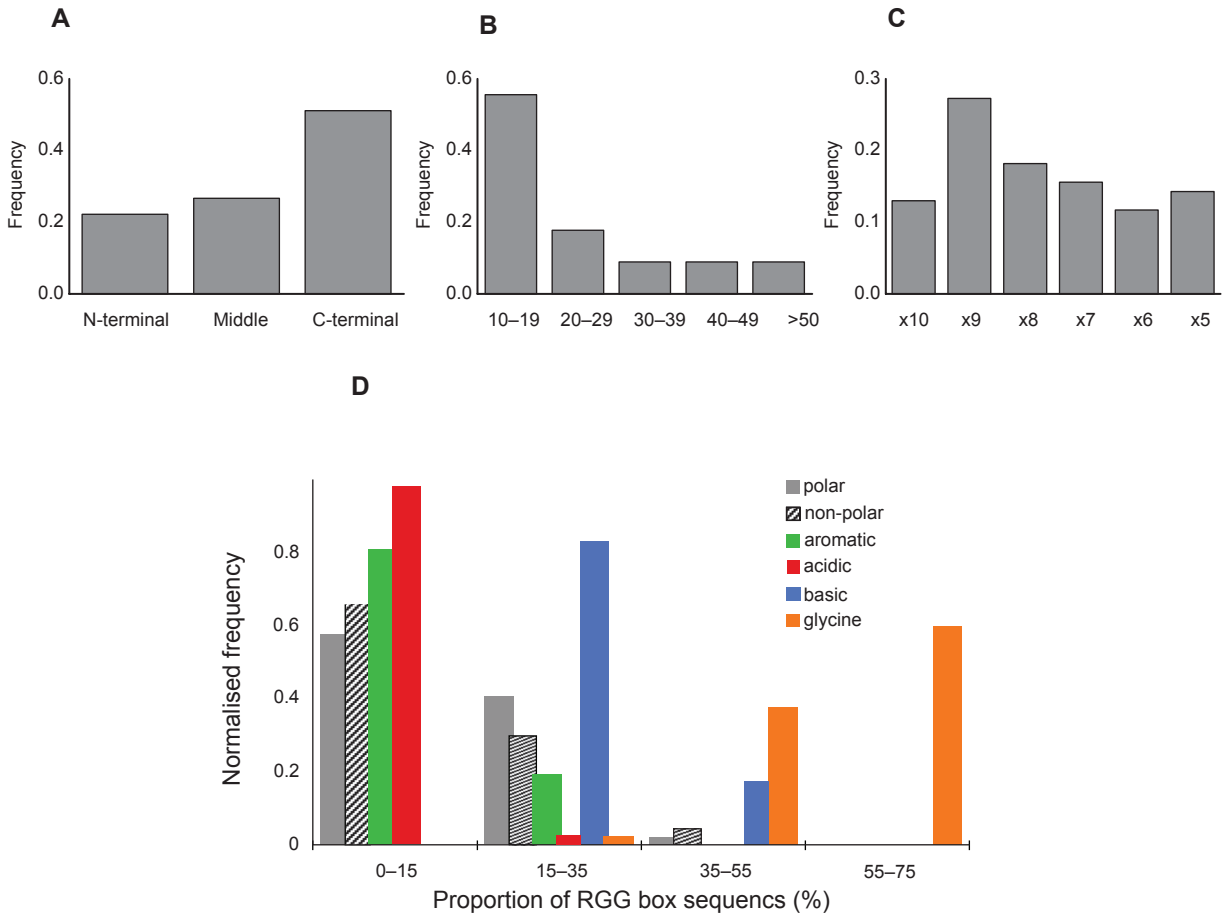

Figure 2. Frequency histograms of structural and compositional features of the 45 RGG sequences surveyed (Table S1). **A)** Position of the RGG box region in the proteins. N-terminal (within the first 35% of the protein sequence), C-terminal (last 35% of the protein), Middle, region between. **B)** Length of the RGG box. **C)** Spacing between RGG repeats where X may be any residue including Arg and Gly. **D)** Amino acid composition in terms of type: basic (R, K, H); acidic (D, E); Gly: aromatic (F, Y, W); non-polar amino acids (A, V, L, I, M, P) and polar (S, T, N, Q, C).

example is the RGG box of the yeast protein Npl3p which docks with the kinase Sky1 (Lukasiewicz et al. 2007). A non-protein example is provided by the intracellular hyaluronan binding protein (HAPB4) (#21, Table S1) which has high sequence similarity to Sho (Fig. 3). The RGG domain of HAPB4 also constitutes a glycosaminoglycan ('GAG') binding motif (R/K–X(7)-R/K) (Yang et al. 1994) and has been found to bind strongly and specifically to hyaluronan and weakly to RNA (Huang et al. 2000). Although it is not surprising to find this motif in an Arg-rich sequence (in fact it is present in most of the proteins included in Table S1), it has particular relevance in the case of Sho, given its cellular location.

The cellular location of Sho will determine its opportunities to bind RNA and whether this is its primary function. We originally predicted Sho to be a GPI-anchored protein (Premzl et al. 2003). This has now been confirmed in mouse (Watts et al. 2007) and for a Sho-like protein (Sho2) (Premzl et al. 2004; Strumbo et al. 2006) in zebrafish (Miesbauer et al. 2006). However, some GPI-anchored proteins, including PrP, undergo anchor cleavage ('shedding'), (Parkin et al. 2004; Zhang et al. 2005) resulting in formation of soluble proteins which can relocate to other cellular destinations and are capable of performing multiple functions (Campana et al. 2005). While the cell surface is one likely location for Sho, it may be a multifunctional protein found in other cellular locations as well, as for PrP. If Sho sheds its GPI anchor or undergoes proteolytic cleavage before attachment to the cell membrane (Watts et al. 2007), the RGG-box domain would be available for functional roles intracellularly. Other RGG-box proteins are known to have multiple cellular locations, for example, nucleolin and the Ewing Sarcoma (EWS) protein (#26, Table S1) are found on the cell surface as well as in the nucleus and cytoplasm. In fact, there is growing evidence that some RNA-binding proteins have additional roles as cell surface receptors (Bajenova et al. 2003; Belyanskaya et al. 2003; Hirano et al. 2005) and in signaling pathways as noted for the ras GTPase activating protein binding protein 1 (Kennedy et al. 2001).

