

Inter-domain Flexibility of Human Ser/Arg-Rich Splicing Factor 1 Allows Variable Spacer Length in Cognate RNA's Bipartite Motifs

Naiduwadura Ivon Upekala De Silva, Talia Fargason, Zihan Zhang, Ting Wang, and Jun Zhang*



Cite This: *Biochemistry* 2022, 61, 2922–2932



Read Online

ACCESS |



Metrics & More

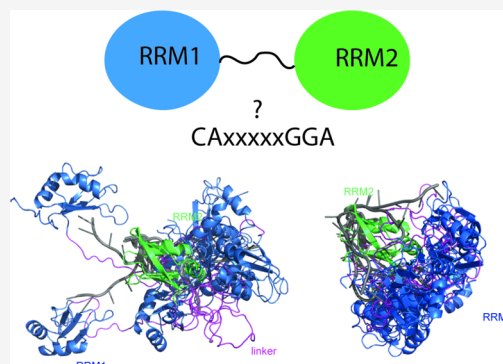


Article Recommendations



Supporting Information

ABSTRACT: Ser/Arg-rich splicing factor 1 (SRSF1 or ASF/SF2) is the prototypical member of SR proteins. SRSF1 binds to exonic splicing enhancers, which prompts inclusion of corresponding exons in the mature mRNA. The RNA-binding domain of SRSF1 consists of tandem RNA-recognition motifs (RRM1 and RRM2) separated by a 30 amino acid long linker. In this study, we investigate roles of RRM1, RRM2, and the linker in RNA binding. We find that although both RRMs are crucial to RNA binding, RRM2 plays the dominant role. The linker mildly contributes to RNA binding and remains flexible in the RNA-bound state. Flexibility of the linker allows the RRM1-cognate motif to be either upstream or downstream of the RRM2-cognate motif. In addition, we find that the spacer length between the bipartite motifs varies from 0 to 10 nucleotides. Our binding assays reveal that SRSF1 prefers RNA sequences with shorter spacers and the RRM1-cognate motif being placed upstream. Restrained by nuclear magnetic resonance data, we simulate RNA-bound complexes and demonstrate how tandem RRMs bind to RNA of different spacer lengths and swapped bipartite motifs. We find that when the RRM1-cognate motif is placed downstream, either the RRM1/RRM2 linker needs to be more extended or RNA needs to form a U turn, which may reduce conformational entropy. Our study suggests that the RNA-binding specificity of SRSF1 is broader than traditionally recapitulated by consensus sequences of 7 to 8 nucleotides. Instead, centered on the RRM2-cognate motif, an RNA fragment encompassing 10-nucleotide upstream and downstream should be scrutinized.



INTRODUCTION

Ser–Arg (SR) proteins are a family of splicing factors that play key roles in constitutive and alternative splicing. The SR family consists of 12 members and is characterized by one to two RNA-recognition motifs (RRM) and a repetitive Arg–Ser dipeptide region (RS).^{1,2} RRM domains of SR proteins specifically bind to exonic splicing enhancer (ESE) RNA elements. RS regions of SR proteins are subject to phosphorylation and mediate protein–protein interactions with U1 snRNP and U2 snRNP.^{3–13} Through these interactions, SR proteins initiate spliceosome assembly and usually facilitate inclusion of the bound exons into mature mRNA.^{10–12} Neighboring exons compete with each other for SR proteins, and therefore, exons that bind to SR proteins tighter are more likely to be included, while exons with weaker ESEs tend to be skipped.^{1,2,14} Therefore, RNA-binding specificity is of fundamental importance to SR functions in alternative splicing.

As the prototypical member of the family, SRSF1 has long been a model for studying functions of SR proteins. SRSF1 contains tandem RRMs (RRM1 and RRM2), followed by an RS tail (Figure 1A). In addition to its roles in RNA splicing, SRSF1 is also essential for genome stability,¹⁵ mRNA transcription,^{16,17} transport,¹⁸ translation,^{19,20} nonsense-mediated mRNA decay,²¹ immune response,²² and regulation of

long noncoding RNA.^{23,24} Investigating SRSF1 is of clinical importance, as it is an oncoprotein involved in processing more than 1500 mRNA transcripts.²⁵ Elevation of the SRSF1 level,^{26–30} mutation of the protein,³¹ or improper phosphorylation of the RS tail^{32–35} can cause various cancers.

RNA-binding specificity of SRSF1 has been investigated by several previous studies using systematic evolution of ligands by exponential enrichment (SELEX) and high throughput sequencing and crosslinking immunoprecipitation (HITS-CLIP).^{36–40} The SELEX study by Tacke and co-workers found that SRSF1 tandem RRMs prefer octamer sequences RGAACAAC (R is purine) and two decamer sequences AGGACAGAGC or AGGACGAAGC.³⁶ However, the SELEX study performed by Liu and co-workers found that tandem RRMs can recognize a broader range of sequences with a consensus sequence SRSASGA (S represents G or C and R represents purine).³⁸ In contrast to these in vitro studies, in

Received: September 30, 2022

Revised: November 17, 2022

Published: December 1, 2022



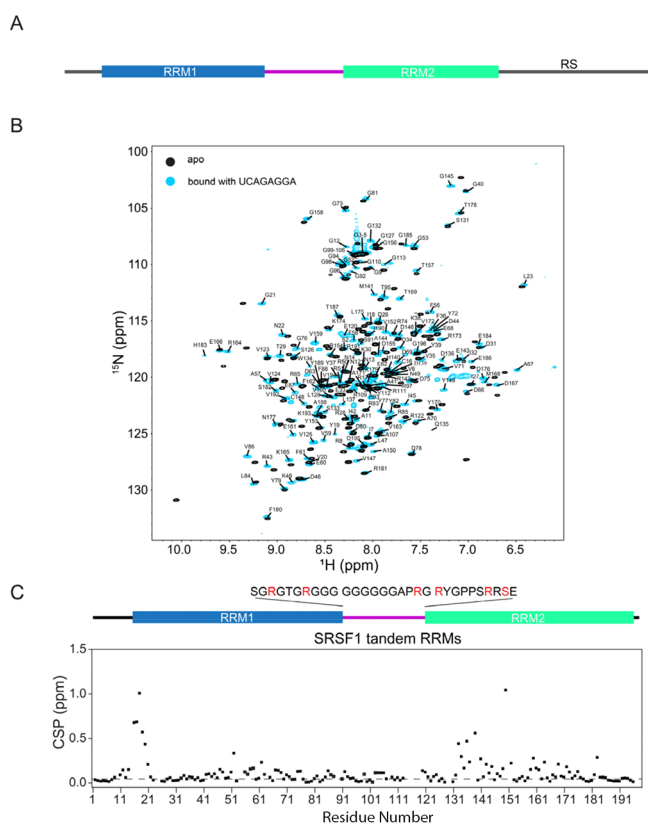


Figure 1. Chemical shift perturbation (CSP) analysis of SRSF1 RNA binding. (A) Domain architecture of full-length SRSF1. (B) HSQC overlaying of apo and RNA-bound SRSF1 tandem RRM1. The RNA sequence is UCAGAGGA. (C) CSP of SRSF1 tandem RRM1 caused by RNA binding. The domain architecture of SRSF1 tandem RRM1 is aligned with the residue number. The dashed line indicates 0.05 ppm. CSP values were calculated as $|\delta^1\text{H}| + 0.1 \times |\delta^{15}\text{N}|$, where $\delta^1\text{H}$ and $\delta^{15}\text{N}$ are the chemical shift differences between the apo and bound states for proton and nitrogen, respectively. The linker residues probed by mutation are shown in red fonts.

vivo HITS-CLIP by Sanford and co-workers revealed that SRSF1-cognate RNA has a consensus sequence of GAA-GAA.^{39,40} Still different from above studies, recent work by Feng found that SRSF1 recognizes short RNA motifs with clusters of GGA sites.⁴¹

