Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The structural basis of SRSF6-mediated splicing decisions governed by tandem-domain RNA recognition and RS domain tuning

Andreas Schlundt

andreas.schlundt@uni-greifswald.de

University of Greifswald https://orcid.org/0000-0003-2254-7560

Julian von Ehr

Goethe University https://orcid.org/0000-0003-4535-661X

Sunesh Sethumadhavan

Goethe University

Friederike Schubert

Max Planck Institute of Biophysics https://orcid.org/0009-0006-8942-2120

Sophie Korn

University of Greifswald

Natalie Meixner

Goethe University https://orcid.org/0009-0000-1043-4118

Lasse Oberstrass

Marburg University https://orcid.org/0009-0005-5112-1456

Bernd Simon

University of Connecticut Health Center https://orcid.org/0000-0003-0164-5516

Julia Weigand

Marburg University

Gerhard Hummer

Max Planck Institute of Biophysics https://orcid.org/0000-0001-7768-746X

Michaela Müller-McNicoll

Goethe-Universität Frankfurt am Main https://orcid.org/0000-0002-7174-8310

Article

Keywords: SRSF6, NMR, alternative splicing, RRM, RNA binding preferences, iCLIP

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9953096/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Why and how cells exploit redundant versus complementary specificities of the 12 splice-regulatory SRSF(1–12) proteins remains enigmatic, but points at subtle details in their RNA binding modules. The resulting need in high-resolution information is especially evident for SRSF6 with no existing structures explaining its pre-mRNA selectivity. Using iCLIP, global splicing analyzes, RNA-Bind-and-Seq, integrated structural biology, molecular simulations, and systematic interaction studies including purified full-length eukaryotic protein, we provide the first atom-resolved basis of RNA-recognition by SRSF6 through RRM-, linker-, and RS-domains. SRSF6 uses its two differently typed RRMs in distinct interactions with constitutive (GAA-rich) or alternative exons (GGA) modulated by the RS domain, explaining previous inconsistencies within SRSF6-targeted motifs. Our findings suggest a mode of redundant and mutually exclusive splicing decisions between the closely related SRSF1 and SRSF6 as a blueprint for integrated splicing regulation by all 12 SRSF proteins.

Introduction

In humans, over 95% of multi-exon genes undergo alternative splicing generating multiple transcript isoforms from the same gene, which is essential for cell identity, development, and stress adaptation[1]. The selection of splice sites is tightly regulated by a network of RNA binding proteins (RBPs), including the serine/arginine-rich splicing factors (SRSFs). SRSFs promote constitutive and alternative splicing by binding to specific sequences in pre-mRNAs and recruiting components of the splicing machinery[2]. Misregulation of SRSF expression or activity, or mutations within their binding motifs, disrupts splicing and contributes to a wide range of diseases, including neurodegeneration and cancer[3,4].

The SRSF protein family consists of 12 canonical members (SRSF1–SRSF12), which share a modular architecture comprising an N-terminal RNA recognition motif (RRM), a variable glycine/arginine-rich linker, and a C-terminal RS domain composed of arginine/serine repeats[5]. Five SRSFs contain a second RRM lacking canonical RNA binding motifs (RNP1/RNP2), known as pseudo RRM (ΨRRM). Besides a general preference for purine-rich triplets, little is known about the drivers of SRSF RNA specificity. Overlapping binding preferences of SRSF proteins are a consequence of their redundant and in parts competitive roles in constitutive splicing. However, SRSF proteins fulfill additional, highly specific functions, such as promoting the inclusion of alternative exons[6]. This suggests a more sophisticated mechanism of specific RNA recognition that allows SRSF proteins to play distinct roles in alternative splicing, including the subordinate recognition of less abundant or context-dependent motifs.

RNA binding of SRSF proteins is determined by the structured RRM domains. Their domain boundaries and tertiary structures play major roles for RNA selectivity. In SRSFs with tandem RRMs, domain arrangement, linker lengths and sequence composition are decisive[7]. An additional layer of regulation is provided through the intrinsically disordered RS domains, which are thought to modulate RNA binding affinity through charge interactions[8,9], but are subject to reversible phosphorylation and differ significantly in length.

A limited number of SRSF RRM structures is available, and from those only five in complex with target RNA motifs (from SRSF1, 2, 3). SRSF1, the best-studied family member, comprises an RRM and a ΨRRM and binds GGA triplets *in vivo* (Fig. 1A) and *in vitro*[10–12]. A high-resolution NMR structure of the ΨRRM in complex with a GGA-motif containing RNA revealed the molecular basis for this selective interaction[10]. In contrast to canonical RRMs, the ΨRRM of SRSF1 engages RNA via a conserved motif, α-helix 1. However, no high-resolution structure exists to date for a tandem RRM bound to a bipartite RNA motif revealing the unambiguous atomic basis for specific transcript regulation. Also, no systematic work has been carried out to address the precise role of domain contributions and relative positioning, the linker, RNA motif composition and nature of RS domain in this context.

SRSF6, a yet understudied member of the SRSF family plays crucial roles in alternative splicing programs during cell differentiation and stress responses, e.g.,hypoxia or DNA damage[13,14]. A co-regulation of specific mRNA targets together with SRSF1 has recently been reported[15]. SRSF1 and SRSF6 share the overall RRM-ΨRRM domain architecture yet differ in the linker and the length of their RS domains. In contrast to SRSF1, *in vitro* and *in cellulo* RNA binding preferences reported for SRSF6 diverge significantly. Earlier *in vitro* derived motifs[16,17], resemble the sequence 'UGCGGAC'[15] that was recently reported to be bound and regulated by SRSF6, basically consisting of C nucleotides flanking a GGA triplet. In contrast, iCLIP studies – including our own – suggested that on a transcriptome-wide scale SRSF6 predominantly occupies GAA triplets[12,18,19] (Fig. 1A). Thus, it remains unclear how SRSF6 recognizes both GGA and GAA triplets, whether different binding modes exist that affect splicing outcomes, and how SRSF1 and SRSF6 compete for target sites. To date, the lack of SRSF6 structural information and the challenges faced during purification of full-length (FL) SRSF proteins have made more detailed, comparative studies impossible.

In this study, we provide the first molecular basis for the discrepant RNA recognition by SRSF6 by presenting high-resolution structures of the individual RRM domains with and without their preferred single-stranded RNA motifs using NMR and crystallography. We find that *in vitro* SRSF6 ΨRRM2 clearly prefers GGA motifs, but GAA motifs are structurally tolerated. Moreover, we show that RRM1 binds C(A)-rich motifs and contributes to binding selectivity and affinity in minimal composite motifs, important for the splicing of alternative exons. Using a combination of NMR, crystallography, molecular modeling, molecular dynamics (MD) simulation, SAXS, and *in vitro* binding assays, we further show cooperative RNA binding of the tandem RRMs. Our data reveal that the interdomain linker slightly contributes to the recognition of combined RNA motifs influenced by motif order and spacing. The first successful purification of human FL SRSF6 proteins from HEK293F cells ensures endogenous phosphorylation levels and allowed us to investigate the discrepancy between *in vitro* and *in vivo* binding selectivity. We uncover that the RS domain plays an important role in modulating the overall RNA affinity of SRSF6, acting as a filter for selectivity not found in SRSF1. Altogether, our data provide a hitherto unknown mode of distinct RNA recognition by SRSF6 and its interference with SRSF1 explaining prior discrepancies in reported motifs. This work underlines the importance of systematic structure-based evaluations of

statistically treated target motifs, but also serves as a blueprint for a general mechanism of SRSF proteins' redundant and competitive pre-mRNA processing.

Material and Methods

SRSF6 cloning, expression, and protein purification from E. coli

All constructs were based on the human SRSF6 amino acid sequence (UniProt[20] ID: Q13247). The cloning strategy has been described by us in detail[21]. Briefly, *E. coli* codon optimized sequences for the SRSF6 tandem RRM construct (1-188) and the individual RRM domains with various boundaries (RRM1: 1–105; ΨRRM2: 72–188, 86–188, and 106–188) were sub-cloned into the bacterial expression plasmid pPK591 (kind gift from Peter Kötter, GU Frankfurt), containing an N-terminal His_6- and SUMO-tag. The RRM1 construct 1–72 was cloned into pET24d-derived expression vector pET-Trx1a (kind gift from Gunter Stier, EMBL/BZH Heidelberg), encoding an N-terminal His_6-tag, a thioredoxin tag (Trx) and a TEV cleavage site. Oligonucleotides used for cloning are listed in Supplementary Table 1. SUMO-tagged proteins contained three non-native N-terminal residues (GAG) after Ulp1 cleavage, the Trx-tagged protein contained four artificial residues (GAMA) after TEV-cleavage. All mutations in SRSF6 were introduced by site directed mutagenesis (Supplementary Tables 2–5). For RNA Bind-n-Seq (RBNS) applications, an N-terminal 2xStrep-tag was added to selected SRSF6 constructs by Gibson assembly[22] (Supplementary Table 6).

Recombinant proteins were purified either under native or denaturing conditions, as described previously[21]. After tag removal, proteins were subjected to size exclusion chromatography with either SEC-buffer 1 (20 mM sodium phosphate, 150 mM NaCl, 2 mM dithiothreitol (DTT), 0.02% NaN_3, pH 6.5) for constructs 1–72 and 106–188, SEC-buffer 2 (20 mM sodium phosphate, 1 M NaCl, 2 mM DTT, 380 mM l-arginine/glutamate (l-R/E), 0.02% NaN_3, pH 7.0) for constructs 1–188, 72–188 and 1–105 or SEC-buffer 3 (20 mM sodium phosphate, 1 M NaCl, 2 mM DTT, 50 mM l-R/E, 0.02% NaN_3, pH 7.0) for construct 86–188. Final samples were buffer exchanged to their respective SEC-buffer but with either 150 mM NaCl or 0 mM NaCl for subsequent experiments. All constructs except 1–105 could be purified under denaturing conditions.

Preparation and labelling of RNA Ligands

Unlabeled and 5'-labeled (fluorescein – Fl or Cyanine 5 – Cy5) RNA oligonucleotides (≤ 12 nucleotides (nt)) were obtained from Sigma-Aldrich or Dharmacon Horizon in deprotected and desalted form (Supplementary Table 7). RNAs were dissolved in nuclease-free water to a final concentration of 3 mM (unlabeled) or 100 µM (labeled). Longer RNAs (> 20 nt) were generated by *in vitro* transcription from synthetic DNA templates (Supplementary Table 8) as described previously[23]. Transcribed RNAs (Supplementary Table 9) were purified by denaturing PAGE and ethanol precipitation. For 5'-labeling of *in*

vitro transcribed RNAs we followed previously established protocols (5'-Fl[24] and 5'-Cy5[25]). Final RNA stocks were resuspended in nuclease-free water and stored at -80°C until use.

In vitro transcription of the RBNS input pool

The RNA input pool was produced by *in vitro* transcription as described before[23]. In brief, as template an oligonucleotide containing a T7 promoter was annealed to an equimolar quantity of an oligonucleotide containing a 20mer random sequence flanked by partial Illumina adapters. Then, 500 fmol of the resulting duplex was transcribed using homemade T7 RNA polymerase. The transcribed RNA pool was purified using denaturing polyacrylamide gel electrophoresis (PAGE) and folded via snap cooling before use in the RBNS assays.

RBNS assays of Strep-tagged SRSF6 protein constructs

RBNS assays were performed using the previously described protocol[23] which is based on Lambert et al[26], with all Strep-tagged SRSF6 constructs expressed in *E. coli*. Briefly, the pulldowns were performed in 25 mM Tris-HCl (pH 7.5), 150 mM KCl, 3 mM $MgCl_2$, 0.01% Tween, 500 µg/ml BSA, 1 mM DTT and 0.16 U/µL Ribonuclease Inhibitor (moloX, Berlin), with 1 µM RNA input pool and varying concentrations of the protein constructs. The specific SRSF6 constructs and the protein concentrations used are listed in Supplementary Table 10. The RNA-protein complexes were captured with MagStrep “type 3” XT beads (IBA Lifesciences), and the bound RNA was extracted. The recovered RNA was reverse transcribed and transformed into Illumina libraries with unique sample indices (Supplementary Table 11). The libraries were then subjected to deep sequencing (GENEWIZ, Leipzig).

Analysis of RBNS data

NGS data were processed with the RBNS pipeline described in [27] available at https://bitbucket.org/pfreese/rbns_pipeline/overview. Sequence context logos were generated with a custom Python script. The script extracts every sequencing read that contains a given short motif and aligns the reads by shifting the motif to the same position while adding gaps, represented by “-” at both ends of the sequences. Next, the nucleotide count is calculated for each position, and a position weight matrix is generated and plotted using Logomaker[28]. Positions with > 75% missing base information in the alignment were excluded from the plot. As background, the bases around the same motif in the input pool were used. The input motifs were either enriched k-mers from the same construct or from constructs with other RRM domains.