Attached to the cell surface, Sho would be positioned to act as a receptor for ligands found at the cell surface, including nucleic acids, as suggested for EWS (Belyanskaya et al. 2001). Sho may, therefore, have a role in cell signaling, similar to PrP which binds the neural cell adhesion molecule and thus participates in the tyrosine kinase fyn signaling pathway leading to neurite outgrowth (Santuccione et al. 2005). Alternatively, in this location Sho may bind other anionic ligands such as the GAG, hyaluronan, which is known to bind another GPI-anchored protein, brevican, and is involved in the structural plasticity of neural tissue

Protein	Start	End	#	Sequence
Q5BIV9_HUMAN	28	42		**R G G** A **R G** S A **R G G** V **R G G**
HABP4_HUMAN	185	199	**10**	**R G G** P **R G G** M **R G R G R G G**
FUS_HUMAN	491	505	**10**	**R G G** F **R G G R G G G** D **R G G**
K1C9_HUMAN	478	492	**9**	**R G G** S **G G** S Y **G R G** S **R G G**
G3BP1_HUMAN	435	449	**9**	**R G G** L **G G G** M **R G** P P **R G G**
LS14A_HUMAN	406	420	**9**	**R G G** Y **R G R G G** L **G** F **R G G**
NOLA1_HUMAN	173	187	**9**	**R G G R G G G R G G G G R G G**
NOLA1_HUMAN	185	199	**9**	**R G G G R G G G** F **R G G R G G**
EWS_HUMAN	309	323	**9**	**R G G** M S **R G G R G G G R G G**
NOLA1_HUMAN	30	44	**8**	**R G G G G G G G G** N F **R G G**
FBRL_HUMAN	24	38	**8**	**R G G R G G** F **G G G R G R G G**
BRWD3_HUMAN	1699	1713	**8**	**R G G G G** T **R G R G R G R G G**
RBP56_HUMAN	337	351	**8**	**R G G** Y **R G R G G** F Q **G R G G**
THOC4_HUMAN	38	52	**7**	**R G G G** A Q A A A **R** V N **R G G**
FBRL_HUMAN	15	29	**7**	**R G G** F **G** D **R G G R G G R G G**
FA98A_HUMAN	352	366	**7**	**R G G** H E Q **G G G R G G R G G**
PP1RA_HUMAN	726	740	**7**	**R G G R** S **G G G** P P N **G R G G**
HNRPG_HUMAN	113	127	**7**	**R G G** S **G G** T **R G** P P S **R G G**
FMR1_HUMAN	534	548	**7**	**R G G G G R G** Q **G G R G R G G**
HNRPU_HUMAN	714	728	**7**	**R G G** N F **R G G** A P **G** N **R G G**
CA077_HUMAN	128	142	**6**	**R G G** M S L **R G G** N L L **R G G**

Figure 3. Alignment of the RGG box of proteins with RGG-X9-RGG spacing. The number of the residue at the start and end of the sequence is given, as well as the total number of exact residue matches (#) to Sho.

(Rauch, 2004). It is interesting to note that PrP also binds GAGs including hyaluronan and heparin (Pan et al. 2002) and that GAGs may facilitate the conversion of the normal cellular PrP to the isoform found in prion disease (Yin et al. 2007).

If Sho were to shed its GPI anchor and re-enter the cell or if a segment of the N-terminal region incorporating the RGG domain was cleaved off prior to expression at the cell surface, the RGG box would be available to interact with cellular RNA. Indeed, as a small protein of no more than 123 residues, Sho would be capable of diffusing in and out of the nucleus (Cyert, 2001) and shuttling RNA from the nucleus to the cytoplasm. This is a function normally performed by RNA-binding proteins involved in neural plasticity, which participate in the biogenesis of mRNA, its transport to dendrites and repression of translation pending appropriate neural stimulation (Ule and Darnell, 2006).

Conclusion

In summary, we have observed that Sho has a conserved RGG-box domain with similar composition to other known RGG-box proteins. We predict that this domain has functional significance and may mediate some of the neural functions already indicated for Sho. Our analysis leads us to postulate that Sho is an RNA-binding protein which may also play a role in cell signaling. Our initial experiments to test the prediction have shown Sho RGG box peptide is competent to bind RNA but further work is required to characterize the interaction.

The discovery of the RGG box in Sho opens new avenues for investigating its function and potential functional overlap with PrP. It is known that PrP plays a role in neural plasticity through its involvement in neural signaling pathways. Here we suggest that Sho may bind mRNA directly and thus play a role in neural plasticity similar to other neural RBPs.

Note

Beck et al. (J. Med. Genet. online 19/9/08) have reported an association of a null allele of *SPRN* with variant CJD.

Acknowledgements

We acknowledge support from the Australian National University Institute for Advanced Studies (IAS) block grant.

Disclosure

The authors report no conflicts of interest.

References

Bajenova, O., Stolper, E., Gapon, S., Sundina, N., Zimmer, R. and Thomas, P. 2003. Surface expression of heterogeneous nuclear RNA binding protein M4 on Kupffer cell relates to its function as a carcinoembryonic antigen receptor. *Exp. Cell. Res.*, 291;228–41.

Belyanskaya, L.L., Delattre, O. and Gehring, H. 2003. Expression and subcellular localization of Ewing sarcoma (EWS) protein is affected by the methylation process. *Exp. Cell. Res.*, 288;374–81.

Belyanskaya, L.L., Gehrig, P.M. and Gehring, H. 2001. Exposure on cell surface and extensive arginine methylation of Ewing sarcoma (EWS) protein. *J. Biol. Chem.*, 276:18681–7.

Bera, A., Roche, A.C. and Nandi, P.K. 2007. Bending and unwinding of nucleic acid by prion protein. *Biochemistry*, 46:1320–8.

Blagosklonny, M.V. and Pardee, A.B. 2002. Conceptual biology: unearthing the gems. *Nature*, 416:73.

Boisvert, F.M., Chenard, C.A. and Richard, S. 2005. Protein interfaces in signaling regulated by arginine methylation. *Sci. STKE.*, 2005:re2.

Burd, C.G. and Dreyfuss, G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 265:615–21.

Campana, V., Sarnataro, D. and Zurzolo, C. 2005. The highways and byways of prion protein trafficking. *Trends Cell. Biol.*, 15:102–11.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, 31:3497–500.

Cyert, M.S. 2001. Regulation of nuclear localization during signaling. *J. Biol. Chem.*, 276:20805–8.

Darnell, J.C., Warren, S.T. and Darnell, R.B. 2004. The fragile X mental retardation protein, FMRP, recognizes G-quartets. *Ment. Retard. Dev. Disabil. Res. Rev.*, 10:49–52.

Dolzhanskaya, N., Merz, G., Aletta, J.M. and Denman, R.B. 2006. Methylation regulates the intracellular protein-protein and protein-RNA interactions of FMRP. *J. Cell. Sci.*, 119:1933–46.

Ghisolfi, L., Joseph, G., Amalric, F. and Erard, M. 1992. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J. Biol. Chem.*, 267:2955–9.

Grossman, A., Zeiler, B. and Sapirstein, V. 2003. Prion protein interactions with nucleic acid: possible models for prion disease and prion function. *Neurochem. Res.*, 28:955–3.

Hedges, S.B. and Kumar, S. 2002. Genomics. Vertebrate genomes compared. *Science*, 297:1283–5.