Structural studies on isolated RRM1s have been performed to elucidate the RNA recognition mechanism of SRSF1. The nuclear magnetic resonance (NMR) structure of the RRM2/RNA complex reveals that RRM2 binds to GGA using a noncanonical RNA-binding site.⁴² RRM2 alone is sufficient for splicing some pre-mRNA transcripts.⁴² However, another study found that both RRM domains are required for optimal RNA binding and concluded that the RRM1/RRM2 linker rich in Arg residues is involved in RNA binding.⁴³ The role of RRM1 in RNA binding was unknown until a recent study determined the RNA-bound structure of isolated RRM1.⁴⁴ This study revealed that SRSF1 RRM1 binds with a 2-nucleotide motif with the first site being C and any nucleotide for the second site.⁴⁴ The same study also found that the C-containing motif for RRM1 can be either 4-nucleotide upstream the 5' of the GGA motif for RRM2 or 6-nucleotide downstream the 3' of the GGA motif, and swapping bipartite motifs does not affect binding affinity.⁴⁴

Despite the extensive research on SRSF1, our understanding of the RNA-binding specificity of SRSF1 is still limited. A key question is why different studies have obtained such divergent sequences for SRSF1-cognate RNA. In addition, what is the role of the RRM1/RRM2 linker in RNA binding, how are the tandem RRM1s arranged to bind RNA, and how does the RNA spacer between the bipartite motifs affect binding? Answering these questions needs studies in the context of tandem RRM1s. Using mutagenesis and fluorescence polarization (FP) binding assays, we find that both RRM1s are required for RNA binding, and RRM2 plays the dominant role. Our NMR relaxation results show that the RRM1/RRM2 linker remains flexible in the RNA-bound state and that its basic residues only moderately contribute to RNA binding. We find that the linker flexibility allows the spacer between bipartite motifs to vary from 0 to 10 nucleotides, in addition to making the bipartite motifs swappable. We also find that SRSF1 tandem RRM1s prefer RNA with RRM1-cognate motif upstream or RNA with shorter spacers. With restraints of NMR paramagnetic relaxation enhancement (PRE), we simulate the conformational heterogeneity of RNA-bound SRSF1 tandem RRM1s. We demonstrate how SRSF1 tandem RRM1s achieve the diversity in RNA binding and explain why SRSF1 prefers RNA with a shorter linker and RRM1-cognate motif upstream. Our findings suggest that RNA-binding specificity of SRSF1 cannot be comprehensively recapitulated by consensus sequences that are traditionally obtained by sequence alignment of motifs with a fixed length. Therefore, our work helps to resolve seeming contradictions among reported studies and advances our understanding of RNA-binding specificity of SRSF1.

MATERIALS AND METHODS

Molecular Cloning and Protein Expression. The DNA encoding human SRSF1 tandem RRM1s (residues 1–196) was subcloned into pSMT3 using BamH I and Hind III. SRSF1 constructs RRM1 (residues 1–109) and RRM2 (residues 110–196) and their mutants were prepared using mutagenesis polymerase chain reaction. All proteins were expressed by BL21-CodonPlus (DE3) cells in LB media or minimal media supplemented with proper isotopes. Cells were cultured at 37 °C to reach an OD₆₀₀ of 0.6, when 0.5 mM isopropylthio-β-galactoside (IPTG) was added to induce protein expression. Cells were further cultured for 16 h at 22 °C. The cells were harvested by centrifugation at 3000 RCF for 15 min and stored at −80 °C before use.

Protein Purification. The cell pellets were resuspended in 20 mM Tris–HCl, pH 7.5, 2 M NaCl, 25 mM imidazole, 0.2 mM TCEP, 1 mM PMSF, 0.5 mg/mL lysozyme, and 1 tablet of protease inhibitor. After three freeze–thaw cycles, the sample was sonicated and centrifuged at 23,710 RCF for 40 min using a Beckman Coulter Avanti JXN26/JA20 centrifuge. The supernatant was loaded onto 5 mL of HisPur Nickel-NTA resin and washed with 200 mL of 20 mM Tris–HCl, pH 7.5, 2 M NaCl, 25 mM imidazole, and 0.2 mM TCEP. The sample was then eluted with 30 mL of 20 mM 2-Morpholinoethanesulfonic acid sodium salt (MES), pH 6.5, 500 mM imidazole, 500 mM Arg/Glu, and 0.2 mM TCEP. The eluted sample was cleaved with 2 μg/mL Ulp1 for 2 h at 37 °C and diluted by threefold with a buffer A of 20 mM MES, pH 6.0, 100 mM Arg/Glu, and 0.1 mM TCEP before being loaded onto a 5-mL HiTrap Heparin column. The sample was eluted over a gradient with a buffer B of 20 mM MES, pH 6.0, 100 mM Arg/

Glu, 0.1 mM TCEP, 2 M NaCl, and 0.02% NaN₃. The tandem RRM and their mutants were eluted around 50% B. Fractions containing the target proteins were pooled, concentrated, and loaded onto a HiLoad 16/600 Superdex 75 pg size exclusion column equilibrated with 0.2 M Arg/Glu, 0.2 M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1 mM TCEP, and 0.02% NaN₃.

Fluorescence Polarization Assays. FP assays were carried out using 10 nM 5' fluorescein-labeled RNA (product of Dharmacon) mixed with SRSF1 constructs at concentrations ranging from 8000 nM to 0.488 nM by twofold serial dilutions in 20 mM HEPES, pH 7.5, 200 mM Arg/Glu, 200 mM NaCl, 0.02% Tween, and 0.1 mM TCEP. One hundred microliters of samples were mixed in black flat-bottom 96-well plates (Costar) by shaking at 100 RPM for 5 min, followed by incubation at 37 °C for 10 min and incubation at 25 °C for 30 min. The FP data were gathered at room temperature using a Cytation 5 Cell Imaging Multimode reader with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The binding affinities were determined using nonlinear regression for one-site interaction using GraphPad Prism 7. The fluorescence polarization F_p was fitted using the quadratic equation below, where the fitting parameters $F_{\min}$, $F_{\max}$, and K_D are the FP baseline, plateau, and dissociation constant, respectively. $[P_T]$ is the total protein concentration, and $[L_T]$ is the total RNA concentration (10 nM). Errors of dissociation constants were calculated based on three independent measurements.

$$F_p = F_{\min} + (F_{\max} - F_{\min}) \times \left\{ \frac{[(P_T) + (L_T) + K_D] - \{[(P_T) + (L_T) + K_D]^2 - 4[P_T][L_T]\}^{0.5}}{2[L_T]} \right\} \quad (1)$$

The above equation can be used to determine binding affinity, assuming that the binding process does not change the quantum yield of RNA. We confirmed that this assumption is true as the total fluorescence intensity of fluorescein-labeled RNA does not change with protein concentrations (Figure S1A).

Differential Scanning Fluorimetry Protein Unfolding. Differential scanning fluorimetry (DSF) experiments were performed in triplicate for the wild-type RRM tandem and its mutants (RRM1 and linker mutants) after diluting to 20.1 μM in 0.2 M Arg/Glu, 0.2 M NaCl, 20 mM Tris-HCl, pH 7.5, 0.1 mM TCEP, and 0.02% NaN₃. Promethius NT.48 nanoDSF (Temper Technologies) was employed to measure the fluorescence emission intensities at 330 and 350 nm while increasing the temperature over a range of 20–90 °C at a rate of 1.0 °C/min. Melting curves were generated and averaged to calculate the first derivative of the fluorescence emission intensity using the software PR. ThermControl v2.1.1. The average ratio of the first derivatives of the fluorescence emission intensities at 330 nm was plotted against temperature to compare thermal stabilities between the wild-type protein and its mutants.

NMR Assignment Experiments. The tandem RRM construct (residues 1–196) was prepared as described above except that the *E. coli* cells were grown in M9 media containing ¹⁵N, ¹³C, and ²H isotopes. The apo protein (436 μM) was purified as described above and exchanged into 20 mM MES, pH 6.2, 380 mM Arg/Glu, and 5% D₂O for NMR measurements. To prepare the 8-mer UCAGAGGA bound tandem RRM, RNA was added to 620 μM. Triple resonance

assignment experiments HNCA, HNCACB, HN(CO)CA, CBCA(CO)NH, HN(CA)CO, and HNCO were performed at 37 °C using a Bruker Avance III-HD 850 MHz spectrometer installed with a cryo-probe. The NMR data were processed using NMRPipe,⁴⁵ and assignment was performed using NMRViewJ.⁴⁶

T_2 Measurements and Paramagnetic Relaxation Enhancement. For T_2 measurements, intensities were gathered at 17.0, 34.0, 67.8, 135.7, 170.0, 203.5, 237.4, and 271.4 ms, in which 34 and 203.5 ms planes were measured as duplicates to estimate error. A recycle delay of 2.0 s was used. T_2 was fitted using a single exponential decay as below:

$$I(t) = I_0 \times e^{-t/T_2} \quad (2)$$

where $I(t)$ is the intensity at time point t , I_0 is the signal intensity at $t = 0$, and T_2 is the transverse relaxation time constant. The error in T_2 measurements was estimated from the duplicate measurements of 34 and 204 ms planes.

