

# The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRM

Suhyung Cho, Amy Hoang, Sutapa Chakrabarti, Nhat Huynh, De-Bin Huang and Gourisankar Ghosh\*

Department of Chemistry & Biochemistry, University of California, San Diego 9500 Gilman Drive, La Jolla, CA 92093, USA

Received June 21, 2011; Revised and Accepted July 27, 2011

## ABSTRACT

**SR proteins promote spliceosome formation by recognizing exonic splicing enhancers (ESEs) during pre-mRNA splicing. Each SR protein binds diverse ESEs using strategies that are yet to be elucidated. Here, we show that the RNA-binding domain (RBD) of SRSF1 optimally binds to decameric purine rich ESE sequences although locations of purines are not stringently specified. The presence of uracils either within or outside of the recognition site is detrimental for binding with SRSF1. The entire RBD, comprised of two RRMs and a glycine-rich linker, is essential for ESE binding. Mutation within each segment reduced or nearly abolished binding, suggesting that these segments mediate cooperative binding. The linker plays a decisive role in organizing ESE binding. The flanking basic regions of the linker appear to communicate with each other in bringing the two RRMs close together to form the complex with RNA. Our study thus suggests semi-conservative adaptable interaction between ESE and SRSF1, and such binding mode is not only essential for the recognition of plethora of physiological ESE sequences but may also be essential for the interaction with various factors during the spliceosome assembly.**

## INTRODUCTION

SR proteins are sequence-specific RNA-binding factors. RNA binding is a requirement for all of their known cellular activities including the spliceosome assembly. RNA binding is mediated by the RNA recognition motifs (RRMs) within the N-terminal portions of SR proteins. The spliceosome assembly is facilitated by the interaction between the RRMs and exonic splicing enhancers (ESEs) (1,2). The C-terminal RS domain(s) of SR

proteins are thought to serve as modifiers of diverse interactions within the spliceosome.

SRSF1, one of the SR proteins, contains two RRMs at its N-terminus and a relatively short RS domain at the C-terminus. The C-terminal RS domain plays a modulatory role in the SRSF1:ESE complex formation through phosphorylation and dephosphorylation of the serine residues (3). The N-terminal RRM-containing region is responsible for sequence-specific RNA binding. The more N-terminal RRM (RRM1) exhibits clear sequence similarity to the canonical RRM consensus by virtue of its RNP1 and RNP2 motifs. The RRM2 of SRSF1 lacks these motifs. SRSF1-specific ESE sequences have been determined by both binding affinity and functional SELEX experiments (4–6). In addition to selection based and physiological ESE identification, significant progress has been made in identifying ESEs from genome-wide sequences through the use of computational methods (7). All these studies suggest that SRSF1 binds a broad spectrum of ESEs with only a loose consensus among these sequences. ESE bound SRSF1 does not only activate but also represses splicing. Other SR proteins also behave similarly as SRSF1 in terms of loose consensus for their respective ESEs and ability to both activate and repress splicing. The NMR solution structure of the single-RRM containing SR protein SRSF3 (SRp20) bound to a 4-nt ESE RNA demonstrated that SRSF3 recognizes the ESE in a semi-sequence specific manner by using conserved motifs and amino acid residues within its RRM (8). This mode of ESE recognition by SRSF3 provides some clues as to how numerous degenerate ESE sequences within various pre-mRNA might be recognized by SR proteins. RNA-bound structures of non-SR protein splicing factors, Sxl (9), U2AF65 (10), HuD (11) and PTB (12), have also been elucidated. In each case, the complex structure contains two RRMs bound to cognate RNA. These structures reveal diverse modes by which individual RRMs articulate with one another in recognizing specific target RNA. In one case, for example, each RRM of U2AF65 binds to a

\*To whom correspondence should be addressed. Tel: +1 858 822 0469; Fax: +1 858 534 7042; Email: gghosh@ucsd.edu

polypyrimidine tract independent of one another. In contrast, the two PTB RRM1 and RRM2 recognize their target RNA as a single unit with extensive interdomain protein–protein contacts. The Sxl RRM1 and RRM2 exhibit clear cooperativity in their binding to RNA. In all of these cases the conserved RNP1 and RNP2 motifs are directly involved in RNA recognition. The RNA recognition sequences of these splicing factors are highly specific in general. However, it is unclear how SR proteins with two RRM1 and RRM2 bind to such a large and diverse collection of cognate sequences.

This study investigates the mechanism of how the SRSF1 RNA-binding domain (RBD) binds to a large repertoire of putative ESEs in cells. We found that the protein optimally binds RNA sequences of 10-nt in length with no stringent position-specific base requirement with the exception of uracil. The presence of uracils both inside and outside of the recognition sequence is detrimental to binding. All three segments of SRSF1, RRM1, RRM2 and the linker are essential for ESE binding. The flexibly linked segments in SRSF1 RBD recognize RNA using cooperative interactions. Our result thus explains how SRSF1 binds to a large number of ESE to promote splicing.

## MATERIALS AND METHODS

### Cloning and protein expression

His-SRSF1 (RBD, 1–196) and GST-SRSF1 (R1, 1–98) mutants were generated using a site-directed mutagenesis kit (Stratagene). DNA fragments corresponding to WT and mutant SRSF1-RBDs were cloned into pET24dTEV vector, and His-SRSF1 (R1, 1–90 and 1–98), His-SRSF1 (LR2, 90–196 and 105–196) and His-SRSF1 R2 (118–196) were expressed by cloning the corresponding DNA fragments in pET15b. GST-SRSF1 R1 construct was made by cloning the DNA fragment in pGEX-4T2 vector. All proteins were expressed in *Escherichia coli* BL21 (DE3) pLysS cells and grown in M9-based minimal media. Cells were induced with 1 mM IPTG at O.D.<sub>600</sub> 0.8 and grown for 3 h at 25°C. The cell pellet in His-SRSF1 constructs was lysed in 20 mM Tris–HCl (pH 7.5), 500 mM NaCl, 50 mM urea, 5 mM imidazole, 10% glycerol, 10 mM β-mercaptoethanol, 1 mM PMSF and 0.1× protease inhibitor cocktail and soluble fractions were loaded onto DEAE column to remove non-specific RNA or DNA at room temperature. The flow through was loaded onto a Ni<sup>2+</sup>–NTA agarose column at room temperature followed by washing and elution in three steps using lysis buffers containing 20 and 250 mM imidazole in the absence of urea. Proteins were further purified by size exclusion chromatography (Superdex 75, 16/60; GE Health care). GST-SRSF1 (RRM1) wt and mutant proteins were lysed in 20 mM Tris–HCl (pH 7.5), 500 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF and 0.1× protease inhibitor cocktail. Soluble fraction was loaded on glutathione S-transferase sepharose column. The proteins were eluted using 20 mM L-glutathione after washing with lysis buffer. Protein was purified