Hirano, K., Miki, Y., Hirai, Y., Sato, R., Itoh, T., Hayashi, A., Yamanaka, M., Eda, S. and Beppu, M. 2005. A multifunctional shuttling protein nucleolin is a macrophage receptor for apoptotic cells. *J. Biol. Chem.*, 280:39284–93.

Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.*, 27:215–9.

Huang, L., Grammatikakis, N., Yoneda, M., Banerjee, S.D. and Toole, B.P. 2000. Molecular characterization of a novel intracellular hyaluronan-binding protein. *J. Biol. Chem.*, 275:29829–39.

Kanaani, J., Prusiner, S.B., Diacovo, J., Baekkeskov, S. and Legname, G. 2005. Recombinant prion protein induces rapid polarization and development of synapses in embryonic rat hippocampal neurons in vitro. *J. Neurochem.*, 95:1373–86.

Kennedy, D., French, J., Guitard, E., Ru, K., Tocque, B. and Mattick, J. 2001. Characterization of G3BPs: tissue specific expression, chromosomal localisation and ras GAP(120) binding studies. *J. Cell. Biochem.*, 84:173–87.

Kiledjian, M. and Dreyfuss, G. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO. J.*, 11:2655–64.

Lee, J. and Bedford, M.T. 2002. PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO. Rep.*, 3:268–73.

Lima, L.M., Cordeiro, Y., Tinoco, L.W., Marques, A.F., Oliveira, C.L., Sampath, S., Kodali, R., Choi, G., Foguel, D., Torriani, I., Caughey, B. and Silva, J.L. 2006. Structural insights into the interaction between prion protein and nucleic acid. *Biochemistry*, 45:9180–7.

Liu, Q. and Dreyfuss, G. 1995. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.*, 15:2800–8.

Lukasiewicz, R., Nolen, B., Adams, J.A. and Ghosh, G. 2007. The RGG domain of Npl3p recruits Sky1p through docking interactions. *J. Mol. Biol.*, 367:249–61.

Mears, W.E. and Rice, S.A. 1996. The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *J. Virol.*, 70:7445–53.

Menon, L. and Mihailescu, M.R. 2007. Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res.*, 35:5379–92.

Miesbauer, M., Bamme, T., Riemer, C., Oidtmann, B., Winklhofer, K.F., Baier, M. and Tatzelt, J. 2006. Prion protein-related proteins from zebrafish are complex glycosylated and contain a glycosylphosphatidylinositol anchor. *Biochem. Biophys. Res. Commun.*, 341:218–24.

Moya, K.L., Hassig, R., Breen, K.C., Volland, H. and Di Giamberardino, L. 2005. Axonal transport of the cellular prion protein is increased during axon regeneration. *J. Neurochem.*, 92:1044–53.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. and Mann, M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127:635–48.

Pan, T., Wong, B.S., Liu, T., Li, R., Petersen, R.B. and Sy, M.S. 2002. Cell-surface prion protein interacts with glycosaminoglycans. *Biochem. J.*, 368:81–90.

Parkin, E.T., Watt, N.T., Turner, A.J. and Hooper, N.M. 2004. Dual mechanisms for shedding of the cellular prion protein. *J. Biol. Chem.*, 279:11170–8.

Passos, D.O., Quaresma, A.J. and Kobarg, J. 2006. The methylation of the C-terminal region of hnRNPQ. (NSAP1) is important for its nuclear localization. *Biochem. Biophys. Res. Commun.*, 346:517–25.

Premzl, M., Gready, J.E., Jermiin, L.S., Simonic, T. and Marshall Graves, J.A. 2004. Evolution of vertebrate genes related to prion and Shadoo proteins-clues from comparative genomic analysis. *Mol. Biol. Evol.*, 21:2210–31.

Premzl, M., Sangiorgio, L., Strumbo, B., Marshall Graves, J.A., Simonic, T. and Gready, J.E. 2003. Shadoo, a new protein highly conserved from fish to mammals and with similarity to prion protein. *Gene*, 314:89–102.

Ramos, A., Hollingworth, D. and Pastore, A. 2003. G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*, 9:1198–207.

Rauch, U. 2004. Extracellular matrix components associated with remodeling processes in brain. *Cell. Mol. Life. Sci.*, 61:2031–45.

Santuccione, A., Sytnyk, V., Leshchyns'ka, I. and Schachner, M. 2005. Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p 59fyn and to enhance neurite outgrowth. *J. Cell. Biol.*, 169:341–54.

Steele, A.D., Emsley, J.G., Ozdinler, P.H., Lindquist, S. and Macklis, J.D. 2006. Prion protein (PrPc) positively regulates neural precursor proliferation during developmental and adult mammalian neurogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, 103:3416–21.

Strumbo, B., L. Sangiorgio, S. Ronchi, J.E. Gready, and T. Simonic. 2006. Cloning and analysis of transcripts and genes encoding fish-specific proteins related to PrP. *Fish Physiol. Biochem.*, 32:339–53.

Uboldi, C., Paulis, M., Guidi, E., Bertoni, A., Meo, G.P., Perucatti, A., Iannuzzi, L., Raimondi, E., Brunner, R.M., Eggen, A. and Ferretti, L. 2006. Cloning of the bovine prion-like Shadoo (SPRN) gene by comparative analysis of the predicted genomic locus. *Mamm. Genome*, 17:1130–9.

Ule, J. and Darnell, R.B. 2006. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.*, 16:102–10.

Watts, J.C., Drisaldi, B., Ng, V., Yang, J., Strome, B., Horne, P., Sy, M.S., Yoong, L., Young, R., Mastrangelo, P., Bergeron, C., Fraser, P.E., Carlson, G.A., Mount, H.T., Schmitt-Ulms, G. and Westaway, D. 2007. The CNS glycoprotein Shadoo has PrP(C)-like protective properties and displays reduced levels in prion infections. *EMBO J.*, 26:4038–50.

Yang, B., Yang, B.L., Savani, R.C. and Turley, E.A. 1994. Identification of a common hyaluronan binding motif in the hyaluronan binding proteins RHAMM, CD44 and link protein. *EMBO J.*, 13:286–96.

Yin, S., Pham, N., Yu, S., Li, C., Wong, P., Chang, B., Kang, S.C., Biasini, E., Tien, P., Harris, D.A. and Sy, M.S. 2007. Human prion proteins with pathogenic mutations share common conformational changes resulting in enhanced binding to glycosaminoglycans. *Proc. Natl. Acad. Sci. U.S.A.*, 104:7546–51.

Yu, M.C., Bachand, F., McBride, A.E., Komili, S., Casolari, J.M. and Silver, P.A. 2004. Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes Dev.*, 18:2024–35.

Zanotti, K.J., Lackey, P.E., Evans, G.L. and Mihailescu, M.R. 2006. Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry*, 45:8319–30.