NMR spectroscopy

NMR experiments were performed at the Frankfurt Center for Biomolecular Magnetic Resonance (BMRZ) using Bruker Avance III/Avance Neo spectrometers of 600, 700, 800, 900 and 950 MHz proton Larmor

frequency, equipped with cryogenic probes and using Z-axis pulsed field gradients. Measurements were performed at 298 K (unless stated otherwise) in different buffers (indicated in the figure legends). Topspin versions 3 and 4 were used for data acquisition and processing. Graphical plots of spectra were created using the program NMRFAM-Sparky v1.470[29]. Analysis and assignments of NMR spectra as well as quantification/plotting of CSPs from titrations were done in the CCPNMR Analysis 2.5 software[30]. Combined ^{1}H-^{15}N-CSPs were calculated in ppm according to Eq. 1.

$$CSP = \sqrt{(0.15 \times \ \delta N)^2 + (\delta H)^2}$$

1

For the assignment of the backbone resonances of the constructs 1–72 (3.6x poly-C RNA), 86–188 (2x CGGACU RNA) a set of HNCA, HNCO, HNCACB/CBCANH, HN(CA)CO, CBCA(CO)NH spectra was used. Hα and Hβ atoms for construct 1–72 were assigned by a combination of HBHANH and HBHA(CO)NH spectra. Side chain resonance assignments were done by a combination of HCCH-TOCSY, 3D NOESY ^{1}H-^{15}N HSQC and 3D NOESY ^{1}H-^{13}C HSQC (aromatic and aliphatic region) spectra. All spectra were recorded at 298 K expect for SRSF6 1–72 + 3.6x Poly-C RNA, which was recorded at 295 K to observe a more stable complex.

NMR structure calculations

For NMR structure calculations of SRSF6 RRM1, ^{1}H-^{1}H distance restraints obtained from 3D-^{1}H-^{13}C-NOESY-HSQCs (individual for aliphatic and aromatic with mixing times of 125/150 and 150 ms, respectively) and a ^{1}H-^{15}N-NOESY-HSQC (mixing time of 150 ms) were used. NOEs were assigned automatically using the CYANA[31] routine (shift tolerances: 0.06 (H), 0.04 (H’), 0.5 (C’/N’)). Additionally, torsion angle restraints based on available secondary chemical shifts and calculated from TALOS-N v.4.12[32] were used for the correct folding of secondary structure elements.

CYANA v3.98.15[33] was used to create an initial 100 structures and refined by 10,000 torsion angle dynamics steps. Seven iterative rounds of structural model building and refinement were followed by a final refinement step, which created an ensemble of 20 structures. From cycle 1 to the final cycle the average backbone RMSD decreased from 2.82 Å to 0.24 Å. The Ramachandran plot of the SRSF6 RRM1 indicates that 80.6 %of the residues are in the most favored regions, 17.9 %in the additionally allowed regions, 1.5 %in the generously allowed regions, and 0 %in the disallowed regions. The final 20 lowest-energy structures were used for deposition in the PDB/BMRB and subsequent structural analysis. For an overview of structural statistics, see Supplementary Table 12.

X-ray crystallography

Initial crystals were obtained from crystallization trials in 96-well SWISSCI plates with commercial screens by the sitting drop vapor diffusion method. Crystals appeared within the first week at 4°C, with only few crystals growing at later time points. Initial crystals were screened and further optimized. Obtained crystals were cryo-protected in mother liquor and frozen in liquid nitrogen until measurement. Datasets were collected at the EMBL P13 beamline at the PETRA III storage ring of the DESY synchrotron[34]. Pre-processed unmerged datasets from autoPROC+STARANISO[35] were further processed in CCP4cloud v1.8[36].

Phases for the RRM2 apo and RNP structure were obtained by molecular replacement (MR) using Phaser[37] from an AlphaFold2[38] model. For the structure of the RRM2 W123A mutant MOLREP[39] was used instead. Structures were build using ModelCraft[40] as an automatic model-building pipeline, refined in BUSTER[41] and REFMAC[42], optimized by PDB-REDO[43] and manually adjusted in Coot[44].

All structures were present as single protein molecule chains per crystal unit. For the RNP crystal, a 1:1 RNA-protein complex was visible in the asymmetric unit with only three out of six nucleotides showing electron density. All depictions of the structures were created in PyMOL version 3.10 (The PyMOL Molecular Graphics System, Schrödinger, LLC.).

For each structure, residues 106–185 (construct 86–188) could be built showing the flexibility of the N-terminal linker and the C-terminus. Additional experimental details are provided in Supplementary Table 13.

Small angular X-ray scattering measurements

Size exclusion coupled to small angular X-ray scattering (SEC-SAXS) measurements were performed at 293 K at the EMBL P12 BioSAXS beamline at PETRA III storage ring of the DESY synchrotron[45] equipped with a Pilatus 6M detector. Samples were stored at 30°C prior to measurements to avoid phase separation behavior when mixed with RNA. All samples were prepared fresh on-site from individual purified components. For this, protein was buffer exchanged to SEC-SAXS-buffer (20 mM HEPES, 150 mM NaCl, 380 mM l-Arg/Glu, 2 mM TCEP, pH 7.0) and concentrated to > 10 mg/ml in 0.5 ml Amicon® ultra centrifugal filter units with a molecular weight cutoff of 10 kDa. Commercial RNA was used at a stock concentration of 3 mM in ultrapure H_2O and mixed with the protein in a 1:1 molar ratio. Final concentrations were adjusted by buffer addition. For each SEC-SAXS sample, a final volume of 35 or 55 µl was prepared, of which 30 or 50 µl, respectively, were loaded onto a Superdex 75 Increase 5/150 GL column (Cytiva™) using an Agilent 1260 Infinity Bio-Inert HPLC system at a flow rate of 0.33–0.35 ml/min. 2,400 or 2,640 successive 2D SAXS data frames of 0.25 s each were collected from the continuously flowing eluate spanning one column volume (3 ml) and delivered directly to the beamline.

SEC-SAXS data averaging and reduction was performed using the ATSAS 3.2.1 software suite[46]. In a first step, CHROMIXS[47] was used to choose the right sample frames (guided by uniform R_g values) and

perform buffer subtraction, which was used to produce the final 1D profiles. These were further processed using the BIOXTAS RAW 2.3.0 software suite[48]. First, the data was converted from a nm to Å scale by dividing q-values by 10. Secondly, every 1D profile was visually analyzed in the I(q) vs q representation and inspected for up- or downturn values in the low q range which indicate aggregation or interparticle interference, respectively. The first 5–15 outlier values in this region were then deleted to improve data quality and Guinier fit values. These manually adjusted 1D datasets were then used for all further analysis.

Modeling of the tandem RRM

For modeling of the tandem RRM, optionally in complex with RNA, initially AlphaFold3 (AF3) protein models[49] were generated with standard settings using the native protein sequence of residues 1-188 with a “GAG” extension at the N-terminus as obtained from the cloning procedure. Additionally, the X-ray structure of SRSF6 RRM2 bound to CGGACU RNA (PDB ID 9HX7) was used. For RRM1, the NMR structure of SRSF1 RRM1 bound to AACAAA RNA (PDB ID 6HPJ) was employed, due to the similar RNA binding mode of SRSF1 and SRSF6 RRM1 and the lack of a high-resolution SRSF6 RRM1-RNA complex structure. For each RNA oligonucleotide used in the modeling, two separate complex models were generated, with the RNA bound either to RRM1 or to RRM2. The RRM2-RNA model was created by superimposing RRM2 from the X-ray structure onto the AF3 model, retaining the “GGA” nucleotides of the X-ray complex and extending both RNA ends. The RRM1-RNA model was constructed by superimposing the SRSF1 RRM1-RNA NMR structure onto the SRSF6 RRM1 AF3 model, keeping the coordinates of the “CA” nucleotides, RNA backbone, and sugars fixed relative to the protein. Missing bases were generated by fixing the position of the glycosidic bond nitrogen and the two adjacent carbons, resulting in an RNA conformation with all bases oriented as in the NMR structure.

The model of the tandem RRM with RRM1 bound to RNA was used as a template to generate an ensemble of structures with randomized linker conformations. Structures were produced by applying random phi/psi rotations to residues in the linker and at the N- and C-termini of the protein[50]. If a random phi/psi rotation resulted in atom clashes between the region C-terminal to the rotated residue and the N-terminal portion of the structure, the rotation was rejected and the original angles were restored. Randomization was followed by a short four-step simulated annealing energy minimization. In a second iteration, the second part of the RNA was removed and positioned according to the RRM2-RNA template model, producing an RNA with a broken bond in the middle. To restore a continuous RNA conformation, a second, longer four-step simulated annealing energy minimization was performed. For each protein-RNA combination, a total ensemble of 5,000 different models was generated.

The complete ensemble was used to select the best-fitting models for the apo protein or each protein-RNA complex based on SAXS data. Each model was fitted to its corresponding processed 1D SAXS profile using the Crysol function implemented in BIOXTAS RAW 2.3.0[48] with standard settings. The

resulting fits were ranked according to their Chi^2 values, and the ten models with the lowest values were selected for further analysis of R_g, D_{max} and structural alignments.

Molecular modeling for simulations

In the absence of an experimental structure of SRSF6 RRM1 with RNA we predicted the starting structures using AF3[49] with different random seeds. We used the sequence of RRM1 (residue 1 to 72) from UniProt (ID Q13247_1) and the following RNA tetramers: GCCG, GUCG, CCCC, CCAC, and CAUC.

For the simulations of the RRM2-RNA complex, the starting coordinates were taken from the crystal structure of SRSF6 RRM2 bound to a GGA RNA motif (PDB ID: 9HX7) published in this work. For simulations of the RRM2 domain bound to a GAA motif, a G2A mutation was introduced using PyMol v3.10 (The PyMOL Molecular Graphics System, Schrödinger, LLC.[51]).

The complexes were solvated with the SPC/E water model[52] in a cubic simulation box with a minimum distance of 12 Å between solute and box edges. KCl was added to neutralize the system and achieve a bulk ion concentration of 150 mM. Molecular interactions were modeled using the ff14sb parameter set for protein with the improved side-chain torsion potentials[53,54] and OL3 parameter set for RNA[55].

Molecular Dynamics simulations

All MD simulations were performed using GROMACS 2025.1[56]. Each system was first energy minimized using the steepest descent algorithm for 50000 steps and then heated gradually from 0 to 310 K in steps of 50 K. Energy minimization was followed by a short (20 ns) equilibration in the NVT ensemble, with position restraints applied on all heavy atoms. Subsequently, the systems were equilibrated for 1 μs in the NPT ensemble, during which pairwise distance restraints were applied to all protein-RNA heavy-atom pairs within 4.5 Å of each other in the initial structure. The force constants for the harmonic restraints were gradually reduced in three stages: 500 ns at 500 kJ mol⁻¹ nm⁻², 250 ns at 200 kJ mol⁻¹ nm⁻², and 250 ns at 50 kJ mol⁻¹ nm⁻². Production MD simulations were performed for 5 μs for each replica. All simulations were performed with a time step of 2 fs and with periodic boundary conditions enabled. The equations of motion were integrated using the leap-frog algorithm. The temperature and pressure were maintained at 310 K ($\tau_t = 0.1\ \mathrm{ps}$) and 1 bar respectively, using the stochastic velocity rescaling (v-rescale) thermostat[57] and stochastic exponential relaxation (c-rescale) barostat[58] with $\tau_p = 5.0\ \mathrm{ps}$, a compressibility of $4.5 \cdot 10^{-5}\mathrm{bar}^{-1}$ and an isotropic pressure coupling.

Long range electrostatics were modeled using the Particle Mash Ewald algorithm[59,60] with a grid spacing of 0.12 nm. Short-range Coulomb and van der Waals interactions were computed up to a distance cutoff of 0.9 nm. All bonds with hydrogen atoms were constrained using the LINCS algorithm[61]. RRM1-RNA simulations were run for 5 μs per system. For the ΨRRM2-RNA simulations five replicas of 5 μs length were performed, yielding an aggregate simulation time of 25 μs per system.

Visualization was performed using VMD v1.9.3[62]. Trajectories were aligned using an in-house program for RMSD-based alignment of heavy atoms. Structural analysis was carried out using the GROMACS rmsd tool, MDAnalysis v2.9.0[63,64] and the NumPy v. 2.2.5[65], SciPy[66], and pandas python v2.2.3[67] packages. Plots were generated using matplotlib[68]. RMSD curves were smoothed using th Savitzky–Golay filter from SciPy.

We estimated off-rates k_{off} for GGA and GAA RNA binding to RRM2 from the five independent replica simulations per system according to Eq. 2. The number of observed unbinding events $\mathrm{N_{unbound}}$ is divided by the time the complex stayed bound, which is defined as the sum of the residence times prior to dissociation t_i and the durations of trajectories in which the complex remained bound throughout T_j.

$$k_{off} = \frac{N_{unbound}}{\sum_i t_i + \sum_j T_j}$$

2

Unbinding of RNA and protein is defined as when the RMSD of heavy atoms exceeded a cutoff of 10 Å for longer than 5 ns. The error is calculated from the number of unbinding events observed according to Eq. (3).

$$Error = \frac{k_{off}}{\sqrt{N_{unbound}}}$$

3

Cell culture and SRSF6 knockdown

HeLa and HEK293F wild-type cells (Thermo Fisher Scientific) were cultivated in 10-cm dishes under humidified conditions at 5% CO_2 and 37°C in Dulbecco's modified Eagle's medium (DMEM) Glutamax (Sigma Aldrich) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS; Thermo Fisher Scientific) and 100 U/ml penicillin–streptomycin (Thermo Fisher Scientific). For *SRSF6* knockdown HeLa wild-type cells were transfected (in triplicate) with two different siRNAs (Sigma Aldrich) at a final concentration of 66 nM using jetPRIME Transfection Reagent (Polyplus). The mission universal negative control siRNA (Sigma Aldrich) was used as control. After 48 h, the cells were harvested for RNA extraction.