further by size exclusion chromatography (Superdex 200, 16/60; GE Health care).

### Filter-binding assay

An amount of 10 fmol of [ $\gamma$ -<sup>32</sup>P]-ATP labeled ESEs RNAs was incubated with SRSF1 in 100 μl binding buffer (20 mM Tris–HCl, 75 mM NaCl, 10% glycerol, 0.1% NP40, 1 mM DTT, 2.5 mM MgCl<sub>2</sub>, 10 U RNase inhibitor) at 25°C for 40 min. The reaction mixtures were diluted 1:10 with 900 μl binding buffer and immediately filtered through nitrocellulose membranes (Millipore, 0.45 μm) at a flow rate of 0.5 ml/min and rinsed with 3 ml binding buffer. Membranes were soaked in scintillation cocktail solution (4 ml) after drying at 60°C for 1 h and then the amount of bound RNA was measured using liquid scintillation counter. The membrane after filtering and washing steps with only probe was determined and used as the base line (0%), and the membrane after just spotting of probe without washing was determined and used as 100% binding. The  $K_d$  was estimated as 50% RNA bound fraction. None of the RNAs used in the binding experiment showed any secondary structure as judged by the RNAstructure (ver.5.03) program.

### GST pull-down assay

GST-fusion proteins of 10 μg were mixed with purified target proteins of 15 μg in buffer containing 20 mM Tris (pH 7.9), 100 mM NaCl, 10% glycerol, 1 mM DTT and 0.05 % NP40 at 4°C for 40 min. The mixture was further incubated with 15 μl glutathione sepharose resin (Amersham) for 30 min at 4°C. Resins were washed three times with 400 μl buffer and the bound protein was eluted by boiling with 4× gel loading dye for 5 min at 80°C and was resolved by SDS–PAGE. Separated proteins were visualized by Coomassie staining.

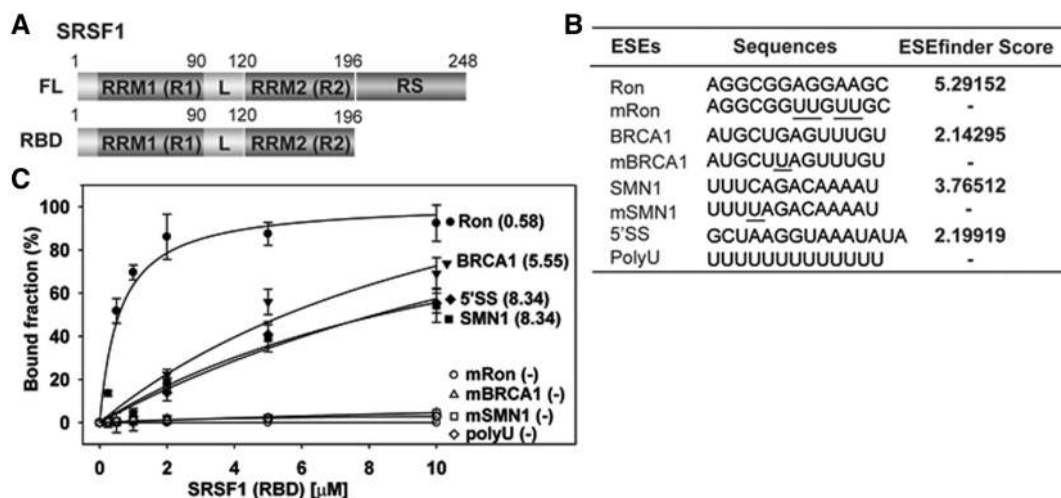
### In vitro splicing assay

β-Globin (*Ron*) in pCDNA with *Ron* ESE sequence was linearized by EcoRI and transcribed with T7 RNA polymerase, in the presence of [ $\alpha$ -<sup>32</sup>P]-UTP as shown previously (3). For *in vitro* splicing, proteins were dialyzed in 20 mM HEPES (pH 8.0), 300 mM KCl, 20% glycerol, 0.5 mM DTT and 0.2 mM EDTA. Pre-mRNA was incubated with HeLa nuclear or cytoplasmic S100 extracts in the presence of wt and mutant SRSF1-RBD as described before (13). Extracted RNA was resolved in denatured 5% acrylamide gel and its phosphorimage was analyzed using typhoon fluorescence scanner (GE healthcare).