Zhang, A.S., West, A.P., Jr., Wyman, A.E., Bjorkman, P.J. and Enns, C.A. 2005. Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. *J. Biol. Chem.*, 280:33885–94.

Zhang, S. and Grosse, F. 1997. Domain structure of human nuclear DNA helicase II (RNA helicase A). *J. Biol. Chem.*, 272:11487–94.

Identification of the RGG Box Motif in Shadoo: RNA-Binding and Signaling Roles?

Susan M. Corley and Jill E. Gready

Supplementary Information

Table SI. Proteins with RGG-box domains Selected on Criteria explained in Methods.

	Database name, name (ID)	# AA	RGG domain (residue numbers)	Other RNA binding motifs[a]	Functions/Comments	R[d]/P[e]
1	SHO_HUMAN, Shadoo (Q5BIV9)	151	RGGARGSARGGVRGG (28–42)		PrP family member. Likely attached to cell membrane by GPI anchor (Premzl et al. 2003).	
2	ROA0_HUMAN, hnRNP A0 (Q13151)	305	RGGNFSGRGGFGGSRGG (192–202)	2 RRM[b]	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Similar to hnRNP A/B but less abundant. Component of ribonucleosomes. (Jurica et al. 2002).	R
3	ROA1_HUMAN, hnRNP AI (P09651)	372	RGGNFSGRGGFGGSRGG (218–234)	2 RRM	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Transport of poly (A) mRNA from nucleus to cytoplasm. (Biamonti et al. 1989; Jurica et al. 2002; Siomi and Dreyfuss, 1995).	R
4	ROA2_HUMAN, hnRNP A2 (P22626)	353	RGGNFGFGDSRGGGGNFG PGPGSNFRGG (203–230)	2 RRM	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Trafficking of RNAs containing the cis-acting A2 response element (A2RE). (Jurica et al. 2002).	R
5	ROA3_HUMAN hnRNP A3 (P51991)	378	RGGGSGNFMGRGGNFGG GGGNFGRGG (216–241)	2 RRM	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Trafficking of RNAs containing the cis-acting A2 response element (A2RE). (Jurica et al. 2002).	R
6	HNRPD_HUMAN, hnRNPD0 (Q14103)	355	RGGFAGRARGRGG (272–284)	2 RRM	Binds to mRNA with AU-rich elements (AREs) in 3'-UTR. Transcription regulator; binds to ds and ss DNA sequences. Possibly Involved in translationally coupled mRNA turnover. (Kajita et al. 1995; Tay et al. 1992).	R
7	HNRPG_HUMAN, hnRNP G (P38159)	391	RGGSGGTRGPPSRGG (113–126)	1 RRM	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. (Jurica et al. 2002).	R

(*Continued*)

Table SI. (*Continued*)

	Database name, name (ID)	# AA	RGG domain (residue numbers)	Other RNA binding motifs[a]	Functions/Comments	R[d]/P[e]
8			RGGGRGGSRSDRGG (373–386)			
9	HNRPK, hnRNP K (P61978)	463	RGGFDRMPPGRGGRPMPP SRRDYDDMSPRRGPPPPPP GRGGRGGSRARNLPLPPPP PPRGG (258–318)	3 KH[C]	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Major poly(C) RNA binding hnRNP. Also binds poly(C) ssDNA. (Jurica et al. 2002).	R
10	HNRPQ_HUMAN, hnRNP Q (060506)	623	RGGPGSARGVRGARGGA QQQRGRGVRGARGGRGG (526–559)	3 RRM	3 isoforms. Pre-mRNA processing. Associated with splicing intermediates and mature mRNA. Interacts preferentially with poly(A) and poly (U) RNA sequences. Region 518–549 (RGRAGYSQRGGPGSARGVRGAR GGAQQQRGRG) sufficient to bind RNA. (Mourelatos et al. 2001).	R
11	HNRPR_HUMAN, hnRNP R (043390)	633	RGGRGGPAQQQRGRGSRG SRGNRGG (543–567)	3 RRM	Found in splicesome C; expected involvement in splicing, pre-mRNA processing. (Hassfeld et al. 1998; Jurica et al. 2002).	R
12	HNRPU_HUMAN, hnRNP U (Q00839)	824	RGGGHRGRGGFNMRGGN FRGGAPGNRGG (701–728)		First discovered that hnRNP U contains a 26-residue peptide (M**RGG**NF**RGG**APGN-**RGG**YNRRG N) sufficient to account for its RNA-binding activity. This novel RNA-binding motif was defined as the RGG box. Found in splicesome C; expected involvement in splicing, pre-mRNA processing. Stabilizes specific mRNAs. Aka SAF-A; has high affinity for DNA with scaffold attachment regions (SAR). (Fackelmayer and Richter, 1994; Helbig and Fackelmayer, 2003; Jurica et al. 2002; Kiledjian and Dreyfuss, 1992; Yugami et al. 2007).	R
13	HNRL1, hnRNP U like protein 1 (Q9BUJ2)	856	RGGGGFRGRGGGGGFQR YENRGPPGGNRGGFQN RGGGSGGGGNYRGG (612–658)		Pre-mRNA processing and transport. Binds poly(G) and poly(C) RNA. Represses transcription driven by viral and cellular promoters. Associated with RBRD7, activates transcription. (Gabler et al. 1998).	R
14	PURG_HUMAN, Purine-rich element-binding protein gamma (Q9UJV8)	347	RGGGGGRGRGG (7–17)		In purine-rich element binding protein family (PUR). Binds ssDNA and RNA. Highly expressed in many tumor lines. (Liu and Johnson, 2002).	R