¹H PRE data were gathered at 37 °C using a Bruker AVANCE III-HD 600 MHz spectrometer installed with a cryo-probe. Native C16 and C148 were mutated to serine, and these mutations did not change RNA-binding affinity compared with the wild-type protein (data not shown). The proteins were prepared as described above except that the *E. coli* cells were grown in M9 media containing ¹⁵N isotopes. Immediately before paramagnetic labeling with S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), TCEP was removed by loading the sample onto a HiPrep 26/10 desalting column (GE) equilibrated with 400 mM Arg/Glu, pH 7.5. The protein was diluted to 40 μM and mixed with 200 μM MTSL for overnight reaction at 25 °C. Unreacted MTSL was removed by loading the sample onto a HiPrep 26/10 desalting column (GE) equilibrated in 20 mM MES, pH 6.2, 380 mM Arg/Glu, and 5% D₂O. The protein samples were concentrated to 200 μM and mixed with 600 μM RNA. We also confirmed that SRSF1 tandem RRM binds with RNA with an affinity of 40 nM in the NMR buffer. The PRE measurements were carried out using a pulse sequence developed by Iwahara et al.⁴⁷ A total of 80 scans were accumulated, and the time interval was set to 15 ms. Diamagnetic data were collected with the sample quenched using 2 mM ascorbic acid. The NMR data were processed using NMRPipe⁴⁵ and analyzed using NMRViewJ.⁴⁶

Small-Angle X-ray Scattering (SAXS). SAXS data (0.013 to 0.4 Å⁻¹) were collected at the SIBYLS beamline (12.3.1) of the Advanced Light Source at room temperature. SRSF1 tandem RRM and RNA UCAGAGGA were mixed into a ratio of 1:2 and exchanged into a buffer containing 20 mM MES, pH 6.2, 380 mM Arg/Glu, and 2 mM DTT by centrifugal filters with a cut-off molecular weight of 10 kDa. Samples were prepared at three different concentrations, and the filter-through solution was used as SAXS reference. SAXS data were analyzed with the ATSAS package, and the Guinier plot was used to determine the radius of gyration.⁴⁸

Xplor-NIH Simulations. The RNA-bound structures of SRSF1 tandem RRM were simulated with Xplor-NIH. RRM1 (residues 14–90) and RRM2 (residues 121–196) were treated as rigid bodies, while the N-terminal region (residues 1–13) and the linker (91–120) were given full degrees of freedom.

We assumed that RRM domains in the tandem interact with RNA in a way similar to their isolated domains. For 16-mer RNA (2, S2, 0, S0, 2, S2, 6, S6, 10, and S10 U), RRM1 and RRM2 were restrained to interact with CA and GGA motifs,

respectively. This was achieved by employing 37 distance restraints measured between the RRM1 and their interacting RNA in the NMR structures of the isolated RRM domains (AAAACA bound RRM1, PDB ID: 6HPJ, UGAAGGGAC bound RRM2, PDB ID: 2M8D). These distances were measured between the RNA-binding residues of each RRM and their cognate RNA motif (CA for RRM1 and GGA for RRM2) that participates in π - π stacking interactions and H bonds. These distance restraints were introduced into Xplor-NIH as ambiguous nuclear Overhauser effect distances. In structure calculation, the degree of freedom was initially randomized, and gradient minimization was performed, followed by a standard simulated annealing protocol. For 16-mer RNA-bound tandem RRMs, 100 conformers were simulated for each RNA-bound tandem RRM. Out of them, top 10 structures with the lowest energy were selected to investigate how the length of the RNA spacer between CA and GGA motifs facilitates RNA binding to the RRM tandem.

PRE data on tandem RRMs bound with uuuCAuuGGAuu or uGGAuuuuuCAu were used as inter-domain distance restraints to simulate 100 structures of 8-member ensembles. In this method, PRE values were used as the distance restraints between amide protons and the oxygen atom bearing the unpaired electron of MTSL conjugated to E120C. The amide protons of the bleached residues (the ones whose resonances disappeared due to close proximity to MTSL) were restrained within 15 Å to MTSL ambiguous distance restraints. Efficiency of MTSL labeling was estimated to be >95% by measuring the residual intensity of the residues close to E120C in structure. We have also confirmed that MTSL labeling did not affect RNA binding. The MTSL paramagnetic probe was represented with three conformers in order to account for its flexibility. In addition to the PRE restraints, RRM1 and RRM2 were restrained to interact with CA and GGA motifs, respectively, as detailed above. The scripts and parameters can be obtained upon request. The lowest energy ensembles were selected to analyze how the domain arrangements change when the CA and GGA motifs were swapped. We selected the ensembles that contain the lowest q factors (≤ 0.1) and the highest correlations values (≥ 0.9) for both PRE and the protein–RNA distance restraints. Therefore, those experimental values are in a well agreement with the restraints calculated by the software.⁴⁹

RESULTS

Roles of RRM1 and the RRM1/RRM2 Linker in RNA Binding. SRSF1 consists of tandem RRMs and an RS domain (Figure 1A). It is believed that RNA binding primarily involves the tandem RRMs, which consist of four segments: the N-terminal region (residues 1–15), RRM1 (residues 16–89), the linker (residues 90–120), and RRM2 (residues 121–193) (Figure 1C). Using filter-binding assays, a previous study has shown that deletion of either RRM completely abolishes RNA binding.⁴³ Mutation of the linker residues R109 and R111 decreases binding affinity by >10-fold, and an S119A mutation completely abolishes RNA binding.⁴³ The filter-binding assay is not optimal for weak interactions, and its bias against weak binding has been noticed in the literature.^{50,51} Although the RNA-bound structure of isolated RRM1 has been determined by NMR,⁴⁴ its RNA-binding role in the tandem RRMs is still unclear.

To determine the roles of RRM1 and the linker in RNA binding, we compared the chemical shifts of the tandem RRMs

in the apo and RNA-bound states (Figure 1B). As chemical shift is sensitive to the local environment of nuclei, chemical shift perturbations (CSPs) can be used to probe conformational changes or proximity of ligands. Significant perturbations were observed for both RRM1 and RRM2 residues when the classic SRSF1 cognate RNA, UCAGAGGA, was added into the tandem RRMs (Figure 1C).⁵² Using FP binding assays, we found that deleting either RRM1 or RRM2 reduced the binding to an extent beyond detection (Figure S1B). The RRM2 residues showing CSPs match the reported RNA-binding sites of isolated RRM2 (Figure 1C).⁴² To confirm our CSP findings on RRM1, we mutated the RRM1 residues that show significant perturbations (Table 1; Figure S1C). Here,

Table 1. Relative Dissociation Constants (K_D) of RRM1 Mutants in SRSF1 Tandem RRMs

RRM1 mutants	relative K_D ^a
R8A	3.9 ± 0.63
R17A	3.3 ± 0.11
Y19A	34 ± 1.8
R50A	3.2 ± 0.72
F56A	17 ± 1.3
F58A	25.7 ± 0.66
R65A	1.9 ± 0.18
R85A	2.1 ± 0.11
R90A	3.1 ± 0.35

^aThe relative K_D values were calculated as the ratios to the K_D of the wild-type protein (24 ± 1.7 nM). The standard error of the mean was estimated from three technical replicates. The RNA sequence was UCAGAGGA in these assays.

residue R65 was selected as a negative control, as it is distal to the canonical RNA-binding site of RRMs. All these mutants were eluted at the same position as the wild-type protein in size exclusion chromatography. Our differential scanning fluorimetry results also suggest that most mutants have a melting temperature of 53 °C, identical to the wild-type protein (Figure S1D). The only exception is S119A, which has a lower melting temperature of 49 °C. Therefore, these mutants maintain the overall structure and the observed changes in RNA-binding affinities are due to loss of important residues for binding, instead of altered protein structure or protein misfolding. Significant binding affinity decreases were observed for mutation of Y19, F56, and F58 (Figures 2A and S1E). Y19 and F58 form stacking interactions with RNA in the isolated RRM1 structure (PDB ID: 6HPJ). F56 is not directly involved in RNA binding, but it is packed against F58 and stabilizes the F58 sidechain. This may explain why mutation of this residue weakens RNA binding. The residues showing large CSPs are located on the β 1 and β 3 strands, which are known as the canonical RNA-binding sites of RRMs (Figure S1F).⁵³ We also found that mutation of basic residues generally decreases RNA binding by 2 to 4 fold. Consistent with the NMR complex structure, our CSP analysis and FP binding assays suggest that residues Y19 and F58 are critical for RNA binding in the tandem RRMs.