## RESULTS

### SRSF1 RBD recognizes ESEs with specificity

To elucidate the mechanism of ESE RNA recognition by SRSF1, we prepared SRSF1 RBD as a highly purified recombinant protein (Supplementary Figure S1 and Figure 1A). We tested the binding affinity of this protein *in vitro* with the ESEs present in the proto-oncogene *Ron* (*Ron*-ESE), which encodes a receptor tyrosine kinase, the



**Figure 1.** ESE binding by SRSF1-RBD. (A) Cartoon representation of SRSF1 domain organization. (B) List of RNA sequences used in *in vitro* binding assay. The bar on the bottom indicates the mutation site. (C) Filter-binding assay showing the binding of SRSF1 RBD to *Ron* ESE, mutant *Ron* ESE (m*Ron*), *BRCA1*, m*BRCA1*, *SMN1*, m*SMN1*, 5'-SS and polyU. Errors (indicated by bar) were obtained from at least three independent experiments. The number in parenthesis denotes the apparent  $K_d$  of binding.

breast cancer-associated gene 1 (*BRCA1*) (*BRCA*-ESE), and the survival motor neuron protein gene 1 (*SMN1*) (*SMN*-ESE) (Figure 1B). In order to investigate the specificity of these ESEs, we also tested binding of the SRSF1 RBD to mutated versions of each of these ESEs. The mutant *Ron* ESE used was tested previously and showed no splicing activity in cells while the naturally occurring *BRCA1*-ESE and *SMN1*-ESE mutants have been linked to disease (Figure 1B) (14–16). A filter-binding (FB) assay was used to evaluate the protein: RNA-binding affinity (Figure 1C). Of the three, the *Ron* ESE exhibited the highest affinity for SRSF1 RBD. However, binding affinity was moderate, measuring in the low micromolar range. The *BRCA1* and *SMN1* ESEs displayed weaker binding affinity. As expected, none of the three mutant ESEs showed any detectable binding (Figure 1C).

We next tested SRSF1 RBD binding to the 5'-SS RNA from the simian virus 40 (SV40) small T antigen (5'-SS) as earlier reports suggested its possible involvement in SRSF1 binding (Figure 1B) (17). We found that the SRSF1 RBD exhibits modest binding to 5'-SS with an affinity that is similar to the *SMN1* ESE (Figure 1C). No interaction was observed when SRSF1 RBD binding to poly-U RNA was assayed as a control (Figure 1C). As an alternative approach, we next ran electrophoretic mobility shift assays (EMSA) to measure the binding affinity of the SRSF1 RBD for *Ron* ESE and the 5'-SS (17) (Supplementary Figure S2A and B). EMSA analysis revealed a similar binding pattern as the FB assay with the *Ron* ESE binding more strongly than 5'-SS to the SRSF1 RBD. Also in agreement with the FB assay results, the poly-U RNA failed to interact with the protein (Supplementary Figure S2C). Therefore, despite their relatively low affinity, SRSF1 RBD binds to ESEs with sequence specificity.

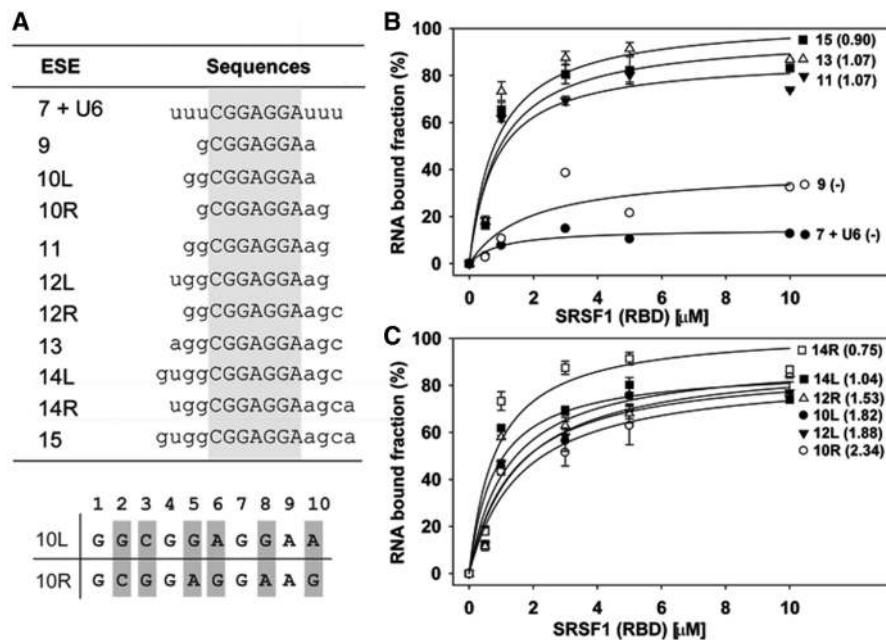
Even though these 5'-SS sequences display some affinity for the SRSF1 RBD, our results do not necessarily imply that SRSF1 binds to these sequences within their natural

cellular context. Our binding data simply suggest that the 5'-SS sequences bear some characteristics of a functional ESE, as predicted by ESEfinder. Moreover, we find that binding affinity does not agree with predictions of the ESEfinder scoring system that were developed based upon ESEs identified by functional SELEX method (Figure 1B and C). This suggests that SRSF1:ESE-binding affinity and splicing efficiency are not necessarily correlated.

#### SRSF1 RBD optimally binds to 10-mer ESE sequences with variable modes

In order to more clearly define the determinants of SRSF1:ESE binding, we employed the *Ron* ESE and SRSF1 RBD and investigated their interactions in greater detail. ESE sequences derived from functional SELEX exhibit only modest conservation with natural ESEs through a region of weak consensus that spans only 7 nt (15,18,19). Curiously, *in vitro* selection for ESEs based solely upon binding affinity identified octameric and decameric consensus sequences (4). To determine whether the SRSF1 RBD binds to RNA sequences that are heptameric or longer, we tested a 13-nt sequence that contained the 7-nt core *Ron* ESE at the center and uridines nucleotides in all the flanking positions as a 7-mer sequence cannot be efficiently radiolabeled (Figure 2A). FB assay revealed little or no binding of this ESE by SRSF1 RBD (Figure 2B). Although a negative role for uracils in the flanking regions cannot be eliminated, this result strongly suggested that SRSF1 mediates base-specific contacts with nucleotides beyond the 7-nt consensus core sequence.

We next tested several RNA sequences that each contained the 7-nt *Ron* ESE core sequence and progressively incorporated natural nucleotide sequences at the flanking regions up to the maximum length of 15 nt (Figure 2A). Initially, we tested five different RNA lengths: 7, 9, 11, 13



**Figure 2.** Determination of optimal ESE length for stable ESE:SRSF1-RBD complex formation. (A) The list of different variants of *Ron* ESE sequences tested for binding (top). Alignment of 10L and 10R sequences (bottom). Highlighted positions denote differences in the nucleotide identity. (B and C) Filter-binding assay showing the binding of SRSF1-RBD to *Ron* ESEs of varying lengths indicated in (A). Error bars represent results from three independent experiments. The number in parenthesis denotes the apparent  $K_d$  of binding.