15	DDX4_HUMAN, DEAD box protein 4 (Q9NQI0)	724	RGGRGSFRGCRGG (147–159)		In DEAD box helicase family. Helicase activity, RNA unwinding, needed in splicing, ribosome biogenesis and RNA degradation. (Castrillon et al. 2000; Luking et al. 1998).	R
16	THOC4_HUMAN, Tho complex subunit 4 (Q86V81)	257	RGGGAQAAARVNRGG (38–52)	1 RRM	In THO/TREX complex, promotes transcriptional activation, recruited to RNA polymerase during elongation. Associated with spliced mRNA; roles in mRNA export and decay. May mediate interactions of proteins and/or RNA. (Strasser et al. 2002; Virbasius et al. 1999).	R/P
17	NOLA1_HUMAN, Nucleolar protein family A member 1 (Q9NY12)	217	RGGGRGGFNRGGGGGGF NRGGSSNHFRGGGGGGGG GNFRGGGRGGFGRGGG RGG (4–57)		Aka GAR1. Required for ribosome biogenesis and telomere maintenance. Processing or intranuclear trafficking of TERC, the RNA component of the telomerase reverse transcriptase (TERT). RGG box accessory to RNA binding. Interaction with SMN1 requires at least one of the RGG-box regions. (Bagni and Lapeyre, 1998; Whitehead et al. 2002).	R
18			RGGGRGGRGGGRGGGG RGGGRGGGFRGGRGG GGGGFRGGRGG (169–210)			
19	SFPQ_HUMAN, Splicing factor proline- and glutamine-rich (P23246)	707	RGGGGGGFHRRGGGGG RGG (9–27)	2 RRM	Pre-mRNA splicing factor. Binds to intronic polypyrimidine tracts. Possible role in nuclear retention of defective RNAs. Regulates basal and cAMP-dependent transcription. (Patton et al. 1993).	R
20	FBRL_HUMAN, Fibrillarin (P22087)	321	RGGGFGGRGGFGDRGG RGGRGGFGGGRGRGG GFRGRGRGG (8–47)		Involved in pre-rRNA processing. Component of box C/D small nucleolar ribonucleoprotein (snoRNP) particles. (Aris and Blobel, 1991; Jansen et al. 1991).	R
21	HABP4_HUMAN, Hyaluronan binding protein (HAPB4, Ki-1/57) (Q5JVS0)	413	RGGPRGGMRGRGRGG (185–199)		This sequence also constitutes a hyaluronan binding motif, (R/K-X(7)-R/K) where X is not acidic (Yang et al. 1994). This domain within HAPB4 has been found to bind strongly and specifically to hyaluronan and weakly to RNA. Involved in mRNA transport, chromatin remodeling, regulation of transcription. Interacts with chromodomain DNA helicase binding protein 3(CHD3). (Kobarg et al. 1997; Lemos et al. 2003; Passos et al. 2006).	R/P
22	PAIRB_HUMAN, Plasminogen activator inhibitor 1 RNA binding protein (Q8NC51)	408	RGGRGGRGGRGRGG (367–380)		Aka CG1–55. Regulation of mRNA stability/ decay. Interacts with CHD3, similar to HABP4. (Lemos et al. 2003).	R/P

(*Continued*)

Table SI. (*Continued*)

	Database name, name (ID)	# AA	RGG domain (residue numbers)	Other RNA binding motifs[a]	Functions/Comments	R[d]/P[e]
23	FUS_HUMAN, RNA binding protein FUS (P35637)	526	RGGGRGGRGGMGGSD RGG (244–261)	1 RRM	Component of nuclear riboprotein complexes. Binds ds and ss DNA. Promotes annealing of complementary ssDNAs. (Rabbitts et al. 1993).	R/P
24			RGGGNGRGGRGRGGPMG RGG (377–396)			
25			RGGRGGYDRGGYRGRGG DRGGFRGGRGGGDRGG (473–505)			
26	EWS_HUMAN, Ewing sarcoma (EWS) protein (Q01844)	656	RGGFDRGGMSRGGRGGG RGGMGSAGERGG (304–332)	1 RRM	Found on cell surface as well as in the nucleus and cytoplasm. Binds RNA. Is a transcriptional activator but this activity can be repressed by the RGG box. May be involved in pre mRNA splicing and transport. It has been suggested that EWS protein may act as a receptor or binding protein for ligands on the cell surface, such as nucleic acids, and thus might mediate extracellular and nuclear events. Interacts with PTK2B/ FAK2 then relocates from cytoplasm to ribosomes. (Belyanskaya et al. 2003; Belyanskaya et al. 2001; Ohno et al. 1994; Plougastel et al. 1993).	R/P
27			RGGPGGMRGGRGGLMD RGGPGGMFRGGRGGD RGGFRGGRGMDRGGFGG GRRGG (565–617)			
28	RB56_HUMAN, TATA-binding protein-associated factor 2N (Q92804)	592	RGGYRGRGGFQGRGG (337–351)	1 RRM	Binds RNA and ssDNA. Transcription regulation. In RNA polymerase II transcriptional multiprotein complex. Similar to EWS and FUS/TLS. (Morohoshi et al. 1996).	R/P
29			RGGGYGGDRGGGYGGD RGGGYGGDRGGYGGD RGGGYGGDRGGYGGD RGGYGGDRGGYGGDRGG YGGDRSRGGYGGDRGG (459–537)			
30	CIRPB_HUMAN, Cold-inducible RNA-binding protein (Q14011)	172	RGGSAGGRGFFRGGRGRG RGFSRGG (94–118)	1 RRM	Cold-induced suppression of cell proliferation. Activates the ERK pathway. (Nishiyama et al. 1997).	

31	PP1RA_HUMAN, Serine/threonine-protein phosphatase 1 regulatory subunit (Q96QC0)	940	RGGPGPGPGPYHRGRGG RGGNEPPPPPPPFRGA RGGRSGGGPPNGRGG (693–740)		Aka p99. Binds mRNA, ssDNA, poly(A) and poly(G). Inhibits phosphatase activities when phosporylated. (Kreivi et al. 1997; Totaro et al. 1998).	R
32	FMR1_HUMAN, Fragile X Mental Retardation Protein (FRMP) (Q06787)	632	RGGGGRGQGGRGRGG (534–548)	2 KH	Binds many mRNA transcripts. Transports mRNA from nucleus to cytoplasm. Involved in neural plasticity through translational repression. (Bagni and Greenough, 2005; Darnell et al. 2001; Ule and Darnell, 2006; Zalfa et al. 2003).	R/P
33	NUCL_HUMAN, Nucleolin (P19338)	710	RGGGRGGFGGRGGGRGG RGGFGGRGRGGFGGRGGF RGGRGG (656–696)	4 RRM	Found on cell surface as well as in the nucleus and cytoplasm. RGG box is necessary for efficient RNA binding and possibly operates by unstacking RNA bases, but the RRMs are required for specific RNA recognition. Duplex DNA, ssDNA and RNA are all effective ligands for nucleolin. Associated with intranucleolar chromatin and preribosomal particles. Binds to histone HI to induce chromatin decondensation. When attached to the cell surface, nucleolin binds the proteins cytokine MK and HB-19 through its RGG box and acts as cell surface receptor. (Ghisolfi et al. 1992; Hirano et al. 2005; Said et al. 2002).	R/P
34	G3BP1_HUMAN, Ras GTPase-activating protein-binding protein 1 (Q13283)	466	RGGLGGGMRGPPRGG (435–449)	1 RRM	G3BP has a role in the ras-signaling pathway affecting cell proliferation and survival as well as being involved in RNA metabolism. Cleaves MYC mRNA And has Helicase activity—unwinds DNA/DNA, RNA/DNA and RNA/RNA. Combining these two functions, it has been suggested the G3BPs are members of a novel subclass of RNA-binding proteins which act at the level of RNA metabolism in response to cell signaling allowing the cell to rapidly control protein activity at a stage after transcription. Also involved in formation of stress granules. (Irvine et al. 2004; Kennedy et al. 2001; Tourriere et al. 2003; Tourriere et al. 2001).	R/P

(*Continued*)

Table SI. (*Continued*)