RNA extraction cDNA preparation

HeLa cells with and without *SRSF6* knockdown or HEK293F cells grown on 10-cm cell-culture dishes to 80% confluence were harvested by scraping and lysed in TRIzol (Thermo Fisher Scientific). RNA was extracted according to the manufacturer's recommendations and treated with TURBO DNase (Thermo Fisher Scientific) for 30 min at 37°C to remove genomic DNA. 2 µg total RNA was reverse transcribed

into cDNA using SuperScript III kit (Life Technologies) with 10 mM dNTP Mix (Thermo Fisher Scientific) and oligo d(T)-18 (Sigma Aldrich).

RNA-seq and splicing analyzes

To identify SRSF6-regulated splicing events an RNA-Seq experiment was performed. 5 µg of total RNA from three control and three knockdown samples were subjected to ribo-depletion and library generation (Novogene, Cambridge). Libraries were sequenced on an Illumina HiSeq 4000 machine (single-end 75 nucleotide reads, 30 million reads, 3 replicates per condition). RNA-seq reads were mapped against the human genome (version hg38) with GENCODE gene annotation using STAR (v2.6.1d[69]) with the following parameters:

STAR \ --runMode alignReads --outSAMattributes All --outSAMtype BAM SortedByCoordinate --outFilterMismatchNmax 999 --outFilterMultimapNmax 1 --outFilterMismatchNoverLmax 0.04.

To detect alternative splicing events, rMATS (v3.2.5[70]), which uses an improved version of the generalized linear mixed model to detect differential AS from RNAseq data with replicates was used. rMATS accounts for exon-specific sequencing coverage in individual samples as well as variation in exon splicing levels among replicates. It was used with parameter –c 0.0001, considering reads mapped to splice junctions and the cutoff was set at a false discovery rate (FDR) < 0.05 and absolute change in 'percent spliced in' (|ΔPSI|) ≥ 5%.

Splice PCR

To monitor changes in the inclusion of exon 6 of *MDM4* and exon 4 of *CLK1*, primers binding in exons flanking these regulated cassette exons (CEs) (Supplementary Table 14) were designed using SnapGene (GSL Biotech LLC, Chicago US). A semi-quantitative PCR was performed using Phusion DNA polymerase (New England Biolabs) for 30 cycles. PCR products were analyzed on a 2% agarose gel and images were acquired with a ChemiDoc MP Imaging System (Bio-Rad). For isoform quantification, a TapeStation (Agilent Technologies) was used. Plots and statistical analysis were done with Prism (GraphPad, US).

Determination of in vivo binding motifs of SRSF6 from iCLIP data

The *in vivo* binding motifs of mouse SRSF1 and SRSF6 across all sites were derived from iCLIP data performed in murine P19 cells as described before[19]. The *in vivo* binding motif of human SRSF6 across all sites was derived from iCLIP data performed in HeLa cells as described recently[13]. To extract functional binding motifs of SRSF6 relevant to alternative splicing, all exons that were skipped or more included after SRSF6 knockdown in HeLa cells were compiled and filtered for SRSF6 iCLIP signal[13]. The sequences of these exons were extracted as fasta files using R and used as input for DREME (Discriminative Regular Expression Motif Elicitation) to obtain enriched motifs (https://meme-suite.org/meme/doc/dreme.html).

Genome editing (SRSF6 KO) with CRISPR-Cas9

SRSF6 KO cells were generated by genome editing using CRISPR/Cas9 and a single guide RNA (sgRNA) as described previously[71]. The sgRNA (Supplementary Table 15) was designed with the IDT design tool (https://eu.idtdna.com/pages/tools/alt-r-crispr-hdr-design-tool). The sgRNA cut within exon 2 of *SRSF6* and generated a frameshift due to non-homologos end joining. 1 µM sgRNA was pre-mixed with Cas9 protein (IDT) and the resulting Cas9-sgRNA RNP complex was transfected to the HEK293F cells grown on a 6-well cell culture plate using Lipofectamine CRISPRMAX (Invitrogen) and incubated for 24 h. The media was exchanged with fresh DMEM media supplemented with 10% FCS and PenStrep as described above. The cells were recovered for 48 h and were directly used for single-cell generation on 96-well plates by limited dilution approach. After further incubation (~10 days), all potential single-cell clones were subjected to immune blot analysis using anti-SRSF6 antibodies. SRSF6 KO was further confirmed by genomic screening by PCR and Sanger sequencing.

Construct design for inducible and stable expression of SRSF6 versions in HEK293F

Sequences encoding SRSF6 WT were amplified by PCR using cDNA prepared from HEK293F WT total RNA. The doxycycline inducible vector pSB-Tet and the plasmid encoding Sleeping Beauty (SB) transposase (SB100X) were purchased from Addgene (Plasmid #60510 and Plasmid #127909, respectively). The ORF encoding GFP was removed from the pSB-Tet parental plasmid by PCR to generate the pSB-Tet(Mod) vector (Supplementary Table 16). SRSF6 was fused C-terminally with an in-frame HRV-3C protease cleavage site followed by the sequence encoding GFP and a Twin-Strep II affinity purification-tag, using modified primers and overlap PCR. All primers are listed in Supplementary Table 17.

For the generation of stable cell lines, adherent HEK293F cells grown on 6-well plates, were co-transfected with pSB-Tet(Mod) plasmids carrying different SRSF6 versions and the SB100x vector using JetPRIME® transfection reagent (Polypus Life Sciences, Sartorius, France) according to the manufacturer's instructions. For each well 2 µg DNA was prepared by mixing 1.9 µg of the pSB-Tet(Mod) vector with 100 ng of the pSB100X vector. 24 h post transfection, cells were subjected to selection by blasticidin (10 µg/ml) for at least 14 days. Media was exchanged every second day until all untransfected control cells were dead. Surviving cells were further cultivated for 2 weeks without selection medium to ensure genomic integration.

Protein extraction and immunoblotting

To control expression of the different SRSF6 variants, 2 x 10^6 cells were pelleted by centrifugation at 500 x g for 5 min. The cell pellets were washed with phosphate-buffered saline (PBS) and resuspended in lysis buffer (20 mM HEPES pH7.4, 150 mM NaCl, and 0.5% SDS) supplemented with 1X complete protease inhibitor mix (Roche) and 2 U/ml DNase. The cells were lysed by pipetting and incubated on ice for 30 min. The lysate was cleared by centrifugation at 15,000 x g for 10 min and the supernatant was supplemented with reducing 1X SDS-PAGE sample loading buffer (Thermo Fischer Scientific) and heated at 95°C for 10 min. 20 µg (~ 10 µl) total protein was separated on a 3–12% SDS-PAGE and transferred to

pre-activated PVDF membrane by electrophoresis for immunoblotting. The membrane was blocked with 5% non-fat milk in Tris-buffered saline supplemented with 0.05% Tween-20 (TBST) for 1 h at room temperature. All antibodies were diluted in 3% milk prepared in TBST. The list of antibodies and their dilutions are listed in Supplementary Table 18.

Expression and purification of SRSF6 protein variants from HEK293F cells

For protein purification, stable HEK293F cell lines (Supplementary Table 19) were adapted to growth in suspension and were cultured in FreeStyle™ (Thermo Fisher Scientific) media in 1000 ml Erlenmeyer flasks under 135 rpm shaking at 8% CO_2 and 37°C. SRSF6 variant expression was induced with 2 µg/ml doxycycline for 48 h. Cells were harvested by centrifugation at 1000 x g for 5 min and cell pellets were stored at -80°C or resuspended in lysis buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 10% Glycerol, 0.5% NP-40) supplemented with 1X complete protease inhibitor mix (Roche), and 5 mM PMSF. Cell lysis was completed by sonication (3 cycles of 20 sec pulse and 5 min pause with 20% amplitude, Branson). Lysates were cleared by centrifugation at 17,000 x g for 30 min and incubated with Strep-Tactin agarose beads (IBA Life Sciences) for 4 h on a rotating wheel. Beads were then collected on a gravity-flow column and washed with wash buffer (20 mM HEPES pH 7.4, 1 M NaCl, 10% Glycerol, and 0.1% DDM (w/v)). Proteins were eluted from the beads by adding 1 ml cleavage buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 10% Glycerol, and 0.1% DDM (w/v), 5 mM DTT, 2 mM EDTA) containing 10 U HRV-3C-protease (Merck Millipore, #71493) and incubated for 12–14 h at 4°C on a rotating wheel. The protein eluates were then incubated with Ni-NTA (IBA Life Sciences) for 1 h to remove the 3C-protease. The purified proteins were then concentrated 3-fold using 10 kDa molecular weight cut-off (MWCO) centrifugal concentrators (Amicon Ultra-0.5 ml, Merck Millipore). The protein samples were centrifuged at 17,000 x g for 10 min to remove any aggregates. The purified proteins were further polished by size exclusion chromatography (SEC) on an ÄKTA-go using a Superose™ 6 10/300 GL (Cytiva) column in SEC buffer (20 mM HEPES pH 7.4, 500 mM NaCl, 10% Glycerol, and 0.1% DDM (w/v), 2 mM DTT, 5 mM PMSF). The elutions were collected in fractions and those containing SRSF6 were combined and concentrated by centrifugation using 10 kDa MWCO concentrators (Amicon Ultra-0.5 ml, Merck Millipore) to 1 mg/ml final concentration. Aliquots of the protein were flash-frozen in liquid-nitrogen and stored at -80°C.

Immunofluorescence staining and confocal imaging

For immunofluorescence (IF) experiments, 12-mm coverslips (Carl Roth) were coated with 0.5% Poly-D-Lysin (PDL; Gibco) for 24 h and dried for 8 h. HEK293F cells were seeded at 50% confluence and allowed to attach for 12 h before inducing the cells with 50 ng/µl Dox for 48 h. After removing the medium and washing the cells with 1x PBS, cells were fixed with 4% PFA (Electron Microscopy Sciences) in PBS for 15 min, washed with PBS, and permeabilized with 70% ethanol for 4 h. After washing with PBS the cells were blocked using immunofluorescence staining (IF) buffer [3% (w/v) bovine serum albumin (BSA), 0.1% (v/v) Triton-X 100 in 1x PBS] for 30 min and incubated with anti-SON antibody (Santa Cruz Biotechnology) for 12 h at 4°C and then labeled with anti-mouse AlexaFluor-647 secondary antibody (Thermo Fischer Scientific) for 2 h at room temperature. Nuclei were stained with Hoechst 34580 (Sigma

Aldrich GmbH; 1:4,000 dilution in 1x PBS) for 30 min. The cells were then mounted on a glass slide with mounting media and cultured overnight before imaging.

Images were acquired with confocal laser-scanning microscope (LSM780; ZEISS) using the Zen 2012 (black edition; 8.0.5.273; ZEISS). Fluorescence signal was detected with an Argon laser (GFP – 488 nm, and Qasar 670 – 647 nm). Images from the same experiment were acquired with the same settings for all conditions. The images were analyzed with ImageJ software.

Fluorescent polarization

Fluorescent polarization (FP) assays were used to quantify RNA binding affinities. 5'-FI labeled RNA and a dilution series of each protein were prepared in FP-buffer 1 (20 mM NaPi, 150 mM NaCl, 380 mM l-Arg/Glu, 2 mM TCEP, 2 mM $MgCl_2$, 0,02% (v/v) Tween-20, pH 7.0), FP-buffer 2 (20 mM NaPi, 150 mM NaCl, 2 mM TCEP, 2 mM MgCl2, 0,02% Tween-20, pH 7.0) or FP-buffer 3 (20 mM NaPi, 150 mM NaCl, 2 mM TCEP, 2 mM MgCl2, 0,02% Tween-20, pH 6.5). 10 µl of each dilution was mixed with 5 µl RNA (0.1 µM) in Greiner flat black 384-well plates with non-binding surface. The plates were incubated for 10 min at room temperature (21–25°C), placed in the TECAN Spark® microplate reader, shaken for 10 s at 810 rpm and measured at 25°C unless specified otherwise. An excitation wavelength of 485 nm (20 nm bandwidth) was used and emission measured at 535 nm (20 nm bandwidth). Gain and plate position were adjusted automatically. Measurements were performed as technical triplicates (unless specified otherwise) and FP values were automatically calculated by the SparkControl software according to Eq. 4, where $F_{\parallel}$ is the fluorescence intensity parallel to the excitation plane and $F_{\perp}$ is the fluorescence intensity perpendicular to the excitation plane.

$$FP = \frac{\left(F_{\parallel} - F_{\perp}\right)}{\left(F_{\parallel} + F_{\perp}\right)}$$

4

Mean FP values with their respective standard deviation were plotted against the logarithmic protein concentration in OriginPro 2021b and fitted with a Hill1 fit, from which -where possible- we derived the K_D values.

Electrophoretic mobility shift assay

Samples derived from FP measurements (described above) were additionally loaded onto native polyacrylamide gels (PAA) for visualization of complex formation via EMSA. 15 µl taken directly after FP were mixed with 3 µl loading dye (60% glycerol, 1x TB (89 mM TRIS base, 89 mM boric acid), 0.02% bromophenol blue) and loaded onto 6% PAA gels. Gels were run in pre-cooled 1x TB buffer for 80 min at 80 V and subsequently imaged within glass plates in the ChemicDoc imager (Biorad) by the “fluorescein blot” option. The exposure time was automatically adjusted for each gel.