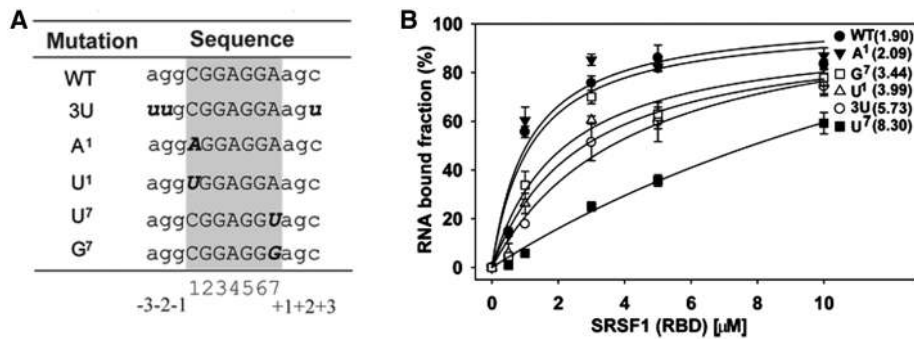
and 15-nt sequences (Figure 2B). All of the RNA possessing 11 natural *Ron* ESE nucleotides or more bound SRSF1 RBD with similar affinities, whereas the 9-nt RNA bound only poorly. These results suggest that the minimal RNA length for native-like *Ron* ESE:SRSF1 RBD interactions might be 10 or 11 nt. To further define the length of recognition sequence, we generated ESE sequences by adding nucleotide(s) at the 5'-end or 3'-end of the 9-nt core to create 10L, 10R, 12L, 12R, 14L and 14R (Figure 2C). Unexpectedly, both 10L and 10R bound the protein with similar affinities and these affinities are comparable to the longer RNA sequences. Sequence comparison of 10L and 10R shows that 6 of 10 positions are different (Figure 2A, bottom). This observation is consistent with the idea that none or only a few base positions within the recognition sequence are stringently fixed to mediate unique interactions with the protein. This mode of binding explains how SRSF1 can accommodate large number of ESE sequences to impose splicing regulation.

#### The presence of uracils within or flanking the ESE reduces binding affinity

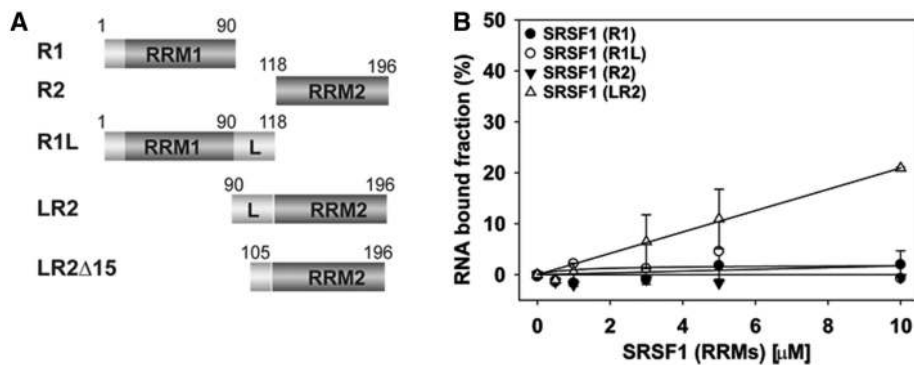
The failure of binding in the 7-mer core *Ron* ESE with flanking uracils to protein suggested that the flanking uracils may also play a negative role in the protein-RNA recognition process. This observation is further supported by the very low binding affinity of SMN1 and BRCA1 ESEs; both of which contain several Us. We further examined the role of uracils within and outside of the core *Ron* ESE sequence in complex formation with the SRSF1 RBD. Having established that a 10-nt

sequence is required for full SRSF1 binding, we altered three flanking nucleotides of the 13-nt long *Ron*-ESE to U (3U) (Figure 3A). Filter-binding assay revealed that the 3U mutant bound with lower affinity to SRSF1 RBD (Figure 3B). Reduced binding by the addition of three Us flanking to the optimized 10-nt sequence suggested that indeed uracils play a negative role in the RNA:protein recognition process. It is well established that poly U sequences are highly flexible as uracils are unstacked compared to adenines in poly A ribonucleotide sequences, which undergo temperature-dependent unfolding (20). Therefore, the possible explanation for negative role of flanking uracils might arise due to the enhanced flexibility of the core RNA sequence in solution resulting in a greater entropic penalty for the complex formation.

We further tested whether uracils within ESE sequences affect SRSF1 RBD:ESE complex formation. We focused on two positions, 1 and 7, within the core sequence of the *Ron* ESE (Figure 3A). Previous reports based on functional splicing assays suggested that position 1 prefers C or G and discriminates against U and A, whereas position 7 prefers A or U (5). We altered position 1 to either A or U, and position 7 to either U or G and used FB assays to measure the binding affinities of these mutant ESEs for SRSF1. We found that, although the presence of U at either of these positions is detrimental to SRSF1 binding, the defect was more severe for U at position 7. Although these binding defects might be due to the loss of direct protein:RNA contact(s), or reduced ability of stacking interactions between the protein and uracils, it is also possible that a U at any position within the protein-binding region increases flexibility of ESE, which in turn negatively affects binding.



**Figure 3.** The role of uridine in SRSF1-RBD binding. (A) The list of the mutated ESE sequences. (B) Filter-binding assay showing the binding of SRSF1-RBD to wt and mutant *Ron* ESE sequences. Errors (shown by bars) were calculated from three independent experiments. The number in parenthesis denotes the apparent  $K_d$  of binding.