	Database name, name (ID)	# AA	RGG domain (residue numbers)	Other RNA binding motifs[a]	Functions/Comments	R[d]/P[e]
35	RGMC_HUMAN, Hemojuvelin (precursor) (Q6ZVN8)	426	RGGGSSGALRGGGGGG RGG (54–72)		In repulsive guidance molecule (RGM) family; RGMa and RGMb involved in neural development. GPI anchored. Interacts with neogenin which regulates shedding of GPI anchor. Binding cytokines BMP2 and BMP4 affects BMP signaling pathway and expression of hepcidin. Function of RGG domain unknown. (Matsunaga and Chedotal, 2004; Zhang et al. 2007; Zhang et al. 2005).	P
36	ZNH14_HUMAN, Zinc finger HIT domain-containing protein 4 (Q9C086)	343	RGGRGGARGERRGG (238–251)		Aka PAPA-1. Induces growth and cell cycle arrest at Gl phase. Interacts with splicing factors altering pre-mRNA splicing. Complexes with other nucleolar proteins. (Kuroda et al. 2004; Maita et al. 2004).	P
37	K1C9_HUMAN, Keratin type 1 cytoskeletal 9 (P35527)	623	RGGSGGSYGRGSRGG (478–492)		Cytoskeletal and microfibrillar keratin. Function in mature or developing palmar and plantar skin (Langbein et al. 1993).	
38	MRE11_HUMAN, Double strand break repair protein MRE 11A (P49959)	708	RGGRGQNSASRGG (577–589)		In MRN complex, role in dsDNA repair, recombination, maintenance of telomere integrity and meiosis. (Petrini et al. 1995).	
39	WBP7_HUMAN WW domain-binding protein 7 (Q9UMN6)	2715	RGGQSSRGGRGG RGRGRGG (281–299)		WW domain-binding (Trithorax homolog 2). Possible transcriptional regulator. (FitzGerald and Diaz, 1999).	
40	BRWD3_HUMAN, Bromodomain and WD repeat-containing protein 3 (Q6RI45)	1802	RGGGGTRGRGRGRGG (1699–1713)		In WD repeat protein family involved in cell-cycle progression, signal transduction, apoptosis, gene regulation. Possible transcription factor with 2 bromodomains and 9 WD repeats. May be involved in Jak/Stat pathway. (Vodermaier, 2001).	
41	CA077_HUMAN, Uncharacterised protein Clorf77 (Q9Y3Y2)	248	RGGVRGRGGPGRGG (153–166)		NA	

42	FA98A_HUMAN, Protein FAM98A	519	RGGHEQGGGRGGRGG YDHGGRGG (352–374)	NA
43	(Q8NCA5) FA98A_HUMAN (Q8NCA5)	519	RGGGRGGRGGRGGRGG (458–473)	NA
44	LS14A_HUMAN, LSM14 protein homolog A (Q8ND56)	463	RGGYRGRGGLGF RGGRGRGGGRGG (406–429)	Putative alpha synuclein binding protein.

[a]RNA-binding motifs in addition to the RGG box.
[b]RRM = 80–90 amino acid sequence containing RNP-1 (octapeptide) and RNP-2 (6 amino acid) consensus sequences.
[c]K homology region as in hnRNP K.
[d]RNA binding.
[e]Protein binding.

Table S2. Phosphorylation sites in RGG box proteins surveyed in this study[a].

Protein	Id	PKC[b]	CK2[C]	TYR[d]	Expt[e]
SHO_HUMAN	Q5BIV9	3	0	0	
ROAO_HUMAN	Q13151	4	3	0	2
ROA1_HUMAN	P09651	10	10	0	9
ROA2_HUMAN	P22626	9	4	0	6
ROA3_HUMAN	P51991	9	7	0	6
HNRPD_HUMAN	Q14103	9	6	1	7
HNRPG_HUMAN	P38159	18	16	1	7
HNRPK	P61978	7	12	1	6
HNRPQ_HUMAN	O60506	7	3	2	2
HNRPR_HUMAN	O43390	5	5	2	
HNRPU_HUMAN	Q00839	10	5	0	5
HNRL1	Q9BUJ2	6	9	0	3
PURG_HUMAN	Q9UJV8	6	2	0	1
DDX4_HUMAN	Q9NQI0	16	14	0	
THOC4_HUMAN	Q86V81	4	5	0	1
NOLA1_HUMAN	Q9NY12	3	1	0	
SFPQ_HUMAN	P23246	7	4	2	1
FBRL_HUMAN	P22087	5	3	0	
HABP4_HUMAN	Q5JVS0	5	8	2	2
PAIRB_HUMAN	Q8NC51	6	10	0	12
FUS_HUMAN	P35637	7	6	1	
EWS_HUMAN	Q01844	4	5	0	
RB56_HUMAN	Q92804	7	12	2	1
CIRPB_HUMAN	Q14011	4	2	1	
PP1RA_HUMAN	Q96QC0	10	12	2	4
FMR1_HUMAN	Q06787	9	12	1	1[f]
NUCL_HUMAN	P19338	8	23	0	14
G3BP1_HUMAN	Q13283	2	6	1	5
RGMC_HUMAN	Q6ZVN8	11	2	0	
ZNH14_HUMAN	Q9C086	3	0	0	
K1C9_HUMAN	P35527	7	14	3	
MRE11_HUMAN	P49959	16	17	0	4
WBP7_HUMAN	Q9UMN6	40	36	3	4
BRWD3_HUMAN	Q6RI45	34	42	4	4
CA077_HUMAN	Q9Y3Y2	4	2	0	1
FA98A_HUMAN	Q8NCA5	5	9	1	
LS14A_HUMAN	Q8ND56	4	7	1	11

[a]Searches were conducted using the ScanProsite program available on the ExPASy Proteomics Server of the Swiss Institute of Bioinformatics website http://au.expasy.org/.
[b]Number of protein kinase C phosphorylation sites (PS00005).
[c]Number of casein kinase II phosphorylation sites (PS00006).
[d]Number of tyrosine kinase phosphorylation sites (PS00007).
[e]As annotated in the SwissProt database.
[f]Mazroui, R., Huot, M.E., Tremblay, S., Boilard, N., Labelle, Y. and Khandjian, E.W. (2003) Fragile X Mental Retardation protein determinants required for its association with polyribosomal mRNPs. *Hum Mol Genet,* 12;3087–96.