Microscale thermophoresis and dual-emission spectral shift measurements

For FL SRSF6 variants, RNA protein binding analysis was carried out using Fl- and Cy5-labeled 12-mer RNA. Fl-labeled RNAs were used for FP measurements as previously described. Cy5-labeled RNAs were used for spectral shift measurements with purified proteins in coated capillaries (MO-K025, NanoTemper Technologies GmbH) on a Monolith-X (NanoTemper Technologies GmbH) instrument equipped with dual-emission detection optics. A 12-point serial dilution of the purified protein was prepared at a final volume of 10 µl in each dilution step with 37.5 µM as highest concentration. Then 2.5 µl of Cy5-labeled RNA was added to all protein dilutions yielding 30 µM as the highest protein concentration. 10 µl of each sample was then loaded into the Monolith X device and fluorescence intensity values were recorded at 650 and 670 nm. The ratiometric reading, 670 nm/650 nm were calculated automatically in the control software, and data was fitted with a 1:1 binding model to derive the K_D values.

For the RRM constructs the same protein dilutions as for FP/EMSA measurements were used with Cy5-labeled RNAs to obtain K_D values for RNAs > 20 nt. 6 µl of each protein dilution were mixed with 6 µl of 12 nM RNA in its respective buffer in 200 µl PCR reaction tubes and transferred by capillary force to Monolith standard capillaries. Measurements were performed in a NanoTemper Monolith X at 25°C with wavelengths at 650 and 670 nm. The excitation power was adjusted automatically for each set of experiments. For MST measurements the IR-laser power was set to medium with an off-on-off time of 5-10-5 s. Quality checks and data analysis were performed directly in the MO.Control 2 software.

Results

SRSF6-regulated exons are enriched in a CUGGAG motif

To resolve the discrepancy between previously reported SRSF6 binding motifs, we asked whether SRSF6 recognizes distinct sequence elements in regulated alternative exons compared with non-regulated constitutive exons. To identify SRSF6-dependent splicing events, we depleted SRSF6 in HeLa cells using a pool of two siRNAs and performed RNA-seq after ribodepletion (48 h, three biological replicates per condition). Differential alternative splicing was analyzed with rMATS[70].

Using a threshold of false discovery rate (FDR) < 0.05 and an absolute change in percent spliced in ($|\Delta PSI|$) ≥ 5%, we identified 1,773 significantly altered alternative splicing events, comprising 1,018 cassette exons (CEs), 143 alternative 3' splice site events (A3SS), 202 alternative 5' splice site events (A5SS), 193 mutually exclusive exons (MXEs), and 217 retained introns (RIs) (Supplementary Fig. 1). CEs represented the largest class of regulated events. Among these, 411 exons showed increased inclusion upon SRSF6 depletion, indicating that SRSF6 normally represses their inclusion, whereas 607 exons showed increased skipping, indicating that SRSF6 normally promotes their inclusion (Supplementary Fig. 1).

To define sequence features associated with direct regulation, we intersected the regulated CEs with our published SRSF6 iCLIP dataset set[13] and performed de novo motif discovery using DREME. Exons whose inclusion is normally repressed by SRSF6 (more included after SRSF6 depletion) were enriched for the motif CCUCAGC, a sequence that does not resemble any previously described SRSF6 binding motif. In contrast, exons whose inclusion is normally promoted by SRSF6 (more skipped after SRSF6 depletion) were strongly enriched for the motif CUGGAG (E-value = $2.1x10^{-5}$; Fig. 1B). This motif contains a GGA core and closely matches the high-affinity sequences identified by *in vitro* selection approaches[15,17]. In contrast, analysis of all SRSF6-bound exons, irrespective of splicing regulation, clearly revealed a predominance of GAA triplets (Fig. 1B), consistent with our previous iCLIP analysis (Fig. 1A)[19]. These findings indicate that the functionally relevant SRSF6 recognition element in regulated exons is a composite motif centered on a GGA core embedded within a C(U)-rich context, whereas widespread binding to isolated GAA triplets likely reflects constitutive, non-regulatory interactions.

Structures of the individual SRSF6 RRM domains

To understand how SRSF6 binds to these two types of and possible other motifs at an atom-resolved level, we next obtained structural information starting from individual SRSF6 domains. The protein comprises a canonical RRM (RRM1) at its N-terminus followed by a pseudo RRM (ΨRRM2) and a C-terminal RS domain (Fig. 1C). We first determined the solution structure of RRM1 using nuclear magnetic resonance (NMR) spectroscopy (Fig. 1D, PDB ID 9R53), represented by an ensemble of the 20 well-converging lowest-energy structures (backbone RMSD: 0.24 Å). This revealed that RRM1 adopts a canonical β1-α1-β2-β3-α2-β4 topology, in which the four β-strands form an anti-parallel β-sheet with RNP motifs located in β3 (RNP1) and β1 (RNP2) (Fig. 1D, Supplementary Figs. 2A, 3A, 4A, and 5A). In comparison to available structures of SRSF RRM1 domains, SRSF6 RRM1 embeds very short loop regions and thus represents a highly compact RRM (Fig. 1E, Supplementary Fig. 2A).

To determine the structure of ΨRRM2 we used X-ray crystallography, because we encountered NMR-unfavorable dynamics in α-helix 1 (α1) and the loop between β2 and β3[21]. Crystals diffracted to a resolution of 2.1 Å and revealed electron density for residues 106–185. In line with NMR secondary chemical shifts (SCS), ΨRRM2 adopts a canonical-like β1-α1-β2-β3-α2-β4 fold (Fig. 1F, PDB ID 9HX5, Supplementary Figs. 2C, 3B, 4B and 5B). Instead of RNP motifs, ΨRRM2 contains a heptapeptide motif (SWQDLKD) located in α1 that is conserved among all five ΨRRM2-containing SRSF proteins (Supplementary Fig. 3B), and which was found to be the main RNA binding site in SRSF1[10]. In agreement with a higher sequence conservation across different SRSF proteins, ΨRRM2 structures converged much more than those of RRM1 domains (Fig. 1E and G, Supplementary Fig. 2).

While we did not succeed in obtaining the structure of the SRSF6 tandem RRMs, we successfully expressed and purified RRM1 + ΨRRM2 with the full native linker region (aa 1-188, Fig. 1C) and assigned around 90% of the ^{1}H and ^{15}N-backbone chemical shifts[21], providing the basis for detailed RNA binding analyzes. NMR SCS of the tandem RRMs revealed domain boundaries for both structured RRMs that

align with those found with the isolated domains. The 37-residue inter domain linker remains unstructured in solution (Supplementary Fig. 5C). Collectively these data provide the first high-resolution structures of the SRSF6 RRM domains and the successful purification of SRSF6 tandem RRMs including the native linker, establishing the basis to interrogate domain-specific contributions and their combinatorial roles in RNA binding.

A: Comparison of *in vivo* binding motifs of mouse SRSF1 (left) and SRSF6 (right) considering all sites using iCLIP data from Müller-McNicoll et al[19]. B: Comparison of *in vivo* binding motifs of human SRSF6 considering all exons (left) or only regulated exons (right: skipped after SRSF6 depletion) integrating RNA-seq data and iCLIP data from Oliveira Freitas Machado et al[13]. C: Domain architecture of SRSF1 and SRSF6. SRSF6 constructs used in this study are shown with respective secondary structures for each domain[21]. Domain boundaries are given by amino acid residues. D: Representative NMR structure of SRSF6 RRM1 with secondary structure elements labeled. E: SRSF6 RRM1 structure from D aligned and overlaid with two other SRSF RRM1 structures. The RMSD of each structure to SRSF6 RRM1 is given. F: Crystal structure of SRSF6 ΨRRM2 with secondary structure elements labeled. G: SRSF6 ΨRRM2 structure from F aligned and overlaid with two other SRSF ΨRRM2 domain structures. The RMSD of each structure to SRSF6 ΨRRM2 is given. The PDB IDs for each structure are given in Supplementary Fig. 2. For the full ensemble of RRM1 and electron density map of ΨRRM2 see Supplementary Fig. 4.

SRSF6 RRM1 binds C-rich RNA

To systematically characterize the RNA binding of SRSF6, we next defined the RNA motif preferences and affinities of individual domains in an unbiased way and performed RNA-bind-and-Seq (RBNS) from random pools[23]. RNA motifs enriched for two variants of RRM1 clustered into similar C/A-rich sequence logos (Fig. 2A and B, Supplementary Fig. 6A and B). This preference was confirmed by NMR spectroscopy using poly-N RNAs, where RRM1 bound best to poly-C, followed by poly-A, whereas no significant interactions with either poly-G and poly-U were observed (Fig. 2C, Supplementary Fig. 7). The near-exclusive fast exchange regime of the interaction with poly-C is indicative of moderate affinities. The resulting CSP plot confirms a canonical RRM binding pattern through the RNP1 and RNP2 motifs, yet reveals an additional contribution from the RRM1 C-terminus (Fig. 2D). Of note, R71 is in intermediate exchange with the RNA suggesting local high-affinity binding. A large downfield shift of R71 further hints towards an H-bond formation with the RNA, in agreement with an Alphafold3 (AF3) model suggesting the backbone N-H of R71 to form an H-bond with one of the cytosines of poly-C (Fig. 2E). To experimentally probe the role of specific interactions we used MD simulations of RRM1with varying nucleotides in a tetramer context (assumed as the minimal RNA length to cover the canonical binding surface) over a total length of 25 µs per system (Fig. 2F and Supplementary Figs. 8 and 9). We found a stable interaction with cytosines in the two central positions as shown by small positional deviations during simulations, majorly based on stacking interactions with Y5 of RRM1 forming a C-specific pocket. In contrast, non-cytosine bases showed significantly stronger deviations, expressed by higher mean RMSD values. This is supported by additional electrostatic interactions between R71 and the cognate nucleotides. In line with overall RMSD values, we found stable center-of-mass (COM) distances between

R71 C5 and cytosine C5 atoms, in agreement with present H-bonds, while a more than twofold larger distance was found with an AU dinucleotide resulting in subsequent unbinding of AU (Supplementary Figs. 8 and 9). Interestingly, within the unstable CAUC-RRM1 complex the flanking cytosine migrated into the binding site during the simulation. The C4-binding mode reproduced the interactions observed for the original C binding pose indicating that the pocket is selective for a cytosine. These data are in line with RRM1's selectivity for C-rich contexts found in RBNS and NMR.

In line with previous studies on canonical RRMs[72], mutation of the aromatic residues within RNP1 (Y37A or F39A) or RNP2 (Y5A) abolished RNA binding completely (Supplementary Fig. 10A-C). In contrast, mutation of R71 to alanine reduced RNA binding only slightly, retaining the RNA interaction mediated through the RNPs (Supplementary Fig. 10D). Taken together, we show that SRSF6 RRM1 binds C(A)-rich RNAs via a canonical RRM interface with a critical role for R71 in RRM1 specific RNA-engagement.

SRSF6 ΨRRM2 binds GGA motifs

We next sought to identify binding motifs of the SRSF6 ΨRRM2. In contrast to RRM1, NMR titration of ΨRRM2 with poly-N RNAs did not reveal clear binding preferences for either of the poly-N RNAs (Supplementary Fig. 11). We found similar, minor CSPs distributed over the entire domain, indicating a weak and non-specific interaction. We thus employed RBNS with random pools to account for higher-complexity binding motifs. The sequences enriched for ΨRRM2 including the linker region were G-rich interspaced with some adenosines, indicating a clear preference for poly-purines (Fig. 2G, Supplementary Fig. 6C and D). We used NMR to decipher whether SRSF6 prefers GAA or GGA and found that ΨRRM2 binds both motifs (Supplementary Fig. 12), but has a clear preference for GGA over GAA (Fig. 2H). Significant CSPs cluster to α1 and β2 – the primary ΨRRM2 RNA binding interface – hinting towards a similar binding mode for GAA and GGA. The preference for GGA over GAA is reflected by the absolute magnitude of CSPs in combination with divergent exchange regimes (slow for GGA vs. fast for GAA, Supplementary Fig. 12B/C), suggesting a higher affinity for GGA. In line with that, a significantly lower K_D for GGA over GAA was confirmed by FP (Fig. 2I). Altogether, these data confirm the *in vitro* SRSF6 preference for GGA over GAA and reveal ΨRRM2 as relevant for the distinction, similar to the SRSF1 ΨRRM2[10]. This underlines the strong conservation and redundancy within the ΨRRMs of SR-proteins but fails to explain why *in vivo* GAA is preferred by SRSF6.

A detailed comparison of -1 positions preceding the GGA motif by FP further confirmed a contribution to the binding affinity of ΨRRM2 (Supplementary Fig. 13). Here, an adenosine (K_D = 1.18 ± 0.19 µM) was preferred over cytosine (K_D = 1.99 ± 0.26 µM) and uridine (K_D = 2.22 ± 0.80 µM), while an additional guanosine abolished binding. However, this might be related to the formation of oligomers/G-quadruplexes in this RNA, which would block access to the single-stranded motif, as indicated by the higher initial FP-values of RNA alone (Supplementary Fig. 13). We conclude that SRSF6 ΨRRM2 employs a non-canonical RNA motif recognition and prefers the optimal recognition motif as [A > C>U > G]GGA *in vitro*.