**Figure 4.** The entire RBD of SRSF1 is required for optimal ESE binding. (A) Cartoon representation of SRSF1 fragments, R1 (RRM1), R2 (RRM2), R1L (RRM1 with linker), LR2 (linker with RRM2) and LR2 $\Delta$ 15, used for RNA-binding experiments. (B) Filter-binding assay showing the binding of His-tagged SRSF1 constructs (R1, R1L, R2 and LR2) to the 13-mer *Ron* ESE. Error bars obtained from three independent experiments.

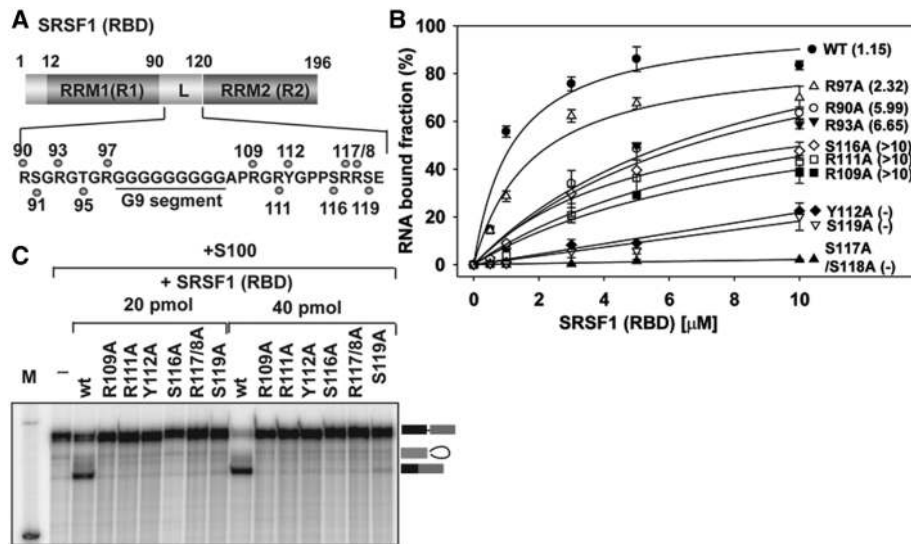
### The linker region plays an essential role in ESE binding

In order to investigate the role played by the SRSF1 RBD in ESE recognition, we generated a series of constructs that encompass individual RRM domains alone, RRM1 (R1) and RRM2 (R2), and with the adjacent linker (L), R1L and LR2 (Figure 4A). Neither R1 (SRSF1 residues 1–90), nor R2, showed any binding to ESE. LR2 (residues 90–196), but not R1L (1–118), showed partial binding (Figure 4B; Supplementary Figure S3A and B). The respective binding affinities displayed by R1 and LR2 fragments defied the standard RNA-binding norms of RRM as R1, but not R2, contains the RNP motifs. However, the binding affinity of LR2 for ESE was significantly lower when compared to that exhibited by the SRSF1 RBD. Taken together, these results suggest that the linker converts R2 into a RNA-binding motif, and that R1 and LR2 mediate cooperative ESE binding. This conclusion is further supported by the fact that no binding was observed when the linker was deleted.

The amino acid composition of the 34-residue long linker indicates it to be highly flexible as it contains 14 glycines. Nine contiguous glycines at the center (G9 segment) separate the two flanking segments containing several arginines interspaced with serines, threonines and tyrosines (Figure 5A). To further elucidate the role of the SRSF1 linker region in ESE binding, we mutated

several residues in both segments of the linker: R90A, R93A, R97A, R109A, R111A, Y112A, S116A, R117A/R118A and S119A (Figure 5A). Upon evaluating the SRSF1 RBD mutants by FB assay we found that each was defective for binding to *Ron* ESE when compared to native SRSF1 RBD (Figure 5B). The R117A/R118A double mutant showed the most severe defect while both Y112A and S119A single mutants exhibited moderate binding. The remaining mutants showed varying degrees of weakened binding affinity. These observations are consistent with a previous crosslinking-based binding assay that showed defective ESE binding by R117A/R118A mutants in RRM2 construct (107–215) (21).

We also examined how the RBD mutants described above affect splicing in S100 splicing complement assay. It was previously shown that RBD alone was able to complement constitutive splicing of  $\beta$ -globin pre-mRNA (3). Therefore, we used the same pre-mRNA in our assay. As expected, wt RBD efficiently complemented splicing; however, none of the linker mutant tested showed any splicing (Figure 5C). Next, we tested the effect of the F56D/F58D mutant that had been previously shown to be defective in splicing. As expected, this mutant failed to complement splicing *in vitro*. These results further support the importance of the linker in ESE binding and consequently in splicing.



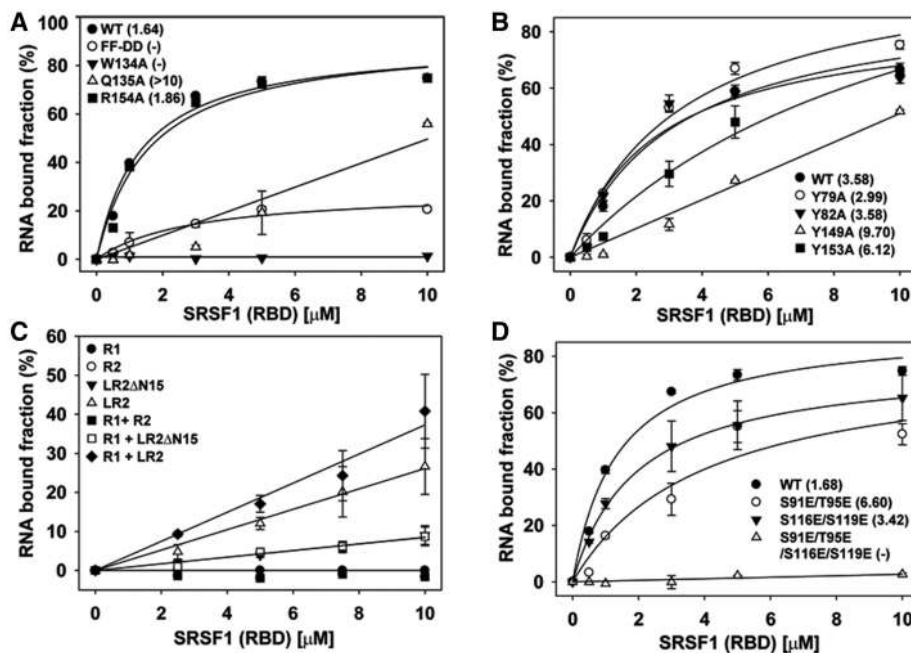
**Figure 5.** The linker in RBD promotes differential ESE binding by SRSF1. (A) The linker sequence and the sites of mutations are denoted in the carton. (B) Filter-binding assay showing the binding of His-SRSF1-RBD wt and different linker mutants. Error bars obtained from three independent experiments. The number in parenthesis denotes the apparent  $K_d$  of binding and (-) indicates not determined. (C) *In vitro* splicing assay of the  $\beta$ -gb (*Ron*) pre-mRNA substrate by SRSF1-RBD and the linker mutants using 20 or 40 pmol with S100 extracts. The template, intermediate and spliced products are marked in the right side of the gel.