We continued to examine the role of the RRM1/RRM2 linker in RNA binding. The linker contains basic residues at the two ends and a 9-Gly segment in the middle (Figure 1C). We found that mutation of basic residues generally decreases binding affinity around twofold (Figure 2B; Table 2), and the linker residues do not display significant CSPs upon RNA

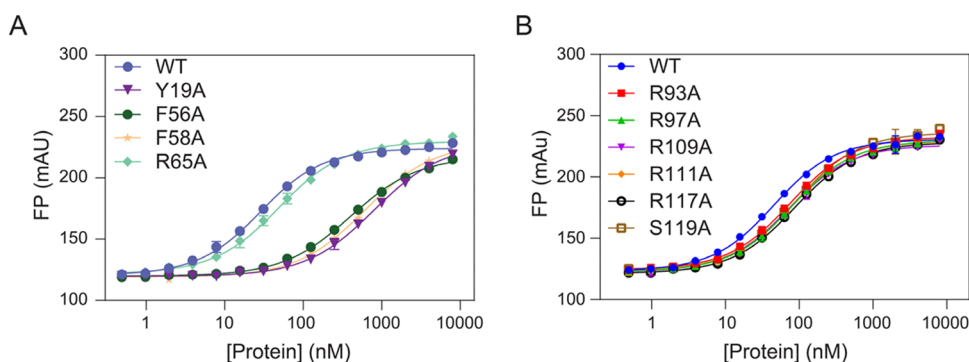


Figure 2. Roles of RRM1 and the RRM1/RRM2 linker in RNA binding. (A) Representative FP binding curves for RRM1 mutants in the tandem. FP binding curves for other mutants are shown in Figure S1C. (B) FP binding curves for the RRM1/RRM2 linker mutants. The standard error of the mean was estimated from three technical replicates.

Table 2. Relative Dissociation Constants of RRM1/RRM2 Linker Mutants in SRSF1 Tandem RRMs

linker mutants	relative K_D^a
R93A	1.75 ± 0.07
R97A	1.9 ± 0.11
R109A	1.7 ± 0.28
R111A	1.83 ± 0.06
R117A	2.1 ± 0.12
S119A	2.1 ± 0.16

^aThe relative K_D values were calculated as the ratios to the K_D of the wild-type protein (40.0 ± 0.52 nM). Note that the K_D of the wild-type protein is different from that in Table 1, as they were performed in different batches, although great care had been taken to create identical binding conditions. The standard error of the mean was estimated from three technical replicates. The RNA sequence was UCAGAGGA in these assays.

binding (Figure 1C). S119A only decreases RNA binding by twofold. Deletion of the 9-Gly segment does not significantly affect RNA binding (data not shown). Therefore, we concluded that while the basic residues in the linker slightly contribute to RNA binding, the major role of the linker in RNA binding is to tether RRM1 and RRM2 to each other.

Flexibility of the RRM1/RRM2 Linker in the RNA Bound State. Using filter-binding assays, a previous study concluded that the RRM1/RRM2 linker is essential for RNA binding.⁴³ However, our results showed that mutation of linker basic residues only slightly decreases RNA binding. Our binding assays imply that the RRM1/RRM2 linker is unlikely

to form a stable interaction with RNA. To characterize the structural dynamics of the linker, we analyzed the ^{15}N transverse relaxation time constant (T_2) for RNA-bound tandem RRMs (Figures 3 and S2). T_2 values are sensitive to molecular tumbling and local structure dynamics. Unstructured protein regions, such as inner loops and flexible termini, have longer T_2 values relative to structured protein regions.⁵⁴ As shown in Figure 3, T_2 values for the N-terminal region (residues 1–11) and the RRM1/RRM2 linker are 3–4 fold longer than those for RRM1 and RRM2. These results suggested that the N-terminal region and the RRM1/RRM2 linker remain unstructured in the RNA-bound state, consistent with our mutagenesis analysis shown in Figure 2B. Compared with the bound state, the T_2 values are overall longer for apo SRSF1 tandem RRMs, which is consistent with the fact that RNA binding slows down overall tumbling of the protein.

Spacer-Length Variability of the Cognate Bipartite Motifs for SRSF1 Tandem RRMs. To investigate the impact of the spacer length between bipartite motifs on binding, we measured binding affinities to RNA sequences with spacers of 0 to 10 nucleotides (0, 2, 6, and 10 U), in addition to swapping CA and GGA (Figure 4 and Table 3). As RRM1 and RRM2 prefer cytidine and purine-rich motifs, respectively, we selected uridine for flanking and spacer regions to avoid interference. We first evaluated the importance of cognate motifs for RRM1 and RRM2. Compared with the reference RNA sequence (2 U), mutation of CA (no-CA) decreases RNA-binding affinity by 10-fold, while mutation of GGA (no-GGA) decreases affinity by 240-fold (Table 3). Therefore, RRM2 plays the dominant role in RNA-binding specificity of SRSF1. It is

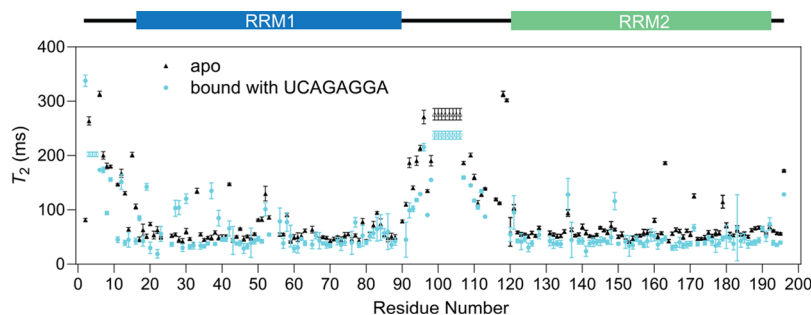


Figure 3. ^{15}N transverse relaxation time constant (T_2) analysis of the SRSF1 tandem RRMs in the apo and bound states. The error was estimated from data fitting. Open symbols indicate overlapping residues, G3–G5 and G99–G106 clusters. Residues in the same cluster are assumed to have the same T_2 relaxation time. The assignment of these two glycine clusters was confirmed by HSQC spectra of constructs with residues 1–13 or 98–106 deleted.

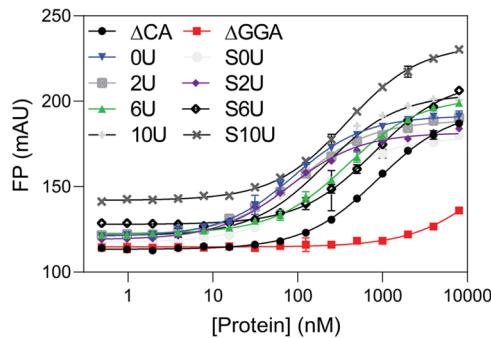


Figure 4. Spacer length variability of SRSF1's bipartite motifs. The standard error of the mean was estimated from three technical replicates.

Table 3. Effects of Motif Swapping and Spacer Length on SRSF1 Binding

RNA	sequence	relative K_D^a
no-CA	uuuuuuuuGGAuuuuu	10 ± 1.1
no-GGA	uuuuuuuCAuuuuuuu	240 ± 65
0 U	uuuuuuCAGGAuuuuu	0.87 ± 0.05
S0 U	uuuuuuGGACAuuuuu	1.72 ± 0.07
2 U	uuuuuCAuuGGAuuuu	1.00 ± 0.05
S2 U	uuuuuGGAuuCAuuuu	0.89 ± 0.05
6 U	uuuCAuuuuuuGGAuu	3.6 ± 0.14
S6 U	uuuGGAuuuuuuCAuu	7.4 ± 0.50
10 U	uCAuuuuuuuuuuGGA	2.1 ± 0.30
S10 U	uGGAuuuuuuuuuuCA	8.0 ± 0.87

^aThe relative K_D values were calculated as the ratios to the K_D of the 2 U RNA (98 ± 4.6 nM). The standard error of the mean was estimated from three technical replicates.

noteworthy that the 10-fold affinity decrease by the CA mutation does not conflict with the result that deletion of RRM1 yields undetectable binding (Figure S1B). These results together suggest that although RRM1 prefers cytidine, it can bind to other nucleotides with a lower affinity.