The highly divergent binding preferences of SRSF6 RRM1 and ΨRRM2 prompted us to next identify RNA motifs preferentially bound by the tandem RRM1-ΨRRM2 construct. RBNS revealed the motif 'CGGACU' among the most enriched motifs by the tandem RRMs (Fig. 2J, Supplementary Fig. 6E), which contains a GGA triplet and resembles the CUGGAG motif derived from our splicing data (Fig. 1B). Given that both composite motifs reflect the binding preferences of domains in isolation (RRM1: C-rich RNA / ΨRRM2: GGA), we conclude that both RRM domains of SRSF6 are required for selective target recognition and that, in this context, GGA is preferred through the ΨRRM2 site. CGGACU and CUGGAG motifs can be considered minimal composite motifs, which may have a particular role in SRSF6-mediated splicing decisions, where CU can be placed on either side of the GGA core.

Notably, we found a remarkable narrowing of linewidths when NMR spectra of ΨRRM2 were recorded in complex with CGGACU RNA compared to the apo protein. Corresponding peaks locate to residues in α1 (BMRB ID 53107), the primary RNA binding interface within pseudo RRMs. This suggests that binding of native target RNA motifs stabilizes α1 in ΨRRM2.

A and B: RBNS sequence logos for RRM1 (1–72) (A) and RRM1 with linker (1-105) (B). C: ^{1}H-^{15}N-HSQCs showing the titration of RRM1 with poly-C RNA. Measurements were performed at 295K in buffer containing 10 mM NaPi, 80 mM NaCl, 4 mM DTT, 0.02% NaN_3 and 7% (v/v) D_2O, pH 6.5. D: CSP-plot based on C for the 4-fold molar excess of poly-C RNA. RRM1 secondary structure elements and RNPs (red) are highlighted at the top. E: AF3 model of RRM1 with CCCC RNA. Only two central bound nucleotides are shown. Secondary structure elements and key residues are labeled. The R71-RNA interaction is highlighted on the right. F: Median RMSD of RNAs second and third residue from MD simulations of SRSF6 RRM1 with RNA tetramers. RMSDs were measured after alignment of protein and RNA heavy atoms. Mean RMSDs over time for each replica are displayed as black crosses. The corresponding RNA sequence used during simulations is given at the bottom. Bar colors correspond to the nucleotide. The median and error are calculated from five replicas. G: RBNS sequence logos for ΨRRM2 with partial linker (86–188). H: CSP-plots of ΨRRM2 in complex with 4-fold molar excess of CGGACU and CGAACU RNA, respectively (Supplementary Fig. 12). I: FP data of ΨRRM2 with CGGACU and CGAACU RNAs. Data points are averages from triplicate measurements with the standard deviation as error. Fits have been performed on the averaged points. J: RBNS sequence logos for RRM1-ΨRRM2.

Binding of CGGACU RNA rigidifies the flexible α-helix 1 in ΨRRM2

To understand the interaction of SRSF6 ΨRRM2 (86–188) with RNA at an atomistic level, we solved its crystal structure in complex with CGGACU at a resolution of 2.3 Å (Fig. 3A-C). As for the apo structure, only residues 106–185 – corresponding to the ΨRRM2 domain boundaries – showed electron density indicating that the interdomain linker remains flexible and does not undergo specific contacts to the RNA 6-mer. We found electron density for the central GGA motif (G2/G3/A4 of the 6mer) of the RNA allowing us to unambiguously model the bases, sugars and phosphate backbone (Supplementary Fig. 4B). The interface reveals the precise orientation and specific contacts between the ΨRRM2 and the three

nucleotides. In detail, residues W123, Q124, K127, D128, R131, A139 and D140 directly contact the RNA supporting the crucial role of α1 in complex formation. The three bases of GGA are forming a pocket-like fold centered on W123, which interacts with G2 through pi-stacking. An additional hydrogen (H) bond is formed between W123-NεHε and N7 of A4. While there is no direct contact from W123 to G3, the close orientation supports a hydrophobic interaction. The adjacent Q124-NεHε forms an H-bond to the O4' of G3 with an additional dipole-dipole interaction between Oε and O2' of A4. Additionally, the side chains of K127 and D128 each form an H-bond to the nucleobases of G2 and G3, respectively, completing the contributions from the conserved ΨRRM2 SWQDLKD heptapeptide motif. Towards the end of α1, R131-NH2 forms an H-bond to N1 of G3. Apart from the residues in α1, A139 and D140 in β2 form additional contacts, both to the Watson-Crick-Franklin face of G2 (Fig. 3D).

In line with our NMR data, a comparison of apo and RNA-bound crystal structures of SRSF6 ΨRRM2 revealed a decrease in b-factors within α1 in the presence of RNA (Supplementary Fig. 14). This further supports the model of a flexible helix present in the unbound state and a rigidification upon RNA binding into a more compact RNP, which might be a prerequisite for selective RNA recognition by SRSF6.

Our crystal structure was further corroborated by 1D-imino proton spectra of ΨRRM2 in complex with CGGACU RNA, which reveal the formation of an H-bonded imino proton (~ 10.7 ppm) upon addition of protein to RNA (Fig. 3E). The hydrogen bonds from either G2 and/or G3 are also present in the crystal. Here N1H of G2 is stabilized by the carbonyl group of A139 and N1H of G3 is in contact with both the carboxyl group of D128 and the guanidine group of R131, which together underpins the native nature of the crystal structure.

A comparison with the ΨRRM of SRSF1 in complex with UGAAGGAC RNA (PDB ID 2M8D) shows that both proteins directly interact with the "GGA" trinucleotide (Supplementary Fig. 15). However, while the two guanines are oriented very similarly in both structures, the adenine shows a clear difference in its position, reflected in a change in direct interactions (Supplementary Fig. 15B/C). While in SRSF6 the adenine is only contacted by W123 and Q124, it interacts with five amino acids in SRSF1, S133, W134, Q135, D136 and H183 (Supplementary Fig. 15D/E), suggesting that SRSF1 has potentially a higher affinity towards GGA than SRSF6.

We next mutated the residues in the SWQLDKD motif in α1 to verify their contributions to RNA binding using FP measurements (Fig. 3F) and structural data (Supplementary Fig. 16). We found that W123 is the most critical amino acid as its mutation to an alanine (W123A) completely abolished RNA binding, while the domain fold is not affected compared to the WT (Supplementary Fig. 16A/B). Likewise, the S122A mutant severely affected RNA binding, however its protein fold was affected as seen from an overlay of ^{1}H-^{15}N-HSQC spectra (Supplementary Fig. 16A). Both the Q124A and K127A mutants showed reduced binding, which is in line with both residues' side chains interacting with the RNA through H-bonds. Interestingly, the D128A mutant showed only slightly reduced affinity, despite the presence of three hydrogen bonds between D128 and G3. In agreement with its orientation away from the RNA, the D125A mutant had no effect on binding.

To uncover the structural determinants that favor the GGA motif, we carried out all-atom MD simulations of ΨRRM2 complexed with either GGA or GAA triplets, with 5 runs of 5 µs production time per system. With GGA the first nucleotide (G2) remained bound throughout the trajectories (Fig. 3G, Supplementary Fig. 17). G2 stacked against W123, matching the crystallographic data (Supplementary Fig. 18A) and reinforced by two H-bonds between the backbone of A139 and the O6 and N1H atom of G2, (Fig. 3D, Supplementary Fig. 18B). In all simulations A4 showed large fluctuations (mean RMSD of 10.99 Å), indicating its subordinate role for stable binding.

The simulations of ΨRRM2 GAA showed significantly reduced stability. After ~ 13 ns the RNA departed from its initial pose. The second nucleotide (A3) formed transient contacts with arginine residues in the linker region rather than with the canonical binding site. G2 rotated and stacked to the opposite side of W123, while also interacting with R131. These interactions were weak, and the RNA exhibited dynamic behavior throughout all replicas.

From the RMSD time series, we extracted dissociation rates k_{off} for each triplet. The GGA complex exhibited a dissociation rate approximately 30-fold lower than that of GAA (Fig. 3G), corresponding to a substantially higher complex stability. This quantitative difference is in full agreement with the experimental binding data derived from FP (Fig. 2I).

Altogether, the crystal structure with NMR measurements, the screen of RNA binding mutants and the MD simulations collectively reveal a specific binding interface of SRSF6 ΨRRM2 with GGA, reasoned by mutual stabilization sequence-specific contacts, while a complex with a single GAA RNA motif appears remarkably less stable.

A: Crystal structure of ΨRRM2 with CGGACU. Only the central “GGA” showed defined electron density (Supplementary Fig. 4B). B and C: Zoom-ins on RNA-protein interaction of ΨRRM2 from crystal structure. D: Contact map of ΨRRM2 interacting with CGGACU RNA. E: ^{1}H-imino proton spectrum of CGGACU RNA upon titration of ΨRRM2. F: Binding affinities of ΨRRM2 WT and mutants obtained from FP assays with CGGACU RNA. The red dotted line marks the affinity of the WT for comparison. For all data points and fits see Supplementary Fig. 16C. G: Mean RMSD over replicas of SRSF6 ΨRRM2 simulations with GGA and GAA fragments. Error bars show the standard error of the mean. Bar colors correspond to the nucleotide. For details, see Supplementary Fig. 17.

Binding of specific RNA targets requires combined RNA-recognition by both RRMs

Our data show that both RRMs bind to RNA individually, but in tandem they engage with minimal composite motifs centered around GGA with CU flanked on either side (Fig. 2J). This suggests that the flexible arrangement of RRM1 and ΨRRM2 with a correctly placed linker contributes to the preferred recognition of such bi-partite binding motifs. To understand the spatial arrangement of the SRSF6 RRMs in the tandem context, we next used small-angle X-ray scattering (SAXS) data of the tandem RRM to filter

conformations from an *in silico* created ensemble (Supplementary Fig. 19) of single RRM structures connected by their native linker (Fig. 4A and B). While the selected rigid body models generally do not imply stably interacting RRMs, they suggest them to be in close proximity and favored orientation to each other (Supplementary Table 20). This finding is in agreement with our previous NMR data indicating transient contacts[21] and suggests a conformational plasticity driven by the interdomain linker, needed for concerted RNA engagement[11,73].

Next, we probed combined RNA binding by SRSF6 tandem RRMs with respect to i) individual domain contributions, ii) the role of motif order, iii) GGA vs. GAA in the center, and iv) the effect of spatial constrains and the impact of nucleotide spacing. To this end, we carried out FP assays in combination with EMSAs using RNAs that contained a combination of RRM1 (CCAC/CA/CU)- and ΨRRM2 (GGA/GAA)-preferred binding motifs (Fig. 4C and Supplementary Fig. 20) spaced by U nucleotides. The FP suggest that the two RRMs can bind simultaneously and in close proximity to individual motifs with short or even no spacing between them (Supplementary Fig. 20A/B, compare ΨRRM2 and tandem RRM binding to CGGACU RNA with a K_D of 1.99 µM and 1.1 µM respectively). With no spacing, the tandem RRMs bind with identical affinity of 0.31 µM to GGA-CCAC or the inversed CCAC-GGA RNAs (Supplementary Fig. 20C/D), indicating no preference for either motif order, while a GAA core results in an non-fittable curve for both the GAA-CCAC and CCAC-GAA RNAs indicating weak binding (Supplementary Fig. 20C). In contrast, single-RRM-motifs e.g.,poly-C/CCAC or GGA-RNAs are bound with lower affinity by the tandem RRMs (Supplementary Fig. 20E/F). Together, these data confirm that both RRMs contribute to binding to combined GGA-CCAC/CCAC-GGA sequences with flexible spacing between the “CCAC” and “GGA” motifs. For RRM1 in the tandem context, the RNA-motif can be further narrowed down to a CA-dinucleotide being efficiently recognized in combination with a close-by GGA for ΨRRM2 recognition (Fig. 4C, Supplementary Fig. 20G). In line with previous data, the substitution of the GGA to a GAA in the 12-mer context resulted in incompletely fittable FP curves caused by low affinity (Supplementary Fig. 20G).

To understand how the tandem RRM interact with RNA we measured SEC-SAXS data for RRM1- ΨRRM2 with an equimolar amount of a combined motif interspaced by 5 U nucleotides (CA5GGA) (Fig. 4D/E) and selected the ten best fitting *in silico* models (Fig. 4E/F). The D_{max} and R_g values for the SRSF6 tandem RRM apo ensemble were very similar over the best ten selected models with values around 9 and 2.5 nm, respectively (Supplementary Table 20), indicating a preferred orientation of the RRMs to each other (Fig. 4F). Importantly, the RNPs of RRM1 and α-helix 1 of ΨRRM2, are positioned in direct proximity (Fig. 4G), potentially allowing RNA-recognition of spatially close RNA motifs as well as cooperativity between RRMs to increase affinity and selectivity.

To substantiate combined and simultaneous binding at residue resolution, we performed NMR titrations of the tandem RRMs with three different RNAs: CCAC (RRM1-only motif), CCAC2GGA (a spaced combined motif), and CGGACU (the RRM2-driven minimal composite motif). We mapped individual CSPs to their involved SRSF6 sites (Supplementary Fig. 21).

The CCAC2GGA RNA caused CSPs confirming binding of both RRMs. The CSP pattern and magnitude for RRM2 is comparable between the CGGACU and CCAC2GGA RNA indicating a similar affinity between these two RNAs and the same mode of interaction. In line with the RBNS data, also the CGGACU RNA caused CSPs in ΨRRM2 and RRM1 indicating an interaction with both RRMs, supporting the identification as the minimal composite motif.