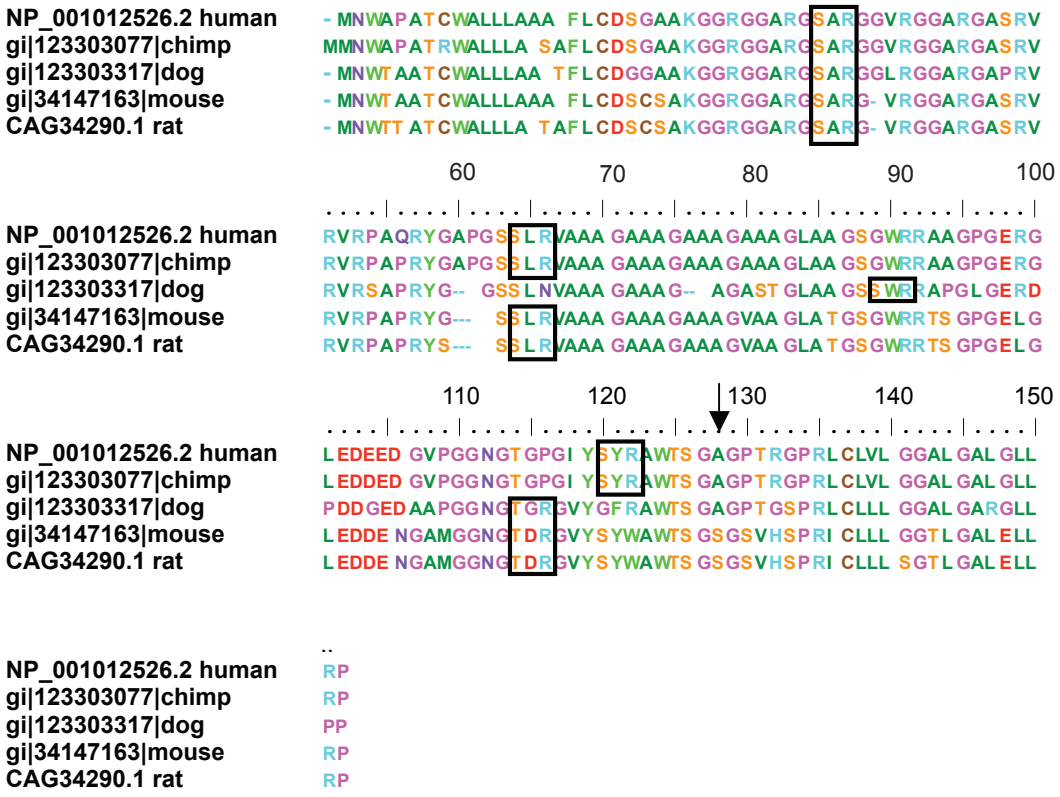

Figure SI. Alignment of Sho sequences of Eutherian mammals. The 3 PKC phosphorylation sites are indicated by boxes. The N-terminal and C-terminal cleavage sites are shown by arrows.

References

Aris, J.P. and Blobel, G. 1991. cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. *Proc. Natl. Acad. Sci. U.S.A.*, 88:931–5.

Bagni, C. and Greenough, W.T. 2005. From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat. Rev. Neurosci.*, 6:376–87.

Bagni, C. and Lapeyre, B. 1998. Garl p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *J. Biol. Chem.*, 273:10868–73.

Belyanskaya, L.L., Delattre, O. and Gehring, H. 2003. Expression and subcellular localization of Ewing sarcoma (EWS) protein is affected by the methylation process. *Exp. Cell. Res.*, 288:374–81.

Belyanskaya, L.L., Gehrig, P.M. and Gehring, H. 2001. Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. *J. Biol. Chem.*, 276:18681–7.

Biamonti, G., Buvoli, M., Bassi, M.T., Morandi, C., Cobianchi, F. and Riva, S. 1989. Isolation of an active gene encoding human hnRNP protein Al. Evidence for alternative splicing. *J. Mol. Biol.*, 207:491–503.

Castrillon, D.H., Quade, B.J., Wang, T.Y., Quigley, C. and Crum, C.P. 2000. The human VASA gene is specifically expressed in the germ cell lineage. *Proc. Natl. Acad. Sci. U.S.A.*, 97:9585–90.

Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell.*, 107:489–99.

Fackelmayer, F.O. and Richter, A. 1994. hnRNP-U/SAF-A is encoded by two differentially polyadenylated mRNAs in human cells. *Biochim. Biophys. Acta.*, 1217:232–4.

FitzGerald, K.T. and Diaz, M.O. 1999. MLL2: A new mammalian member of the trx/MLL family of genes. *Genomics.*, 59:187–92.

Gabler, S., Schütt, H., Groitl, P., Wolf, H., Shenk, T., and Dobner, T. 1998. E1B 55-kilodalton-associated protein: a cellular protein with RNA-binding activity implicated in nucleocytoplasmic transport of adenovirus and cellular mRNAs. *J. Virol.*, 72:7960–71.

Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F. and Erard, M. 1992. Concerted activities of the RNA recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with pre-ribosomal RNA. *Eur. J. Biochem.*, 209:541–8.

Hassfeld, W., Chan, E.K., Mathison, D.A., Portman, D., Dreyfuss, G., Steiner, G. and Tan, E.M. 1998. Molecular definition of heterogeneous nuclear ribonucleoprotein R (hnRNP R.) using autoimmune antibody: immunological relationship with hnRNP P. *Nucleic. Acids. Res.*, 26:439–45.

Helbig, R. and Fackelmayer, F.O. 2003. Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma.*, 111:173–82.

Hirano, K., Miki, Y., Hirai, Y., Sato, R., Itoh, T., Hayashi, A., Yamanaka, M., Eda, S. and Beppu, M. 2005. A multifunctional shuttling protein nucleolin is a macrophage receptor for apoptotic cells. *J. Biol. Chem.*, 280:39284–93.

Irvine, K., Stirling, R., Hume, D. and Kennedy, D. 2004. Rasputin, more promiscuous than ever: a review of G3BP. *Int. J. Dev. Biol.*, 48:1065–77.

Jansen, R.P., Hurt, E.C., Kern, H., Lehtonen, H., Carmo-Fonseca, M., Lapeyre, B. and Tollervey, D. 1991. Evolutionary conservation of the human nucleolar protein fibrillarin and its functional expression in yeast. *J. Cell. Biol.*, 113:715–29.

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N. and Moore, M.J. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna.*, 8:426–39.

Kajita, Y., Nakayama, J., Aizawa, M. and Ishikawa, F. 1995. The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein DO. Common modular structure and binding properties of the 2xRBD-Gly family. *J. Biol. Chem.*, 270:22167–75.

Kennedy, D., French, J., Guitard, E., Ru, K., Tocque, B. and Mattick, J. 2001. Characterization of G3BPs: tissue specific expression, chromosomal localisation and rasGAP(120) binding studies. *J. Cell. Biochem.*, 84:173–87.

Kiledjian, M. and Dreyfuss, G. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *Embo. J.*, 11:2655–64.

Kobarg, J., Schnittger, S., Fonatsch, C., Lemke, H., Bowen, M.A., Buck, F. and Hansen, H.P. 1997. Characterization, mapping and partial cDNA sequence of the 57-kD intracellular Ki-1 antigen. *Exp. Clin. Immunogenet.*, 14:273–80.