### The linker region mediates cooperative binding interactions between ESE and SRSF1-RBD

Relatively lesser role of R1 in ESE binding than R2 made us wonder if the most common RNA-binding motif (RNP1), present only in R1 is involved in ESE recognition. To examine the role of RNP1 in R1, conserved RNP1 residues, F56 and F58, were mutated to aspartates and the double mutant was tested for ESE binding by FB assay. We found that the RBD (F56D/F58D) mutant binds *Ron* ESE poorly implicating the involvement of these two residues in cognate ESE binding (Figure 6A). This result also explains why the F56D/F58D mutant impairs splicing (22,23). Intriguingly, the apparent binding affinity of this RBD mutant is comparable to that of the LR2 fragment suggesting that cooperation between R1 to the rest of the protein in ESE binding is mediated through its conserved RNP1 motif. We next investigated how residues in the R2 domain also participate in ESE binding. R2 domain does not contain consensus RNP motifs but it includes a conserved heptapeptide sequence SWQDLKD, which has been implicated in RNA binding (21). As indicated by the RRM2 structure, W134, Q135 and R154, all reside close to one another on the same positively charged face of R2 (21). Therefore, we hypothesized that this face might likely be involved in ESE recognition. Although R154A was not defective in ESE binding, the W134A and Q135A mutants exhibited highly defective in ESE binding (Figure 6A). Dramatic defect in ESE binding by these two single mutants suggest cooperation between the protein segments in ESE recognition. Moreover, the involvement of four aromatic residues in ESE binding led us to propose that these residues might make stacking interactions as commonly observed in other RNA-protein complexes. To further test extensive nature of aromatic-RNA

contacts, we mutated four other tyrosines (Y79A and Y82A in R1, Y149A and Y153A in R2) located in the same face as F56/F58 or W134. Tyrosine mutants in the R2 (Y149A and Y153A), but not in the R1, showed defects in ESE binding (Figure 6B). This observation further emphasizes more significant role of the R2 in ESE binding.

Since LR2 retains only partial ESE binding, the linker appears to play a special role in bringing R1 to complete the binding event. We hypothesized that the two ends of the linker separated by the central glycine-rich segment are involved in the R1-R2 cooperation. We have measured the binding of LR2 fragments to ESE both in the absence and presence of R1. An enhancement of ESE binding was observed when LR2 was mixed with R1 (Figure 6C and Supplementary Figure S3C-D). However, this enhancement was not observed when the N-terminal part of the linker was removed (LR2 $\Delta$ N15) (Figure 6C). To further test the coupling between the two segments of the linker, we created two double mutants one located in the N-terminal part (S91E/T95E) and the other in the C-terminal part of the linker. Both mutants were constructed in the context of the entire RBD (1-196). The S91E/T95E double mutant showed partial defect in ESE binding (Figure 6D). S119A single mutant was highly defective in ESE binding (Figure 5B), while S116E/S119E double mutant was only marginally defective (Figure 6D). However, when both double mutants were combined (S91E/T95E/S116E/S119E), the resultant quadruple mutant showed no measurable binding affinity (Figure 6D). That is, the defect is not additive but cooperative. Taken together these results suggest that the two end of the linker cooperate with each other in ESE binding and together they cooperate with the R2. Finally, this LR2-ESE subcomplex completes the binding process by recruiting the RNP1 motif of the R1



**Figure 6.** The linker region mediates cooperative binding interactions between ESE and SRSF1-RBD. (A) Filter-binding assay of His-tagged SRSF1-RBD wt and mutants, F56D/F58D from R1, W134A, Q135A and R154A from R2 to *Ron* ESE of 13 nt. Error bars obtained from three independent experiments. (B) Filter-binding assay of His-tagged SRSF1-RBD wt, and four different tyrosine to alanine mutants, Y79A and Y82A from R1, and Y149A and Y153A from R2 to 13-mer *Ron* ESE. Error bars obtained from three independent three experiments. (C) Filter-binding assay showing cooperative interactions between ESE and mixtures of R1 and different constructs of R2 (R2, LR2Δ5 and LR2) of SRSF1. Error bars obtained from three independent three experiments. (D) Filter-binding assay showing cooperation between the N-terminal and C-terminal of the linker of SRSF1 by using three different mutants (S91E/T95E from N-terminal, S116E/S119E from C-terminal, and S91E/T95E/S116E/S119E). The number in parenthesis denotes the apparent  $K_d$  of binding and (–) indicates not determined.

domain. Our model further highlights a novel extensive role played by a large linker in RNA binding.

We further tested if the direct interactions between the two RRM domains also play a role in cooperative interaction using GST pull-down assay. In this assay, interactions between GST-R1 and wt or mutants LR2 were tested in the absence or in the presence of *Ron*-ESE. We found that the two wt fragments do not interact in the absence of RNA (Supplementary Figure S4A). They interact only in the presence of native *Ron*-ESE suggesting that the RNA mediates the binding of two protein fragments (Supplementary Figure S4A). Significant weakening of LR2 retention by the linker mutations further confirms that the linker interaction with RNA is critical for the recruitment of R1 and R2 (Supplementary Figure S4B). However, this result does not preclude if ESE induces further contact between the two RRM domains.