Altogether, the combination of SEC-SAXS with *in silico* modeling yields the first experimentally supported structural ensembles of SRSF6 tandem RRM, both in its apo state and in an RNA-bound form. They confirm that specific RNA targets require the contribution of both RRMs of SRSF6 and that GGA is preferred over GAA in this context. Minimal composite RNA motifs as short as 6 nucleotides are recognized and bound by the two RRMs in agreement with the sequences found with RBNS e.g., CGGACU (Fig. 2J; Supplementary Fig. 6). For RRM1 the CSP magnitude is clearly elevated for the CCAC2GGA RNA compared to CCAC. This suggests that the ΨRRM2 of SRSF6 is the main driving force for RNA binding, but it is supported by the cooperative binding of RRM1 to combined motifs in pre-mRNA targets.

A: Crysol fits of best 10 selected models (out of 5000) to SEC-SAXS data of the tandem RRM. Chi^2-values were between 2.26 and 2.84. B: Best 10 apo models based on filtering against SEC-SAXS data for the tandem RRM. Models are structurally aligned on RRM1 and shown in cartoon representation. C: FP data of SRSF6 tandem RRM (1-188) with different RNAs. Experiments were performed as triplicates. D: Selected frames from SEC-SAXS runs of the tandem RRM alone or in complex with CA5GGA RNA. Intensity is normalized for easier comparison. E: Crysol fits of best 10 (out of 5000) selected models to SEC-SAXS data of the tandem RRM in complex with CA5GGA RNA. Chi^2-values were between 2.32 and 2.72. F: Best 10 models for the tandem RRM in complex with CA5GGA RNA as derived from filtering against SEC-SAXS data. Models are structurally aligned on RRM1. The RNA is depicted in grey. Both protein and RNA are shown in cartoon representation. G: Detailed view of m1170 (from E/F) in semi-transparent cartoon representation with selected RNA-interacting residues depicted as sticks and labeled. Bound nucleotides are also depicted as sticks with the side chain present, while the rest of the RNA is shown as lines. For further details, see Supplementary Table 20.

The SRSF6 linker provides charge-driven benefits to RNA binding

We next investigated the influence of interdomain linker length and residues on RNA binding. For this, we generated four constructs with different linker lengths starting at residues 72, 86, 96 and 106 followed by the ΨRRM2 and tested them by FP for binding to the CGGACU RNA (Supplementary Fig. 22A). Interestingly, the 86–188 construct showed the highest affinity (K_D=1.99 μM), followed by 72–188 (K_D=3.4 μM), 96–188 (K_D=4.9 μM) and 106–188 (K_D=11.8 μM). This suggests that the linker contributes to RNA binding. To test this further, we used a linker peptide (residues 73–105) and performed EMSAs with two RNAs of different length (Supplementary Fig. 22B). As expected, the linker had no intrinsic sequence specificity and interacted with either RNA but showed an increased affinity for the longer RNA.

This hints to an avidity effect likely driven by charge-interaction of the arginine-residues with the negatively charged RNA-backbone. To test this, we introduced charge-reverting mutations (8 arginines, R-E) in the linker in the context of the tandem construct and used them in FP assays with two different RNAs (Supplementary Fig. 23). While the mutations R71E, R74E R75E, R94E R95E and R99E showed a pronounced loss of RNA binding, mutation of R77E and R85E, which are located towards the middle of the linker, had no effect. This was similar for both tested 12-mer RNAs (CCACGGA/GGACCAC) that contained single RNA motifs for RRM1 (CCAC) and ΨRRM2 (GGA, Supplementary Fig. 23B/C) in a poly-U context (Supplementary Table 7). Additionally, deletion of the entire linker in the tandem RRM (del73-105, Supplementary Fig. 23B-C) resulted in a near-complete loss of affinity towards tested target RNAs. These data suggest that besides positioning of the two RRMs on target RNAs, the linker also provides charge-driven benefits to RNA binding, with arginines in close proximity to the RRMs being most critical (Supplementary Fig. 23A).

SRSF6 binding to clustered GAA triplets rescues lower binding affinity

Our data so far established that RRM1 binds preferably to CA-rich RNAs and ΨRRM2 to GGA within minimal composite motifs that resemble the binding motifs in exons regulated by SRSF6. However, they do not explain why all *in vivo* studies reported the binding of SRSF6 to GAA triplets. In our systematic structure-supported *in vitro* data, GAA showed consistently significantly weaker binding (Fig. 2H/I, Supplementary Figs. 12 and 20). To address this discrepancy, we selected two exons of the *MDM4* transcript that are both bound by SRSF6 according to our iCLIP data[13]. However, while inclusion of exon 6 is reduced upon SRSF6 depletion, exon 7 is not. In line with being a regulated SRSF6 target, exon 6 contains two minimal composite motifs centered around GGA (Supplementary Fig. 24A). In contrast, the non-regulated exon 7 contains five GAA motifs (Fig. 5A and Supplementary Fig. 24B). We split exon 7 RNA in two parts of equal length - the first half contains four GAA motifs and the second half only one GAA motif - and compared SRSF6 binding affinities (Fig. 5B, Supplementary Fig. 24C) using dual emission spectral shift assays. Interestingly, the tandem RRM showed a 4.5-fold higher binding affinity for the first RNA half with four GAA motifs (MDM4_E7_1; K_D=590 nM) compared to the second half with a single GAA motif (MDM4_E7_6, K_D=2,660 nM; Fig. 5B). This indicates that SRSF6 prefers clusters of GAA motifs over single GAA motifs as previously suggested[18]. To corroborate this idea, we further split the first half in two parts containing either one (MDM4_E7_2) or three GAA motifs (MDM4_E7_3). These RNAs confirmed the clear preference of SRSF6 for GAA clusters, and it indicates that avidity of SRSF6 drives the increase in RNA affinity.

Strikingly, binding affinity increases nearly two-fold when we replaced a single GAA motif by GGA (MDM4_E7_4). In contrast, exchanging multiple GAA to GGA motifs showed no effect (see RNAs 3 and 5). This indicates that repeated GAA triplets can compensate for the lower preference of single GAA compared to GGA.

To test this further we used RNAs containing CA- or poly-C repeats for RRM1 binding and GGA- or GAA-repeats for ΨRRM2 binding (Fig. 5C, Supplementary Fig. 24D). This revealed that the SRSF6 tandem RRMs bind equally well to GAA and GGA when they occur in repeats and are flanked by CA. Moreover, we noticed a preference of RRM1 for poly-C over CA-repeats (compare C^6GAA^2 and CA^3GAA^2), and in this context GAA repeats even outcompete GGA (compare C^6GGA^2 and C^6GAA^2). Of note however, we found C^6GGA^2 to form secondary structure as observed by NMR (Supplementary Fig. 25), which might impair binding by SRSF6 and in fact may represent a still underexplored parameter in RNA selectivity. Altogether, our data suggests that SRSF6 binds equally well to combined motifs in regulated exons and repeated GAA motifs in non-regulated exons. Given that non-regulated exons with GAA clusters outnumber SRSF6-regulated exons by far, GAA appears transcriptome-wide as preferred binding motif of the protein. This is in stark contrast to SRSF1 that always binds to GGA triplets - regulated or non-regulated exons, *in vitro* or *in vivo* alike[12] - and suggests different modes of RNA selectivity between SRSF1 and 6, with the latter distinguishing regulated from non-regulated exons.

A: Sequence of *MDM4* exon 7 and the six respective RNA variants generated from this. For more details, see Supplementary Fig. 24B. B: Barplot of K_D values derived from dual emission spectral shift assay (see Supplementary Fig. 24C) with SRSF6 tandem RRM and *MDM4* exon 7 variants. C: Barplot of K_D values derived from FP assays (see Supplementary Fig. 24D) with SRSF6 tandem RRM and four different designed RNAs as indicated below the x-axis. Note that for C^6GGA^2 we observed secondary structure formation by NMR (see scheme on top and Supplementary Fig. 25).

The SRSF6 RS domain acts as a selectivity filter for GGA vs. GAA

Our data explain why GAA triplets appear as preferred motifs in SRSF6 iCLIP data, however, they fail to explain why GGA is not found in them (Fig. 1A). One possibility could be that SRSF6 is outcompeted by SRSF1 for binding to GGA. To test this directly, we compared the binding affinities of *E. coli*-derived SRSF6 and SRSF1 tandem RRMs to GGACCAC or GAACCAC RNAs using FP (Fig. 6A). This revealed a clear preference of both proteins for GGA over GAA, but SRSF1 showed indeed a 2-fold higher affinity for GGA than SRSF6 (K_D=0.16 vs. 0.33 µM), while SRSF6 showed a slightly higher affinity for GAA than SRSF1 (K_D=11.2 vs. 7.6 µM).

To test whether the preference of both proteins for GGA holds true in the native setting, we established the expression and purification of human FL, endogenously modified SRSF1 and SRSF6 from HEK293F cells (Fig. 6B-D). RNA binding studies were performed using FP assays (Fig. 6A, Supplementary Fig. 26) or dual emission spectral shift measurements (Supplementary Fig. 27). We first quantified and compared the affinities of FL-SRSF6 (1-344) and SRSF1 (1-248) to GGA- or GAA-CCAC RNAs with the respective tandem RRMs (Fig. 6A). Both FL-proteins still preferred GGA over GAA (Fig. 6A). However, the affinities of FL-SRSF1 and SRSF6 for GGA-CCAC were 20- and 14-fold lower than those of the respective tandem RRMs lacking the RS domain. This was in stark contrast to GAACCAC RNA, which was bound

with low affinity by SRSF1 regardless whether or not the phosphorylated RS domain was present (11.2 vs. 10.0 μM). The affinity of FL-SRSF6 for GAACCAC decreased by ~ 2-fold compared to the tandem RRM alone (7.6–12.6 μM, Fig. 6D), which suggests an intricate interplay of relative affinities of GGA/GAA, SRSF protein and respective RS type and length and motif distinction.

To confirm this hypothesis further, we repeated the experiments with four additional RNAs (Supplementary Fig. 27). Interestingly, for CGGACU RNA the affinity of FL-SRSF6 (3.5 μM) was only 3.5-fold lower than that of the SRSF6 tandem RRM, indicating that the inhibitory effect of the RS domain on the minimal composite motif was less pronounced here. In contrast, GGA/GAA-repeat containing RNAs (CA^3GGA^2 and CA^3GAA^2) showed again a 6-20-fold decrease in K_D values when the RS domain is present. Notably, we observed near-identical binding of FL-SRSF1 and SRSF6 to all RNAs that we compared (Supplementary Fig. 27). Altogether, these data indicate that the phosphorylated RS domains markedly reduce the affinities of SRSF1 and SRSF6 to RNA, but preserve their preferences for GGA over GAA. The reduction is much less pronounced on regulatory sites and likely serves as a selectivity filter amplifying the differences in affinity for natural target and non-target RNAs. It also equalizes the binding affinity of SRSF1 and SRSF6 to GGA, suggesting that the capacity of SRSF1 to outcompete SRSF6 at GGA sites is not due to higher affinity.

The RRM1 of SRSF6 is critical for the splicing of target exons containing minimal composite motifs

Finally, we tested the contributions of individual SRSF6 domains to splicing of target exons in HEK293F cells. For this we established a SRSF6 knock-out HEK293F cell line ($SRSF6^{KO}$) and functionally complemented it with GFP-tagged SRSF6 variants inducibly expressed from a Tet-On promoter (Fig. 6B-E, Supplementary Table 17). This system permits endogenous expression levels of SRSF6 variants upon Dox induction. We introduced point mutations into the RRMs that completely abolished RNA binding in our NMR experiments (Supplementary Figs. 10 and 16). In addition, we generated a deletion mutant (Δ189–344) where the RS domain is missing (Fig. 6F).

We selected two SRSF6-bound and regulated exons from our RNA-seq/iCLIP dataset – exon 6 of *hMDM4*, and exon 4 of *hCLK1* (Fig. 6G, Supplementary Fig. 28) [13] – and analyzed changes in exon inclusion by semi-quantitative splice PCRs. *hMDM4* exon 6 contains two GGA embedding motifs (CAGGAUC and AUGGAU), and its inclusion was reduced in $SRSF6^{KO}$ cells as expected (Fig. 6G). Complementing the KO cells with $SRSF6^{WT}$ restored exon 6 inclusion (Fig. 6G). In contrast, the RRM1 Y5/Y37/F39A mutant and the double RRM1-ΨRRM2 mutant were unable to restore exon 6 inclusion, indicating that binding of the RRM1 domain is critical for binding and splicing. In line with its role in recruiting the spliceosome, the RS domain deletion mutant also failed to restore exon 6 inclusion. Surprisingly, the ΨRRM2 W123A mutant rescued exon inclusion similar to WT SRSF6, suggesting that the W123A mutation alone is not able to abolish RNA binding *in vivo* and binding to the composite motifs in exon 6 and recruitment of the spliceosome is maintained.

In the case of *hCLK1* exon 4, SRSF6 binds close to the 3'-splice site[13] and normally suppresses inclusion of exon 4. In line with this suppressive role, exon 4 is more included in SRSF6^{KO} cells and re-addition of FL *SRSF6*^{WT} led to exon 4 skipping. Similar to *hMDM4*, complementation with the RRM1 mutant, the double RRM1-ΨRRM2 mutant and the RS deletion mutant failed to rescue exon 4 inclusion levels, while the ΨRRM2 single mutant was able to rescue the KO (Supplementary Fig. 28). Nevertheless, these data indicate that RRM1 and the RS domain of SRSF6 are critical for the splicing of target exons.