Kreivi, J.P., Trinkle-Mulcahy, L., Lyon, C.E., Morrice, N.A., Cohen, P. and Lamond, A.I. 1997. Purification and characterisation of p99, a nuclear modulator of protein phosphatase 1 activity. *FEBS Lett.*, 420:57–62.

Kuroda, T.S., Maita, H., Tabata, T., Taira, T., Kitaura, H., Ariga, H. and Iguchi-Ariga, S.M. 2004. A novel nucleolar protein, PAPA-1, induces growth arrest as a result of cell cycle arrest at the Gl phase. *Gene.*, 340:83–98.

Langbein, L., Heid, H.W., Moll, I. and Franke, W.W. 1993. Molecular characterization of the body site-specific human epidermal cytokeratin 9: cDNA cloning, amino acid sequence, and tissue specificity of gene expression. *Differentiation*, 55:57–71.

Lemos, T.A., Passos, D.O., Nery, F.C. and Kobarg, J. 2003. Characterization of a new family of proteins that interact with the C-terminal region of the chromatin-remodeling factor CHD-3. *FEBS Lett.*, 533:14–20.

Liu, H. and Johnson, E.M. 2002. Distinct proteins encoded by alternative transcripts of the PURG gene, located contrapodal to WRN. on chromosome 8, determined by differential termination/polyadenylation. *Nucleic. Acids. Res.*, 30:2417–26.

Luking, A., Stahl, U. and Schmidt, U. 1998. The protein family of RNA helicases. *Crit. Rev. Biochem. Mol. Biol.*, 33:259–96.

Maita, H., Kitaura, H., Keen, T.J., Inglehearn, C.F., Ariga, H., Iguchi-Ariga, S.M. 2004. PAP-1, the mutated gene underlying the RP9 form of dominant retinitis pigmentosa, is a splicing factor. *Exp. Cell. Res.*, 300:283–96.

Matsunaga, E. and Chedotal, A. 2004. Repulsive guidance molecule/neogenin: a novel ligand-receptor system playing multiple roles in neural development. *Dev. Growth Differ.*, 46:481–6.

Morohoshi, F., Arai, K., Takahashi, E.I., Tanigami, A. and Ohki, M. 1996. Cloning and mapping of a human RBP56 gene encoding a putative RNA binding protein similar to FUS/TLS and EWS proteins. *Genomics.*, 38:51–7.

Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. 2001. SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.*, 20:5443–52.

Nishiyama, H., Higashitsuji, H., Yokoi, H., Itoh, K., Danno, S., Matsuda, T. and Fujita, J. 1997. Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. *Gene.*, 204:115–20.

Ohno, T., Ouchida, M., Lee, L., Gatalica, Z., Rao, V.N. and Reddy, E.S. 1994. 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.*, 9:3087–97.

Passos, D.O., Bressan, G.C., Nery, F.C. and Kobarg, J. 2006. Ki-1/57 interacts with PRMT1 and is a substrate for arginine methylation. *Febs. J.*, 273:3946–61.

Patton, J.G., Porro, E.B., Galceran, J., Tempst, P. and Nadal-Ginard, B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes. Dev.*, 7:393–406.

Petrini, J.H., Walsh, M.E., DiMare, C., Chen, X.N., Korenberg, J.R. and Weaver, D.T. 1995. Isolation and characterization of the human MRE11 homologue. *Genomics.*, 29:80–6.

Plougastel, B., Zucman, J., Peter, M., Thomas, G. and Delattre, O. 1993. Genomic structure of the EWS gene and its relationship to EWSR.1, a site of tumor-associated chromosome translocation. *Genomics.*, 18:609–15.

Premzl, M., Sangiorgio, L., Strumbo, B., Marshall Graves, J.A., Simonic, T. and Gready, J.E. 2003. Shadoo, a new protein highly conserved from fish to mammals and with similarity to prion protein. *Gene*, 314:89–102.

Rabbitts, T.H., Forster, A., Larson, R. and Nathan, P. 1993. Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12; 16) in malignant liposarcoma. *Nat. Genet.*, 4:175–80.

Said, E.A., Krust, B., Nisole, S., Svab, J., Briand, J.P. and Hovanessian, A.G. 2002. The anti-HIV cytokine midkine binds the cell surface-expressed nucleolin as a low affinity receptor. *J. Biol. Chem.*, 111:31492–502.

Siomi, H. and Dreyfuss, G. 1995. A nuclear localization domain in the hnRNP Al protein. *J. Cell. Biol.*, 129:551–60.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R. and Hurt, E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, 417:304–8.

Tay, N., Chan, S.H. and Ren, E.C. 1992. Identification and cloning of a novel heterogeneous nuclear ribonucleoprotein C-like protein that functions as a transcriptional activator of the hepatitis B. virus enhancer II. *J. Virol*, 66:6841–8.

Totaro, A., Grifa, A., Carella, M., Rommens, J.M., Valentino, M.A., Roetto, A., Zelante, L. and Gasparini, P. 1998. Cloning of a new gene (FB.19) within HLA class I region. *Biochem. Biophys. Res. Commun.*, 250:555–7.

Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J.M., Bertrand, E. and Tazi, J. 2003. The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J. Cell. Biol.*, 160:823–31.

Tourriere, H., Gallouzi, I.E., Chebli, K., Capony, J.P., Mouaikel, J., van der Geer, P. and Tazi, J. 2001. RasGAP-associated endoribonuclease G3Bp: selective RNA degradation and phosphorylation-dependent localization. *Mol. Cell. Biol.*, 21:7747–60.

Ule, J. and Darnell, R.B. 2006. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.*, 16:102–10.

Virbasius, C.M., Wagner, S. and Green, M.R. 1999. A human nuclear-localized chaperone that regulates dimerization, DNA binding, and transcriptional activity of bZIP proteins. *Mol. Cell.*, 4:219–28.

Vodermaier, H.C. 2001. Cell. cycle: Waiters serving the Destruction machinery. *Curr. Biol.*, 11:R834–37.

Whitehead, S.E., Jones, K.W., Zhang, X., Cheng, X., Terns, R.M. and Terns, M.P. 2002. Determinants of the interaction of the spinal muscular atrophy disease protein SMN. with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GARl. *J. Biol. Chem.*, 277:48087–93.

Yugami, M., Kabe, Y., Yamaguchi, Y., Wada, T. and Handa, H. 2007. hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett.*, 581:1–7.

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., Oostra, B. and Bagni, C. 2003. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell.*, 112:317–27.

Zhang, A.S., Anderson, S.A., Meyers, K.R., Hernandez, C., Eisenstein, R.S. and Enns, C.A. 2007. Evidence that inhibition of hemojuvelin shedding in response to iron is mediated through neogenin. *J. Biol. Chem.*, 282:12547–56.

Zhang, A.S., West, A.P. Jr., Wyman, A.E., Bjorkman, P.J. and Enns, C.A. 2005. Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. *J. Biol. Chem.*, 280:33885–94.