We then systematically investigated the effect of spacer length on binding (Figure 4). We found that RNA sequences with no spacer (0 U) and a 2-nucleotide spacer (2 U) have similar binding affinities to SRSF1. However, increasing spacer to 6 or 10 nucleotides weakens binding affinity by 2 to 3 fold (6 and 10 U). Swapping bipartite motifs has a negligible effect on binding for RNA with 2-nucleotide spacers (2 and S2 U). For the rest of the cases, moving the CA motif downstream decreases binding affinities by 2 to 4 fold. This finding is different from the previous study, which showed that placing CA 4-nucleotide upstream or 6-nucleotide downstream of GGA has no impact on RNA binding. It is noteworthy that RNA sequences of different lengths were used in that study. Using RNA sequences of the same length, we compared binding affinities for motif-swapped RNA (Table 4). To ensure that the nucleotide type in the spacer is not the origin of the binding affinity difference for motif-swapped RNA, we also compared RNA with UU and GA spacers. For both pairs of RNA, we found that SRSF1 prefers RNA sequences with CA being upstream. The 9-Gly region in the RRM linker is believed to play an important role in endowing domain flexibility. Therefore, we further measured binding affinities for the construct with the 9-Gly region deleted. We found that the trend is similar to what we found for the wild-type protein

Table 4. Effects of Motif Swapping on SRSF1 Binding

RNA	sequence	relative K_D^a
12mer_1	uuuCAgaGGAuu	1.00 ± 0.06
12mer_S1	aGGAuuuuuCAg	2.70 ± 0.05
2	uuuCAuuGGAuu	1.72 ± 0.04
S2	uGGAuuuuuCAu	6.08 ± 0.06

^aThe relative K_D values were calculated as the ratios to the K_D of the 12mer_1 RNA (44 ± 2.5 nM). The standard error of the mean was estimated from three technical replicates.

(Table S1). In summary, our binding assays suggest that SRSF1 prefers RNA sequences with shorter motif spacers and RNA sequences with the upstream RRM1 cognate motif.

Simulation of RNA-Bound Structure of SRSF1 Tandem RRM. We have shown that the linker remains flexible in the RNA-bound state and that the spacer between the bipartite motifs has a variable length. These findings indicate the inter-domain flexibility of SRSF1 in the RNA-bound state. We also compared ¹⁵N-HSQC spectra of SRSF1 bound to motif-swapped RNA sequences (uuuCAuuGGAuu and uGGAuuuuuCAu) and found that these two spectra are almost identical (Figure S3A) with the most significant differences in the linker residues and non-RNA binding residues of the RRM1 (Figure S3B). These results suggest that when bipartite motifs are swapped, RRM1 and RRM2 still recognize the CA and GGA motifs, respectively. CSPs in the linker suggest conformational differences for complexes bound to motif-swapped RNA.

To characterize the dynamic nature of the complex, we collected PRE data for RNA-bound SRSF1. PRE provides distance information up to 25 Å about the labeling site to sites of interest. The sites within 15 Å to the spin label are significantly attenuated, and their resonance peaks disappear from NMR spectra (bleached). The sites 15 to 25 Å away from the spin label demonstrate moderate perturbations, and their distances to the spin label are related to PRE according to the Solomon equation.⁵⁵ The nitroxide paramagnetic compound (MTSL) was labeled at E120C, which is on the edge of RRM2 and not involved in RNA binding. We have also confirmed that MTSL labeling had no impact on RNA binding (data not shown). This labeling strategy will provide the information about the relative position between RRM1 and RRM2. As oligomerization or aggregation can complicate PRE data interpretation, we also performed SAXS and analyzed the radius of gyration using the Guinier plot. As shown by Figure S4, radius of gyration values do not change for the complex at different sample concentrations, suggesting that our samples were free of aggregation.

As shown by Figure 5A,B, paramagnetic perturbations are spread around RRM1 for the SRSF1 RRM tandem in complex with uuuCAuuGGAuu or uGGAuuuuuCAu. This scattered perturbation pattern cannot be explained by a single conformation. This happens when the tandem RRM1s have no fixed relative orientation. In addition, the perturbation patterns are different for SRSF1 bound with the two RNA, suggesting that domain arrangement is different when these bipartite motifs are swapped. Based on Figure S3A, we assume that RRM1s in the tandem bind to CA and GGA in the same manner as they do in isolated domains. With this assumption and PRE restraints, we performed Xplor-NIH simulations and found that the PRE data can only be fitted with a conformational ensemble instead of a single conformation

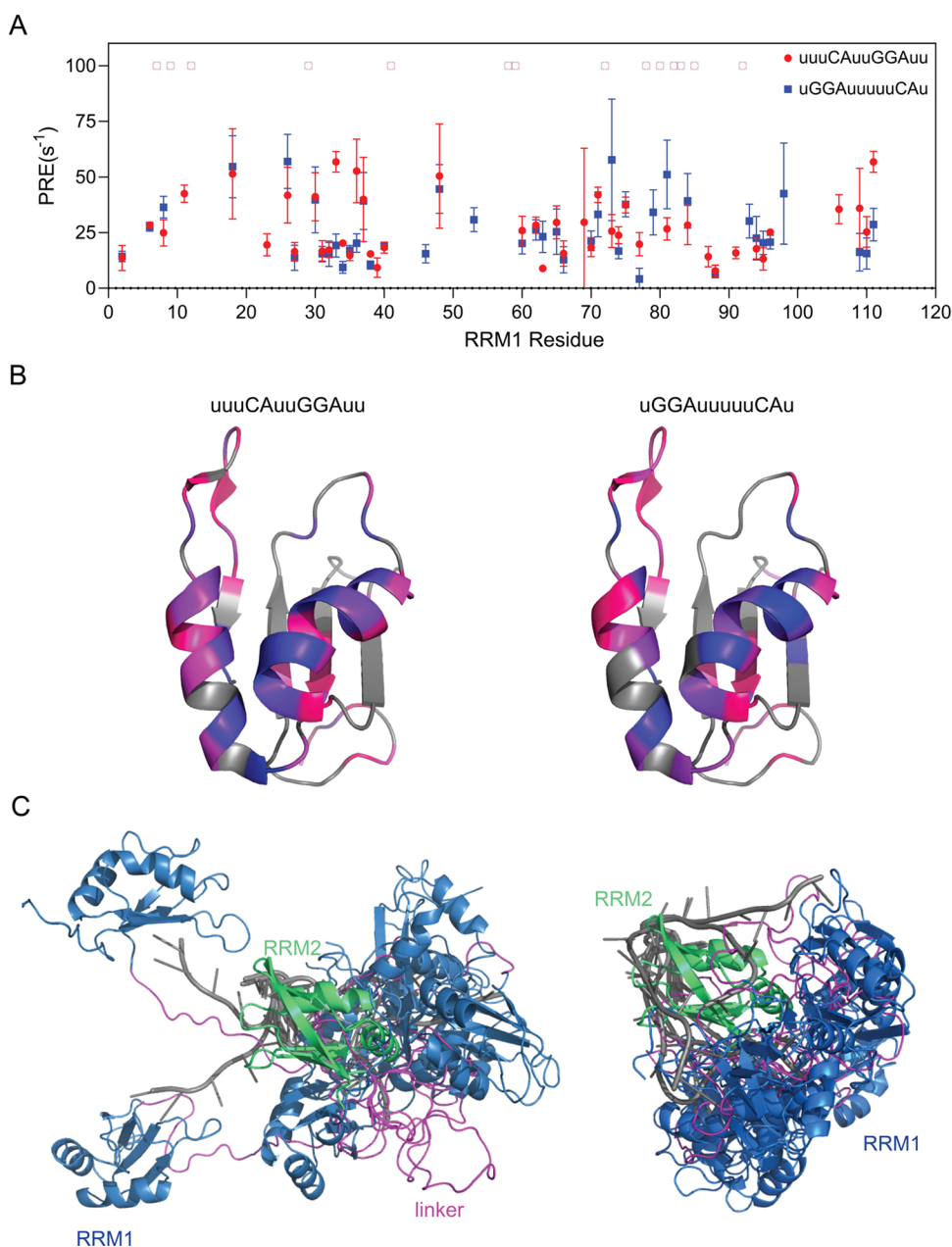


Figure 5. Ensemble simulation of RNA-bound SRSF1 restrained by PRE data. (A) RRM1 backbone amide ^1H -PRE values in the RNA-bound SRSF1 tandem RRM1s bound to motif-swapped RNA. The MTSL was labeled on RRM2 residue E120C. Open symbols indicate residues whose PRE values are too large to be accurately determined. (B) Plot of PRE values in panel A onto RRM1 structure (PDB ID: 6HPJ). Gray indicates residues whose PRE values are unavailable due to overlapping or missing resonances. (C) Xplor-NIH ensemble simulation of RNA-bound SRSF1, in which RRM1, RRM2, the linker, and RNA are shown in blue, green, magenta, and gray, respectively. An ensemble of eight conformers was simulated to fit the PRE data shown in panel A for each RNA. All conformers were aligned based on RRM2 (residues 121–196).