Interestingly, SRSF1 levels were increased in SRSF6^{KO} cells, suggesting reciprocal cross-regulation between SRSF1 and SRSF6. To test whether elevated SRSF1 levels influence the splicing of SRSF6 target exons, we overexpressed GFP-tagged SRSF1 in HEK293F wild-type cells and analyzed splicing as described above (Fig. 6G and Supplementary Fig. 28). Increased SRSF1 expression reduced inclusion of *MDM4* exon 6 even more strongly than loss of SRSF6 (Fig. 6G). A similar effect was observed for *CLK1* exon 4 (Supplementary Fig. 28).

These findings indicate that SRSF1 and SRSF6 act antagonistically on regulated exons containing composite GGA motifs and that their expression levels are tightly controlled through feedback regulation. Notably, increasing SRSF1 levels in the presence of endogenous SRSF6 had a stronger effect on exon inclusion than deletion of SRSF6 alone, suggesting that SRSF1 can effectively outcompete SRSF6 for binding to composite GGA motifs *in vivo*.

A: Comparison of SRSF1 and SRSF6 RNA binding selectivity for the GGACCAC and GAACCAC motifs, determined by FP. Tandem RRMs were derived from *E. coli*, FL proteins were purified from HEK293F cells (Supplementary Fig. 26). B: Schematic representation of the doxycycline (Dox)-inducible, stable expression construct for FL SRSF1 and SRSF6. The domain architecture is given for both FL proteins. C: Confocal immunofluorescence images showing the intracellular localization of eGFP-tagged SRSF1 and SRSF6 in nuclear speckles. Nuclear speckle marker SON (red) and nucleus (DAPI, blue). Scale bars = 5 µm. D: SDS-PAGE showing purified FL SRSF1 and SRSF6 from HEK293F cells. E: CRISPR-Cas9 genome editing strategy to knock out (KO) endogenous SRSF6. A single-guide RNA (sgRNA, red) targets the region between the start codon and the Cas9 cut site (scissors) near exon 2 of the *SRSF6* gene. F: Western blot analysis confirming the KO of endogenous SRSF6 in HEK293F cells. SRSF6^{KO} cells were complemented with wild-type SRSF6 or respective RRM1-, RRM2- or RS domain mutants. Protein expression was induced for 48 h and assessed by western blotting using α-SRSF6 antibodies. α-actin antibodies were used as loading control. G: Model of human MDM4 (*hMDM4*) showing exons 4 to 8. Alternative exon 6 is highlighted in blue. Semi-quantitative RT-PCR of *hMDM4* splice isoforms either including (top) or skipping (bottom) alternative exon 6. Bar graph showing the percentage of exon 6 inclusion (blue) based on densitometric analysis of the RT-PCR products (n = 3). Error bars represent standard deviations. The dotted line indicates the inclusion level in parental HEK293F cells.

Discussion

SRSFs regulate a vast number of RNA transcripts and exhibit substantial, but little-understood functional redundancy. At the same time, individual SRSF proteins possess pronounced target selectivity, a property that is particularly critical during alternative splicing. How selective exon regulation can be achieved despite widespread transcriptome-wide RNA binding has remained a central unresolved question[19,74], which is especially relevant within a densely crowded cellular environment, where RBPs continuously compete for partially overlapping RNA motifs. Here, we address this question, focusing on SRSF6, and our findings suggest that functional target selectivity emerges from the local sequence context and requires the flexible and cooperative contribution from both RRM domains (Fig. 7). Previous studies on SRSF6 RNA recognition have revealed an apparent discrepancy between RNA motifs identified *in vitro*, which predominantly enriched CU/GGA-containing sequences, and transcriptome-wide *in vivo* approaches, such as iCLIP, which instead identified GAA repeat motifs[13,18]. Our structural and functional analyzes provide a framework that reconciles these observations and indicates a critical role of the RS linker domain for selectivity.

We here showed that ΨRRM2 drives SRSF6 binding to purine-rich (GA) motifs. We confirmed a strong preference for GGA over GAA, well in line with the previously *in vitro* obtained motifs[10,75,76]. Our complex structure of the SRSF6 ΨRRM2 with RNA demonstrates that the GGA motif is bound via the conserved heptapeptide RNA binding interface in α1, similar to the previously described binding of the SRSF1 ΨRRM to GGA[10]. However, qualitative and quantitative deviations within the contact-network account for a slightly higher affinity of SRSF1 for GGA as compared to the ΨRRM2 of SRSF6. We found enhanced binding to RNA motifs in the SRSF6 tandem RRM context when GGA is embedded or flanked by C(A)-rich stretches. This is due to the binding contributions of RRM1 – a canonical, but extraordinarily compact RRM domain as shown by our high-resolution NMR structure. Similarly to SRSF1[11], SRSF6 RRM1 specifically bound short C(C/A) dinucleotides, which in the tandem context increased affinities. We were able to define a high-affinity minimal composite motif for SRSF6, CGGACU, which represents its intrinsic binding selectivity and is bound through the cooperative engagement of both RRMs–RRM1 and ΨRRM2–supported by the charged interdomain linker. The composite CGGACU motif identified here closely resembles motifs enriched within SRSF6 regulated alternative exons. For this motif, we found a rigidification of the ΨRRM2 α1 region upon RNA binding, suggesting this may contribute to the distinction of constitutive-versus-alternative motifs by SRSF6, and to the discrimination from other SRSFs. At the same time, ensemble modelling supported by SAXS revealed pronounced conformational plasticity of the tandem RRMs, mediated by the interdomain linker, similarly to the binding mechanism of SRSF1[11,77]. This flexibility likely enables recognition of bipartite RNA motifs in either orientation, and across variable spacing constraints, including motifs in very close proximity. Still, our data show that both apo and RNA-bound SRSF6 tandem RRMs adopt preferred conformations suggesting transient interactions of the two domains. The very same has recently been found for SRSF1[78] and underlines the high degree of redundancy between SRSF1 and 6, while differences in the interdomain linker and the RS domain nature may account for the unique splice activities of either.

Rather than defining a rigid sequence code, the composite motif described in this study likely reflects the broader combinatorial RNA binding preferences of SRSF6 across diverse sequence contexts. In this model, ΨRRM2 recognizes GGA motifs, whereas RRM1 functions as a context-selectivity module that engages C/A-rich sequence elements surrounding the GGA core and thereby helps discriminate functional (minimal) composite motifs from generic purine-rich sequences. This cooperativity is critical for SRSF6 activity, as we showed that an RNA binding deficient RRM1 disproportionately affects the splicing of SRSF6 target exons, even when ΨRRM2-mediated binding is retained.

While we found SRSF6 bound single GAA triplets with low affinity compared to GGA, avid or multivalent binding to GAA repeats equalized apparent affinities (Fig. 7). This finding strongly underpins the relevance of motifs derived from iCLIP[13,18] and suggests avidity-driven binding to clustered GAA sites, rather than an intrinsic sequence preference. SRSF6 is a core nuclear speckle protein (Fig. 6C) and might be involved in a process called interstasis, whereby transcripts rich in GAA repeats are selectively retained within nuclear speckles[79,80]. The observed preference of SRSF6 for GAA repeats might be related to RNA retention and compartmentalization, rather than to alternative splicing. This suggests that nuclear organization contributes to transcriptome-wide occupancy, which is usually not considered in *in vitro* studies, and could account for the apparent paradox. Thus, more generally we should interpret iCLIP-derived motif preferences of a given RBP with care, as they do not necessarily represent its select regulatory target sequences, but merely the statistical occupation of RNA motifs. In addition, specifically or uniquely regulated RNA motifs do not necessarily exhibit the highest affinity, underlining the problematic interweaving of specificity and affinity, not only between RNAs and RBPs[81,82].

Discrimination of splicing decisions along with GGA vs. GAA should require different modes of interaction with transcripts. While beyond this study, we envision that e.g., the combined recognition of composite motifs involves RRM1 and ΨRRM2, consistent with a defined multi-domain configuration. In contrast, binding to GAA-rich elements associated with constitutive exons might be largely mediated by ΨRRM2 alone as it requires avidity. At such sites, RRM1 could be free to interact with other partners, and for example promote avidity or exclude competing RBPs. In this model, RRM1 would have a dual function - sequence specific RNA binding or a context-dependent protein interaction platform, marking target sites as either high-affinity motifs for alternative splicing or constitutively processed exons via GAA repeats. The involvement of canonical RRM domains in protein interactions has been found before[83], e.g.,for PABPC1[84], Sxl[85], U2AF65[86], SPF45[87].

SRSF proteins are challenging with regard to structural and biochemical analysis because of their charged, disordered RS domains. We here established a protocol to purify SRSF FL-proteins from human cells, which allowed for a detailed investigation of RNA binding in context of their native protein modifications. Our data show that the phosphorylated RS domain significantly reduced RNA binding. This is most pronounced for suboptimal binding sites but has less severe effects on binding to minimal composite motifs. This suggests that the phosphorylated RS domain acts as a selectivity filter or proofreading module, suppressing promiscuous RNA interactions while allowing stable engagement at functional sites. This aligns with earlier work demonstrating that RS phosphorylation modulates RNA

binding, protein interactions and SR protein activity in cells[88–90]. Thus, the balance between broad occupancy and specific binding might be actively enforced, rather than passively determined by intrinsic RRM affinity and could be spatially modulated in nuclear speckles, where phosphorylation dynamics and protein concentration differ from the nucleoplasm[91,92].

A direct comparison showed that despite similar domain architecture, SRSF6 and SRSF1 exhibit distinct binding and regulatory behaviors. SRSF1 also binds to GGA-centered motifs with high affinity, but here *in vitro* and *in vivo* methods are in complete agreement[10,12]. In contrast, SRSF6 is less competitive at binding to isolated GGA motifs but engages efficiently at composite motifs or GAA motif clusters. This indicates that both SRSFs compete for the binding space, and while SRSF1 dominates at canonical GGA sites, SRSF6 binds only efficiently in the context of combinatorial recognition or of avidity effects[13,18]. Adding to this, the length and phosphorylation state of the RS domain, nuclear speckle residency and RNA structure might further modulate this competition. For example, GGA-rich sequences frequently form secondary structures, including G-quadruplexes, and SRSF1 was found to bind and remodel such elements[8,9], while SRSF6 was not. Possibly, RNA structure selectively favors SRSF1 at GGA-rich sites, thereby indirectly enriching SRSF6 at GAA-rich sites that remain unstructured or at composite motifs *in vivo*. While to be studied in more detail, our systematic binding data are well in support of this idea (Fig. 5C) and provide a plausible explanation for the observed differences in *in vivo* binding landscapes of SRSF1 and 6.

Our data suggest that SRSF6 binding and function also emerges from a dynamic competition and compensatory regulation within the SR protein network. We observed that SRSF1 is upregulated in SRSF6 KO cells, and SRSF1 overexpression phenocopies splicing observed in SRSF6 KO cells. This indicates a functional cross-regulation whereby SRSF1 antagonizes SRSF6. Given that both proteins recognize GGA-containing elements, SRSF1 might occupy composite motifs that are normally bound by SRSF6, but the regulatory outcome is different. This model is consistent with the higher intrinsic affinity of SRSF1 for GGA motifs and suggests that loss of SRSF6 shifts the balance of occupancy toward SRSF1, effectively rewiring splicing decisions without changes in RNA sequence. In this scenario, composite motifs would represent shared but differentially interpreted binding platforms and splicing output.

The SRSF6 protein uses its tandem RRM domain for combined recognition of GGA-centered composite motifs in regulated exons or GAA motifs in non-regulated exons leading to alternative and constitutive splicing decisions, respectively. A higher intrinsic affinity for single GGA motifs is mediated by ΨRRM2 supported by contributions of RRM1 in flanking regions. GAA-repeats found in non-regulated exons can equalize binding to sufficiently high affinity, while the selectivity is provided through a filter function by the RS domain including its potential phosphorylation states. Selectively bound regulated GGA exons are competed for with SRSF1, which shows a higher abundance and mobility in cells providing an explanation for its dominant occupation of GGA sites in splicing decisions, which yet remain the

selectively targeted exons for SRSF6, but with its occupancy hidden among the predominant localization to more abundant GAA-rich constitutively spliced exons.

Declarations

Acknowledgements

We thank C. de Oliveira Freitas Machado for generating the SRSF6 KD RNA-seq data and A. di Liddo and K. Zarnack for the splicing analyzes. We thank all members of the Schlundt and Müller-McNicoll labs for support and discussions and K.Targaczewski for technical support. This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC-3094–533751785 to A.S. and M.M.-M., and through funding for project number 583887849 to A.S. and M.M.-M.). S.M.K. acknowledges support through a Feodor Lynen Fellowship by the Humboldt Foundation. We acknowledge excellent support and infrastructure for NMR measurements at the Frankfurt BMRZ supported by the state of Hesse. We thank S. Kaiser for access to her BioRad gel documentation device and K. Pos for letting us measure at his Spark Tecan plate reader. We thank M. McDowell for access to the NanoTemper Monolith X. We thank S. Knapp and A. Krämer for access and support at the crystallization facility in the Buchmann Institute for Molecular Life Sciences.

Data availability

NMR assignments have been deposited under BMRB accession codes 53130 (RRM1 + poly-C RNA) and 53107 (ΨRRM2 + CGGACU RNA). Structures have been deposited under PDB IDs pdb_00009r53 (RRM1 WT apo), pdb_00009hx5 (ΨRRM2 WT apo), pdb_00009hx7 (ΨRRM2 WT + CGGACU RNA) and pdb_00009hx8 (ΨRRM2 W123A). SEC-SAXS data has been deposited under the SASBDB[93] accession code SASDZP7. All iCLIP data are available at the Gene Expression Omnibus (GEO) under the accession numbers GSE198309 (SRSF6-GFP in HeLa cells), GSE69689 (SRSF1 and SRSF6 in murine P19 cells).