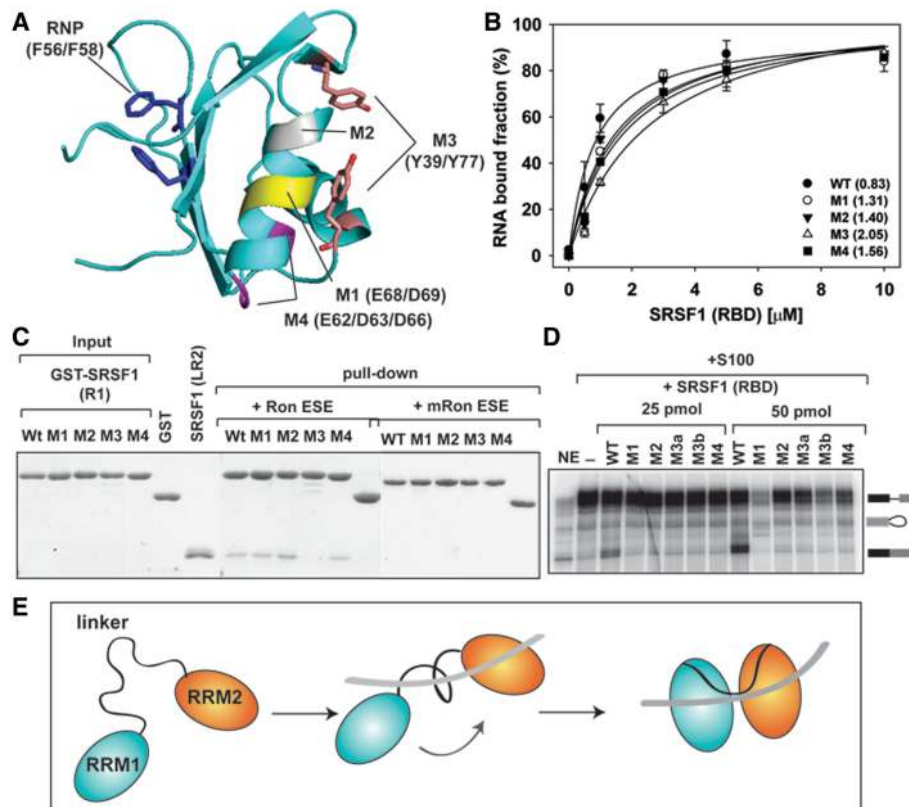
### Cooperative binding brings the two RRM domains in close proximity

To further investigate if the two SRSF1 RRM domains directly cooperate in ESE binding, we first created a model of the ESE:R1 complex based upon the NMR structures of the SRSF3:RNA complex (8) and free SRSF1-RRM1 (PDB, 1X4A) assuming these canonical RRM domains of these two SR proteins bind RNA using a similar mode. For the protein-protein interaction, exposed surface of R1 away from the putative RNA-binding surface might be involved. We identified four surface-exposed patches

that might play such a role (Figure 7A). These patches are composed of residues D68/D69, Y72, Y39/Y77 and D62/D63/D66. We mutated residues in each patch to create mutants M1 (E68A/D69A), M2 (Y72A), M3 (Y39E/Y77A) and M4 (E62A/D63A/D66A). FB assay revealed only minor defects in ESE binding by these mutants (Figure 7B). Our results, therefore, suggest that these patches are not involved in ESE binding.

To test if the two domains interact with each other when present as separate fragments, we carried out GST pull-down experiments using GST-R1 and LR2 as described earlier (Figure 7C). We found that M1, M2 and M4 showed no defect in LR2 retention suggesting that the global RNA-binding modes by R1 and LR2 were preserved (Figure 7C). However, M3 was drastically defective in LR2 retention activity through ESE binding. Our result thus suggests that both or one of the two residues, Y39 and Y77, play a role in ESE binding when two RRM domains are not covalently linked. This result also indicates plasticity in ESE binding by the protein. We propose that when the intact protein binds ESE, two RRM domains do not directly bind but they might lie in close proximity.

We have further tested these RRM1 mutants for their ability to complement *in vitro* splicing (Figure 7D). We found that these are severely defective in splicing even though they are able to bind ESE. We cannot predict the precise reason for their defectiveness. It appears that these residues play roles in the spliceosome assembly in steps other than the RNA recognition.



**Figure 7.** Residues remote from the RNA-binding surface affect ESE recognition (A) The ribbon presentation of the RRM1 domain of SRSF1. The RNP1 residues F56 and F58 are shown in blue. Backbones of residues in four exposed surfaces are denoted by different colors (M1 (E68A/D69A; yellow), M2 (Y72A; gray), M3 (Y39E/Y77A; pink) and M4 (E62A/D63A/D66A; magenta). Additionally, side chains of two M3 residues are shown. (B) Filter-binding assay of His-tagged SRSF1 RBD wt and mutants, (M1, M2, M3 and M4) to *Ron* ESE of 13 nt. Error bars obtained from three independent experiments. The number in parenthesis denotes the apparent  $K_d$  of binding. (C) GST pull-down assay was performed to examine the interaction between 10  $\mu$ g wt GST-SRSF1 (R1) wt or four mutants (M1, M2, M3 and M4) and 15  $\mu$ g His-SRSF1 (LR2) in each the presence of wt *Ron* ESE or mutant *Ron* ESE. As a control, GST protein was used instead of GST-SRSF1 (R1) with wt His-SRSF1 (LR2). SDS-PAGE was resolved in 12.5% acrylamide gel and stained by Coomassie blue. (D) *In vitro* splicing of the  $\beta$ -gb (*Ron*) pre-mRNA substrate using wt SRSF1-RBD and SRSF1-RBD containing M1, M2, M3a (Y39E), M3b (Y77A) and M4 mutants. An amount of 25 and 50 pmol S100 extracts were used for the splicing reactions. (E) A cartoon depicting the mechanism of ESE binding by SRSF1-RBD. In the absence of ESE, the linker domain of the SRSF1 is flexible and probably remains unstructured. However, in the presence of ESE, LR2 binds to ESE and then R1 makes contact to ESE. This binding mode brings R1 and LR2 adjacent each other.