(Figure 5B). When bound to uGGAuuuuuCAu, the complex adopts a more confined ensemble. In addition, forming turns in RNA seems required when the CA motif is placed downstream.

In addition to the canonical RNA sequences with spacers of 2 nucleotides, we found that SRSF1 can also bind to RNA with spacers of 0 to 10 nucleotides. To demonstrate how this sequence diversity is achieved, we simulated the complex structure under the assumption that RRM1 and RRM2 bind to CA and GGA, respectively (Figure 6). Our simulations show that SRSF1 tandem RRM1s can form plausible complexes with all these RNA. For clarity, only the lowest energy conformations are shown. The binding diversity is achieved

by the flexibility in both the RRM1/RRM2 linker and RNA backbones. For RNA without spacers, bases of CA and GGA motifs point opposite directions for SRSF1 binding. For RNA with 6-nucleotide or 10-nucleotide spacers, the RRM1/RRM2 linker tends to assume a more stretched conformation when the RRM1-binding motif CA is downstream. Alternatively, U turns in RNA backbones are required to bind with SRSF1 (Figure 6B). In our simulation, we found that the linker takes relaxed conformation for most cases. This is consistent with our finding that deletion of 9 gly in the linker has a negligible impact on RNA binding (Table S1). This suggests that the linker provides ample flexibility even without the 9-Gly region.

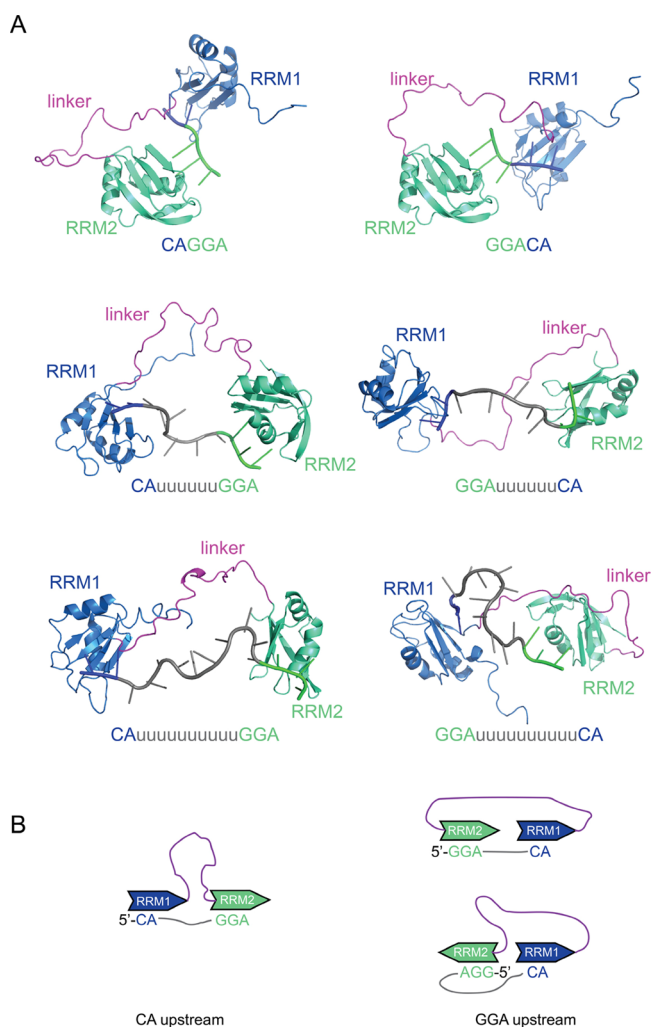


Figure 6. Xplor-NIH simulation of SRSF1 bound to RNA with different spacer lengths. (A) Lowest energy conformation for each complex. RRM1, RRM2, and the linker are shown in blue, green, and magenta, respectively. Sequences of bound RNA are shown below. RRM1 and RRM2 are restrained to interact with CA and GGA motifs, respectively. (B) Schematic illustration of RRM and RNA motif arrangement. The color scheme is the same as panel A. When the CA motif is swapped to downstream of GGA, either a more stretched linker conformation or an RNA U turn is needed for SRSF1 binding.

DISCUSSION

Extensive efforts have been made to determine the RNA-binding specificity of SRSF1 during the past decades. A recent study has found that RRM1 recognizes C-containing dinucleotide motifs, which can be either 4-nucleotide upstream or 6-nucleotide downstream of RRM2-binding motifs.⁴⁴ The same study has also concluded that swapping the two motifs does not affect binding affinity. The RNA with CA at the 5' end in that study was 3 nucleotides shorter than the RNA with CA at the 3' end.⁴⁴ Therefore, the impact of RNA length on binding cannot be ruled out. Using the RNA of the same length, we found that SRSF1 tandem RRMs showed a reproducible two- to three-fold preference for RNA sequences with the CA motif at the 5' end, except for RNA with 2-nucleotide spacers. When the RRM1-cognate motif is 3' to the RRM2-cognate motif, SRSF1 binding requires either turns in RNA or stretched conformation in the linker. These conformational entropy losses may explain the moderate but

reproducible binding affinity difference between the bipartite motif-swapped RNA sequences.

The 30-amino acid linker between the tandem RRMs is responsible for the flexibility of RNA binding. As shown by our T_2 analysis, the linker remains flexible in the RNA-bound state. A previous study has proposed that the linker is involved in RNA binding, as mutation of some linker residues dramatically reduces RNA binding.⁴³ However, we found that the major role of the linker is to tether the two RRM domains while still allowing flexibility. We showed that mutation of basic linker residues only slightly reduces RNA binding. Our binding results are consistent with our NMR T_2 analysis. These five Arg residues are scattered in the linker. If all of them interact with RNA at the same moment, the flexibility of the linker should be reduced dramatically. However, our T_2 analysis suggests that the linker is very flexible. Therefore, only a small portion of them transiently interact with RNA.

The flexibility of the RRM1/RRM2 loop allows considerable length variation in the spacer between the bipartite motifs of the SRSF1 tandem RRMs. We discovered that when the RRM1-cognate motif is upstream, increasing the spacer length up to 10 nucleotides only attenuates affinity by 2 to 4 fold. It is imaginable that the spacer can be even longer with binding affinity further decreased. In contrast, when the RRM1-cognate motif is downstream, a spacer of 10 nucleotides attenuates the affinity by eightfold. This binding affinity is similar to the RNA without C-containing motifs for RRM1 (Table 3). This suggests that RRM1's preference for C-containing motifs is negligible when they are more than 10 nucleotides downstream of RRM2-cognate motifs. The binding-affinity decrease along with spacer length can be explained by the penalty of loop-closure entropy, which arises from the restraint of motions for the two termini of a loop.⁵⁶ Binding reduces the entropy of the RNA and the RRM1/RRM2 loop, and this penalty increases gradually along with the spacer length.⁵⁶ Therefore, it is predictable that increasing the spacer length will gradually reduce RNA-binding affinity, instead of abolishing binding abruptly at a certain spacer length. When the RRM1-cognate motif is placed downstream, either the linker needs to be more extended or the RNA needs to form a U turn. Either situation decreases the conformational freedom to a greater extent, which explains SRSF1's preference for C-upstream RNA. This preference for C-upstream RNA is not changed when the 9-Gly region of the RRM linker is deleted, which suggests that the linker has provided ample inter-domain flexibility.

Considering the fact that neighboring ESEs compete with each other for SRSF1 binding, determining the relative RNA-binding affinity of SRSF1 to various splicing enhancers is critical to understanding its roles in alternative splicing. Existing high-throughput methods such as SELEX or HITS-CLIP identify consensus sequences for RNA-binding proteins using sequence alignment, which is usually based on assumption of a fixed motif length. However, SRSF1's bipartite motifs are swappable and have a variable spacer length from 0 to 10 nucleotides. Previous studies have identified divergent consensus sequences for SRSF1, such as RGAACAAC, AGGACAGAGC, AGGACGAAGC, and SRSASGA, where S represents G or C and R represents purine.^{36–40} Among these sequences, C appears both upstream and downstream of purine-rich motifs. Moreover, the distance of motif "C" to purine-rich motifs also varies, which is consistent with our findings. In addition, we found that RRM1 prefers cytidine over uridine by only 10-fold. SRSF1 has considerable affinities

to RNA sequences without cytidine. Our result explains the recent in vivo finding that clusters of GGA sites are sufficient to interact with SRSF1.⁴¹ In summary, the inter-domain flexibility explains the discrepancies in SRSF1-binding RNA consensus sequences identified by different studies.

The inter-domain flexibility may also be present in other SR proteins with tandem RRM, such as SRSF4, SRSF5, SRSF6, and SRSF9, whose tandem-RRM linkers range from 22 to 38 amino acids. A more advanced method called RNA Bind-n-Seq (RBNS) has been developed and successfully applied to several RNA-binding proteins in profiling the landscape of RNA-binding specificity.^{57,58} However, this method assumes that the length of cognate RNA motifs is a constant, which is not applicable for RNA binding with SRSF1 and other SR proteins with tandem RRM. Therefore, new methods with improved procedures and/or algorithms are needed to determine the RNA-binding specificity of SR proteins with tandem RRM.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.2c00565>.

Impact of 9-Gly deletion in the RRM linker, roles of RRM1 and the RRM1/RRM2 linker in RNA binding, *T*₂ analysis results, and SAXS data (PDF)

Accession Codes

NMR assignment data have been submitted to BMRB (ID: 51256).

■ AUTHOR INFORMATION

Corresponding Author

Jun Zhang – Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294-1240, United States;
 ● orcid.org/0000-0002-5842-7424; Phone: 1-205-934-2139; Email: zhanguab@uab.edu; Fax: 1-205-934-2543

Authors

Naiduwadura Ivon Upekala De Silva – Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294-1240, United States
Talia Fargason – Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294-1240, United States
Zihan Zhang – Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294-1240, United States
Ting Wang – Department of Chemistry, College of Arts and Sciences, University of Alabama at Birmingham, Birmingham, Alabama 35294-1240, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.biochem.2c00565>

Author Contributions

J.Z. and N.D. designed the research, analyzed the data, and wrote the manuscript. N.D., T.F., Z.Z., and T.W. performed research.

Funding

This work was supported by the National Science Foundation (MCB 2024964 to Zhang J.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We want to thank the manager of UAB Central Alabama High-Field NMR Facility Dr. Ron Shin, the director of the NMR facility Dr. William Placzek, and the director of UAB Structural Biology Core Facility Dr. Champion Deivanayagam and Dr. Zhengrong Yang for technical support. We also want to acknowledge Dr. Charles D. Schwieters at NIH for his help with Xplor-NIH. We thank the staff of SIBYLS for assistance with SAXS data collection.

■ REFERENCES

- (1) Manley, J. L.; Krainer, A. R. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* **2010**, *24*, 1073–1074.
- (2) Zahler, A. M.; Lane, W. S.; Stolk, J. A.; Roth, M. B. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **1992**, *6*, 837–847.
- (3) Velazquez-Dones, A.; Hagopian, J. C.; Ma, C. T.; Zhong, X. Y.; Zhou, H.; Ghosh, G.; Fu, X. D.; Adams, J. A. Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *J. Biol. Chem.* **2005**, *280*, 41761–41768.
- (4) Hagopian, J. C.; Ma, C. T.; Meade, B. R.; Albuquerque, C. P.; Ngo, J. C.; Ghosh, G.; Jennings, P. A.; Fu, X. D.; Adams, J. A. Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1. *J. Mol. Biol.* **2008**, *382*, 894–909.
- (5) Aubol, B. E.; Plocinik, R. M.; Hagopian, J. C.; Ma, C. T.; McGlone, M. L.; Bandyopadhyay, R.; Fu, X. D.; Adams, J. A. Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *J. Mol. Biol.* **2013**, *425*, 2894–2909.
- (6) Lai, M. C.; Tarn, W. Y. Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J. Biol. Chem.* **2004**, *279*, 31745–31749.
- (7) Huang, Y.; Gattoni, R.; Stevenin, J.; Steitz, J. A. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell* **2003**, *11*, 837–843.
- (8) Stojdl, D. F.; Bell, J. C. SR protein kinases: the splice of life. *Biochem. Cell Biol.* **1999**, *77*, 293–298.
- (9) Mermoud, J. E.; Cohen, P. T.; Lamond, A. I. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* **1994**, *13*, 5679–5688.
- (10) Cho, S.; Hoang, A.; Sinha, R.; Zhong, X. Y.; Fu, X. D.; Krainer, A. R.; Ghosh, G. 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. U. S. A.* **2011**, *108*, 8233–8238.
- (11) Blencowe, B. J.; Bowman, J. A.; McCracken, S.; Rosonina, E. SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.* **1999**, *77*, 277–291.
- (12) Boucher, L.; Ouzounis, C. A.; Enright, A. J.; Blencowe, B. J. A genome-wide survey of RS domain proteins. *RNA* **2001**, *7*, 1693–1701.
- (13) Kellenberger, E.; Stier, G.; Sattler, M. Induced folding of the U2AF35 RRM upon binding to U2AF65. *FEBS Lett.* **2002**, *528*, 171–176.
- (14) Zhou, Z.; Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **2013**, *122*, 191–207.
- (15) Li, X.; Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **2005**, *122*, 365–378.
- (16) Das, R.; Yu, J.; Zhang, Z.; Gygi, M. P.; Krainer, A. R.; Gygi, S. P.; Reed, R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **2007**, *26*, 867–881.