## DISCUSSION

SRSF1 regulates splicing by binding to a broad spectrum of ESE sequences. SELEX experiments in which selection was based either purely on binding or splicing (functional) activity identified significantly different SRSF1-specific RNA sequences: RGAAGAAC and AGGACRRAGC obtained through binding SELEX (4) and SRSASGA (S = C or G) by functional SELEX (5). Recent CLIP method identified an even more diverse consensus sequence for SRSF1 (UGRWG, R:purine; W:A/G) (24). Results presented here explain how diverse ESE sequences can be recognized: we show that the decameric ESE sequences, but not the shorter sequences as generally thought, are optimal for SRSF1 binding. Our finding that the 10th nucleotide can be added to the either end to obtain maximal binding affinity strongly points toward a semi-conservative binding mode since seven positions are different between these two sequences. A and G substitute each other in these differed positions. Since A and G are decorated with different functional groups and their

hydrogen bonding capacity is different, we conclude that the complex formation is dominated by stacking interactions. Therefore, we suggest sequence-specific hydrogen bonding contacts between the protein and RNA might not be as distinctive a feature as the stacking interactions between bases and aromatic side chains. Purines are better suited for stacking than pyrimidines, this explains why SRSF1-specific ESEs are dominated by the purine residues. The presence of uracil both inside and flanking the recognition sequence is less permissible to SRSF1 binding. The presence of uracil destabilizes the complex, perhaps, due to its higher flexibility compared to other bases in the oligonucleotide and/or due to its reduced stacking interactions with the protein. The NMR solution structure of the single RRM-containing SR protein, SRSF3 (SRp20), bound to a 4-nt (CAUC) ESE shows interesting property of the complex: this structure shows protein bound to the RNA primarily through stacking interactions with only one base-specific hydrogen-bonding contact (8). However, the significance of this base-specific contact awaits



further investigation. Our results when considered through this structural study a novel RNA: protein recognition strategy can be predicted between SR proteins and ESE sequences. All three segments of the RBD; the linker and the RRM are required for ESE binding. Most of the single or double mutations drastically reduced binding affinity suggesting that the RNA contacting residues are in a part of interaction network and recognize RNA in a cooperative manner. These observations suggest a novel ESE-binding mechanism by SRSF1 RBD where the distal protein segments assemble around the target RNA stabilize the protein–RNA complex through strongly cooperative intra-molecular protein–protein and inter-molecular protein–RNA contacts. Since the linker is flexible in nature, it can bind to a variety of sequences with low sequence specificity before properly orienting the RRM to make specific and semi-specific contacts. In such a binding, protein–protein contact between the linker and RRM2 or RRM1 may contribute significant binding energy in forming the complex. This explains why so many single or double mutants practically abolish the complex formation. This coupled binding also explains how small variation in the linker can greatly inhibit the association process. For instance, the modification of the linker would greatly affect RNA binding. Indeed, it has been shown that methylation of arginines (R93, R97 and R109) impact on splicing (25). It is possible that phosphorylation of serines and threonine will reduce RNA binding and hence would negatively affect splicing.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Drs Simpson Joseph, Joseph Adams, Xiang-Dong Fu and Tom Huxford for their comments on the manuscript.

## FUNDING

Funding for open access charge: National Institutes of Health (NIH) in U.S. (grant number GM084277 to G.G.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.*, **25**, 106–110.
- Stojdl, D.F. and Bell, J.C. (1999) SR protein kinases: the splice of life. *Biochem. Cell Biol.*, **77**, 293–298.
- Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R. and Ghosh, G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl Acad. Sci. USA*, **108**, 8233–8238.
- Tacke, R. and Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.*, **14**, 3540–3551.
- Liu, H.X., Zhang, M. and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.*, **12**, 1998–2012.
- Schaal, T.D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell Biol.*, **19**, 1705–1719.
- Zhang, X.H. and Chasin, L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241–1250.
- Hargous, Y., Hautbergue, G.M., Tintaru, A.M., Skrisovska, L., Golovanov, A.P., Stevenin, J., Lian, L.Y., Wilson, S.A. and Allain, F.H. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.*, **25**, 5126–5137.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579–585.
- Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R. and Kielkopf, C.L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell*, **23**, 49–59.
- Wang, X. and Tanaka Hall, T.M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.*, **8**, 141–145.
- Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. et al. (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054–2057.
- Mayed, A. and Krainer, A.R. (1999) Mammalian in vitro splicing assays. *Methods Mol. Biol.*, **118**, 315–321.
- Liu, H.X., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat. Genet.*, **27**, 55–58.
- Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.*, **30**, 377–384.
- Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S. and Biamonti, G. (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell*, **20**, 881–890.
- Zuo, P. and Manley, J.L. (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc. Natl Acad. Sci. USA*, **91**, 3363–3367.
- Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q. and Krainer, A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.*, **15**, 2490–2508.
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568–3571.
- Inners, L.D. and Felsenfeld, G. (1970) Conformation of polyribouridylic acid in solution. *J. Mol. Biol.*, **50**, 373–389.
- Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Hung, M.L., Lian, L.Y., Craven, C.J. and Wilson, S.A. (2007) Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.*, **8**, 756–762.
- Caceres, J.F. and Krainer, A.R. (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J.*, **12**, 4715–4726.
- Zuo, P. and Manley, J.L. (1993) Functional domains of the human splicing factor ASF/SF2. *EMBO J.*, **12**, 4727–4737.
- Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W. and Caceres, J.F. (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One*, **3**, e3369.
- Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M.P. and Krainer, A.R. (2010) Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol. Cell Biol.*, **30**, 2762–2774.