- (17) Lin, S.; Coutinho-Mansfield, G.; Wang, D.; Pandit, S.; Fu, X. D. The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* **2008**, *15*, 819–826.
- (18) Huang, Y.; Steitz, J. A. SRprises along a messenger's journey. *Mol. Cell* **2005**, *17*, 613–615.
- (19) Sanford, J. R.; Gray, N. K.; Beckmann, K.; Caceres, J. F. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* **2004**, *18*, 755–768.
- (20) Michlewski, G.; Sanford, J. R.; Caceres, J. F. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol. Cell* **2008**, *30*, 179–189.
- (21) Zhang, Z.; Krainer, A. R. Involvement of SR proteins in mRNA surveillance. *Mol. Cell* **2004**, *16*, 597–607.
- (22) Kanehiro, Y.; Todo, K.; Negishi, M.; Fukuoka, J.; Gan, W.; Hikasa, T.; Kaga, Y.; Takemoto, M.; Magari, M.; Li, X.; Manley, J. L.; Ohmori, H.; Kanayama, N. Activation-induced cytidine deaminase (AID)-dependent somatic hypermutation requires a splice isoform of the serine/arginine-rich (SR) protein SRSF1. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1216–1221.
- (23) Hu, Z. Y.; Wang, X. Y.; Guo, W. B.; Xie, L. Y.; Huang, Y. Q.; Liu, Y. P.; Xiao, L. W.; Li, S. N.; Zhu, H. F.; Li, Z. G.; Kan, H. Long non-coding RNA MALAT1 increases AKAP-9 expression by promoting SRPK1-catalyzed SRSF1 phosphorylation in colorectal cancer cells. *Oncotarget* **2016**, *7*, 11733–11743.
- (24) Malakar, P.; Shilo, A.; Mogilevsky, A.; Stein, I.; Pikarsky, E.; Nevo, Y.; Benyamini, H.; Elgavish, S.; Zong, X.; Prasanth, K. V.; Karni, R. Long Noncoding RNA MALAT1 Promotes Hepatocellular Carcinoma Development by SRSF1 Upregulation and mTOR Activation. *Cancer Res.* **2017**, *77*, 1155–1167.
- (25) Maslon, M. M.; Heras, S. R.; Bellora, N.; Eyra, E.; Caceres, J. F. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife* **2014**, *3*, No. e02028.
- (26) Karni, R.; de Stanchina, E.; Lowe, S. W.; Sinha, R.; Mu, D.; Krainer, A. R. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* **2007**, *14*, 185–193.
- (27) He, X.; Ee, P. L.; Coon, J. S.; Beck, W. T. Alternative splicing of the multidrug resistance protein 1/ATP binding cassette transporter subfamily gene in ovarian cancer creates functional splice variants and is associated with increased expression of the splicing factors PTB and SRp20. *Clin. Cancer Res.* **2004**, *10*, 4652–4660.
- (28) Fischer, D. C.; Noack, K.; Runnebaum, I. B.; Watermann, D. O.; Kieback, D. G.; Stamm, S.; Stickeler, E. Expression of splicing factors in human ovarian cancer. *Oncol. Rep.* **2004**, *11*, 1085–1090.
- (29) Kozlovski, I.; Siegfried, Z.; Amar-Schwartz, A.; Karni, R. The role of RNA alternative splicing in regulating cancer metabolism. *Hum. Genet.* **2017**, *136*, 1113–1127.
- (30) Anczukow, O.; Rosenberg, A. Z.; Akerman, M.; Das, S.; Zhan, L.; Karni, R.; Muthuswamy, S. K.; Krainer, A. R. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* **2012**, *19*, 220–228.
- (31) Forbes, S. A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C. G.; Ward, S.; Dawson, E.; Ponting, L.; Stefancsik, R.; Harsha, B.; Kok, C. Y.; Jia, M.; Jubb, H.; Sondka, Z.; Thompson, S.; De, T.; Campbell, P. J. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2017**, *45*, D777–D783.
- (32) Blaustein, M.; Pelisch, F.; Tanos, T.; Munoz, M. J.; Wengier, D.; Quadrana, L.; Sanford, J. R.; Muschietti, J. P.; Kornblihtt, A. R.; Caceres, J. F.; Coso, O. A.; Srebrow, A. Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat. Struct. Mol. Biol.* **2005**, *12*, 1037–1044.
- (33) Zhou, Z.; Qiu, J.; Liu, W.; Zhou, Y.; Plocinik, R. M.; Li, H.; Hu, Q.; Ghosh, G.; Adams, J. A.; Rosenfeld, M. G.; Fu, X. D. The Akt-SRPK-SR axis constitutes a major pathway in transducing EGF signaling to regulate alternative splicing in the nucleus. *Mol. Cell* **2012**, *47*, 422–433.
- (34) Jang, S. W.; Liu, X.; Fu, H.; Rees, H.; Yepes, M.; Levey, A.; Ye, K. Interaction of Akt-phosphorylated SRPK2 with 14-3-3 mediates cell cycle and cell death in neurons. *J. Biol. Chem.* **2009**, *284*, 24512–24525.
- (35) Jiang, K.; Patel, N. A.; Watson, J. E.; Apostolatos, H.; Kleiman, E.; Hanson, O.; Hagiwara, M.; Cooper, D. R. Akt2 regulation of Cdc2-like kinases (Clk/Sty), serine/arginine-rich (SR) protein phosphorylation, and insulin-induced alternative splicing of PKC β -tall messenger ribonucleic acid. *Endocrinology* **2009**, *150*, 2087–2097.
- (36) Tacke, R.; Manley, J. L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **1995**, *14*, 3540–3551.
- (37) Schaal, T. D.; Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **1999**, *19*, 1705–1719.
- (38) Liu, H. X.; Zhang, M.; Krainer, A. R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **1998**, *12*, 1998–2012.
- (39) Sanford, J. R.; Wang, X.; Mort, M.; Vanduy, N.; Cooper, D. N.; Mooney, S. D.; Edenberg, H. J.; Liu, Y. Splicing factor SRSF1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **2009**, *19*, 381–394.
- (40) Sanford, J. R.; Coutinho, P.; Hackett, J. A.; Wang, X.; Ranahan, W.; Caceres, J. F. Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One* **2008**, *3*, No. e3369.
- (41) Feng, H.; Bao, S.; Rahman, M. A.; Weyn-Vanhenhenryck, S. M.; Khan, A.; Wong, J.; Shah, A.; Flynn, E. D.; Krainer, A. R.; Zhang, C. Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Mol. Cell* **2019**, *74*, No. e1186.
- (42) Clery, A.; Sinha, R.; Anczukow, O.; Corrionero, A.; Moursy, A.; Daubner, G. M.; Valcarcel, J.; Krainer, A. R.; Allain, F. H. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2802–E2811.
- (43) Cho, S.; Hoang, A.; Chakrabarti, S.; Huynh, N.; Huang, D. B.; Ghosh, G. The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRM. *Nucleic Acids Res.* **2011**, *39*, 9413–9421.
- (44) Clery, A.; Krepl, M.; Nguyen, C. K. X.; Moursy, A.; Jorjani, H.; Katsantoni, M.; Okoniewski, M.; Mittal, N.; Zavolan, M.; Sponer, J.; Allain, F. H. Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. *Nat. Commun.* **2021**, *12*, 428.
- (45) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (46) Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **2004**, *278*, 313–352.
- (47) Iwahara, J.; Tang, C.; Marius Clore, G. Practical aspects of (1) H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **2007**, *184*, 185–195.
- (48) Petoukhov, M. V.; Konarev, P. V.; Kikhney, A. G.; Svergun, D. I. ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **2007**, *40*, s223–s228.
- (49) Tang, C.; Ghirlardo, R.; Clore, G. M. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **2008**, *130*, 4048–4056.
- (50) Jarmoskaite, I.; AlSadhan, I.; Vaidyanathan, P. P.; Herschlag, D. How to measure and evaluate binding affinities. *eLife* **2020**, *9*, No. e57264.
- (51) Hulme, E. C.; Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **2010**, *161*, 1219–1237.
- (52) Anczukow, O.; Akerman, M.; Clery, A.; Wu, J.; Shen, C.; Shirole, N. H.; Raimar, A.; Sun, S.; Jensen, M. A.; Hua, Y.; Allain, F. H.; Krainer, A. R. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol. Cell* **2015**, *60*, 105–117.

(53) Maris, C.; Dominguez, C.; Allain, F. H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118–2131.

(54) Zhang, J.; Petit, C. M.; King, D. S.; Lee, A. L. Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). *J. Biol. Chem.* **2011**, *286*, 41776–41785.

(55) Clore, G. M.; Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **2009**, *109*, 4108–4139.

(56) Nagi, A. D.; Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Folding Des.* **1997**, *2*, 67–75.

(57) Lambert, N.; Robertson, A.; Jangi, M.; McGeary, S.; Sharp, P. A.; Burge, C. B. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* **2014**, *54*, 887–900.

(58) Zaharias, S.; Zhang, Z.; Davis, K.; Fargason, T.; Cashman, D.; Yu, T.; Zhang, J. Intrinsically disordered electronegative clusters improve stability and binding specificity of RNA-binding proteins. *J. Biol. Chem.* **2021**, *297*, No. 100945.

Recommended by ACS

Interaction of hnRNPB1 with Helix-12 of hHOTAIR Reveals the Distinctive Mode of RNA Recognition That Enables the Structural Rearrangement by LCD

Ajit Kumar, Niyati Jain, *et al.*

JUNE 12, 2023
BIOCHEMISTRY

READ 

Site-Selective Tyrosine Phosphorylation in the Activation of the p50 Subunit of NF-κB for DNA Binding and Transcription

Shengxi Chen, Sidney M. Hecht, *et al.*

DECEMBER 19, 2022
ACS CHEMICAL BIOLOGY

READ 

Comparative Biochemical Studies of Disease-Associated Human Dicer Mutations on Processing of a Pre-microRNA and snoRNA

Rachel M. Torrez, Amanda L. Garner, *et al.*

MAY 02, 2023
BIOCHEMISTRY

READ 

B-box1 Domain of MID1 Interacts with the Ube2D1 E2 Enzyme Differently Than RING E3 Ligases

Anupreet Kaur, Michael A. Massiah, *et al.*

FEBRUARY 23, 2023
BIOCHEMISTRY

READ 

Get More Suggestions >