References

1. Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular cell* **76,** 329–345; 10.1016/j.molcel.2019.09.017 (2019).
2. Slišković, I., Eich, H. & Müller-McNicoll, M. Exploring the multifunctionality of SR proteins. *Biochemical Society transactions* **50,** 187–198; 10.1042/BST20210325 (2022).
3. Bradley, R. K. & Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nature reviews. Cancer* **23,** 135–155; 10.1038/s41568-022-00541-7 (2023).
4. Wagner, R. E. & Frye, M. Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease. *BioEssays : news and reviews in molecular, cellular and developmental biology* **43,** e2000242; 10.1002/bies.202000242 (2021).

5. Manley, J. L. & Krainer, A. R. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & development* **24,** 1073–1074; 10.1101/gad.1934910 (2010).
6. Wegener, M. & Müller-McNicoll, M. View from an mRNP: The Roles of SR Proteins in Assembly, Maturation and Turnover. *Advances in experimental medicine and biology* **1203,** 83–112; 10.1007/978-3-030-31434-7_3 (2019).
7. Cléry, A., Boudet, J. & Allain, F. H.-T. Single-stranded nucleic acid recognition: is there a code after all? *Structure* **21,** 4–6; 10.1016/j.str.2012.12.006 (2013).
8. Silva, N. I. U. de *et al.* Unearthing a novel function of SRSF1 in binding and unfolding of RNA G-quadruplexes. *Nucleic Acids Res* **52,** 4676–4690; 10.1093/nar/gkae213 (2024).
9. Silva, N. I. U. de *et al.* Sequence determinants of RNA G-quadruplex unfolding by Arg-rich regions. *FEBS letters*; 10.1002/1873-3468.70274 (2026).
10. Cléry, A. *et al.* Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proceedings of the National Academy of Sciences of the United States of America* **110,** E2802-11; 10.1073/pnas.1303445110 (2013).
11. Cléry, A. *et al.* Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. *Nature communications* **12,** 428; 10.1038/s41467-020-20481-w (2021).
12. Feng, H. *et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Molecular cell* **74,** 1189-1204.e6; 10.1016/j.molcel.2019.02.002 (2019).
13. Oliveira Freitas Machado, C. de *et al.* Poison cassette exon splicing of SRSF6 regulates nuclear speckle dispersal and the response to hypoxia. *Nucleic acids research* **51,** 870–890; 10.1093/nar/gkac1225 (2023).
14. Yang, X. *et al.* SRSF6 regulates alternative splicing of genes involved in DNA damage response and DNA repair in HeLa cells. *Oncology reports* **44,** 1851–1862; 10.3892/or.2020.7750 (2020).
15. Nasrin, F. *et al.* SRSF6 and SRSF1 coordinately enhance the inclusion of human MUSK exon 10 to generate a Wnt-sensitive MuSK isoform. *NAR molecular medicine* **2,** ugaf007; 10.1093/narmme/ugaf007 (2025).
16. Liu, H. X., Zhang, M. & Krainer, A. R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & development* **12,** 1998–2012; 10.1101/gad.12.13.1998 (1998).
17. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499,** 172–177; 10.1038/nature12311 (2013).
18. Alvelos, M. I. *et al.* The RNA-binding profile of the splicing factor SRSF6 in immortalized human pancreatic β-cells. *Life science alliance* **4**; 10.26508/lsa.202000825 (2021).
19. Müller-McNicoll, M. *et al.* SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & development* **30,** 553–566; 10.1101/gad.276477.115 (2016).

20. The UniProt Consortium *et al.* UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* **53,** D609-D617; 10.1093/nar/gkae1010 (2025).
21. von Ehr, J. & Schlundt, A. 1H, 13C, 15N backbone chemical shift assignments and relaxation analysis of the single and tandem RRMs of the human serine-arginine rich splicing factor 6 (SRSF6). *Journal of Magnetic Resonance,* 108079; 10.1016/j.jmr.2026.108079 (2026).
22. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6,** 343–345; 10.1038/nmeth.1318 (2009).
23. von Ehr, J. *et al.* Arid5a uses disordered extensions of its core ARID domain for distinct DNA- and RNA-recognition and gene regulation. *The Journal of biological chemistry* **300,** 107457; 10.1016/j.jbc.2024.107457 (2024).
24. Tants, J.-N., Oberstrass, L., Weigand, J. E. & Schlundt, A. Structure and RNA-binding of the helically extended Roquin CCCH-type zinc finger. *Nucleic Acids Res* **52,** 9838–9853; 10.1093/nar/gkae555 (2024).
25. Dhamotharan, K. *et al.* A core network in the SARS-CoV-2 nucleocapsid NTD mediates structural integrity and selective RNA-binding. *Nature communications* **15,** 10656; 10.1038/s41467-024-55024-0 (2024).
26. Lambert, N. J., Robertson, A. D. & Burge, C. B. RNA Bind-n-Seq: Measuring the Binding Affinity Landscape of RNA-Binding Proteins. *Methods in enzymology* **558,** 465–493; 10.1016/bs.mie.2015.02.007 (2015).
27. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular cell* **70,** 854-867.e9; 10.1016/j.molcel.2018.05.001 (2018).
28. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36,** 2272–2274; 10.1093/bioinformatics/btz921] (2020).
29. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327; 10.1093/bioinformatics/btu830 (2015).
30. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696; 10.1002/prot.20449 (2005).
31. Güntert, P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)* **278,** 353–378; 10.1385/1-59259-809-9:353 (2004).
32. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of biomolecular NMR* **56,** 227–241; 10.1007/s10858-013-9741-y (2013).
33. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *Journal of biomolecular NMR* **62,** 453–471; 10.1007/s10858-015-9924-9 (2015).
34. Cianci, M. *et al.* P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *Journal of synchrotron radiation* **24,** 323–332; 10.1107/S1600577516016465 (2017).

35. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta crystallographica. Section D, Biological crystallography* **67,** 293–302; 10.1107/S0907444911007773 (2011).
36. Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *Acta crystallographica. Section D, Structural biology* **79,** 449–461; 10.1107/S2059798323003595 (2023).
37. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40,** 658–674; 10.1107/S0021889807021206 (2007).
38. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589; 10.1038/s41586-021-03819-2 (2021).
39. Vagin, A. & Teplyakov, A. MOLREP : an Automated Program for Molecular Replacement. *Journal of applied crystallography* **30,** 1022–1025; 10.1107/S0021889897006766 (1997).
40. Bond, P. S. & Cowtan, K. D. ModelCraft: an advanced automated model-building pipeline using Buccaneer. *Acta crystallographica. Section D, Structural biology* **78,** 1090–1098; 10.1107/S2059798322007732 (2022).
41. Bricogne, G. *et al. BUSTER* (Cambridge, United Kingdom: Global Phasing Ltd, 2017).
42. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67,** 355–367; 10.1107/S0907444911001314 (2011).
43. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1,** 213–220; 10.1107/S2052252514009324 (2014).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66,** 486–501; 10.1107/S0907444910007493 (2010).
45. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *Journal of applied crystallography* **48,** 431–443; 10.1107/S160057671500254X (2015).
46. Manalastas-Cantos, K. *et al.* ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of applied crystallography* **54,** 343–355; 10.1107/S1600576720013412 (2021).
47. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34,** 1944–1946; 10.1093/bioinformatics/btx846 (2018).
48. Hopkins, J. B. BioXTAS RAW 2: new developments for a free open-source program for small-angle scattering data reduction and analysis. *Journal of applied crystallography* **57,** 194–208; 10.1107/S1600576723011019 (2024).
49. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630,** 493–500; 10.1038/s41586-024-07487-w (2024).

50. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angewandte Chemie (International ed. in English)* **49,** 1967–1970; 10.1002/anie.200906147 (2010).
51. Schrödinger, L. *The PyMOL Molecular Graphics System ,*
52. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91,** 6269–6271; 10.1021/j100308a038 (1987).
53. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958; 10.1002/prot.22711 (2010).
54. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation* **11,** 3696–3713; 10.1021/acs.jctc.5b00255 (2015).
55. Zgarbová, M. *et al.* Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of chemical theory and computation* **7,** 2886–2902; 10.1021/ct200162x (2011).
56. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25; 10.1016/j.softx.2015.06.001 (2015).
57. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126,** 14101; 10.1063/1.2408420 (2007).
58. Bernetti, M. & Bussi, G. Pressure control using stochastic cell rescaling. *The Journal of chemical physics* **153,** 114107; 10.1063/5.0020514 (2020).
59. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *The Journal of chemical physics* **98,** 10089–10092; 10.1063/1.464397 (1993).
60. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103,** 8577–8593; 10.1063/1.470117 (1995).
61. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472; 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H (1997).
62. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14,** 33-8, 27-8; 10.1016/0263-7855(96)00018-5 (1996).
63. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference* (SciPy2016), pp. 98–105.
64. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32,** 2319–2327; 10.1002/jcc.21787 (2011).
65. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585,** 357–362; 10.1038/s41586-020-2649-2 (2020).

66. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17,** 261–272; 10.1038/s41592-019-0686-2 (2020).
67. The pandas development team. *pandas-dev/pandas: Pandas* (Zenodo, 2026).
68. The Matplotlib Development Team. *Matplotlib: Visualization with Python* (Zenodo, 2026).
69. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21; 10.1093/bioinformatics/bts635 (2013).
70. Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E5593-601; 10.1073/pnas.1419161111 (2014).
71. Okuda, E. K. *et al.* Rapid depletion and super-resolution microscopy reveal dual roles of SRSF5 in coordinating nuclear speckle-paraspeckle crosstalk during cellular stress. *Nucleic Acids Res* **53**; 10.1093/nar/gkaf713 (2025).
72. Schneider, T. *et al.* Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nature communications* **10,** 2266; 10.1038/s41467-019-09769-8 (2019).
73. Cho, S. *et al.* The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs. *Nucleic Acids Res* **39,** 9413–9421; 10.1093/nar/gkr663 (2011).
74. Pandit, S. *et al.* Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell* **50,** 223–235; 10.1016/j.molcel.2013.03.001 (2013).
75. Änkö, M.-L. Regulation of gene expression programmes by serine-arginine rich splicing factors. *Seminars in cell & developmental biology* **32,** 11–21; 10.1016/j.semcdb.2014.03.011 (2014).
76. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research* **31,** 3568–3571; 10.1093/nar/gkg616 (2003).
77. Silva, N. I. U. de, Fargason, T., Zhang, Z., Wang, T. & Zhang, J. Inter-domain Flexibility of Human Ser/Arg-Rich Splicing Factor 1 Allows Variable Spacer Length in Cognate RNA's Bipartite Motifs. *Biochemistry* **61,** 2922–2932; 10.1021/acs.biochem.2c00565 (2022).
78. Nguyen, C. K. X. *et al.* Characterization of flexible RNA binding by tandem RNA recognition motifs through integrative ensemble modelling. *Nucleic Acids Res* **54**; 10.1093/nar/gkag269 (2026).
79. Müller-McNicoll, M. Getting sticky: How nuclear speckles tune the condensation-prone proteome. *Molecular cell* **85,** 4116–4118; 10.1016/j.molcel.2025.10.029 (2025).
80. Faraway, R. *et al.* Collective homeostasis of condensation-prone proteins via their mRNAs. *Nature* **647,** 798–808; 10.1038/s41586-025-09568-w (2025).
81. Webster, M. W., Stowell, J. A. & Passmore, L. A. RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *eLife* **8**; 10.7554/eLife.40670 (2019).
82. Liberles, D. A. & Meyer, M. M. How Binding Affinity and Binding Specificity Map to Sequence Space. *Journal of molecular evolution* **94,** 279–285; 10.1007/s00239-026-10308-5 (2026).
83. Cléry, A., Blatter, M. & Allain, F. H.-T. RNA recognition motifs: boring? Not quite. *Current opinion in structural biology* **18,** 290–298; 10.1016/j.sbi.2008.04.002 (2008).

84. Sagae, T. *et al.* Paip2A inhibits translation by competitively binding to the RNA recognition motifs of PABPC1 and promoting its dissociation from the poly(A) tail. *The Journal of biological chemistry* **298,** 101844; 10.1016/j.jbc.2022.101844 (2022).
85. Hennig, J. *et al.* Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature* **515,** 287–290; 10.1038/nature13693 (2014).
86. Voith von Voithenberg, L. *et al.* Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E7169-E7175; 10.1073/pnas.1605873113 (2016).
87. Corsini, L. *et al.* U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nature structural & molecular biology* **14,** 620–629; 10.1038/nsmb1260 (2007).
88. Wu, J. Y. & Maniatis, T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75,** 1061–1070; 10.1016/0092-8674(93)90316-I (1993).
89. Tacke, R., Chen, Y. & Manley, J. L. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 1148–1153; 10.1073/pnas.94.4.1148 (1997).
90. Xiao, S. H. & Manley, J. L. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & development* **11,** 334–344; 10.1101/gad.11.3.334 (1997).
91. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harbor perspectives in biology* **3**; 10.1101/cshperspect.a000646 (2011).
92. Lázaro, B. *et al.* Nuclear speckle dynamics are controlled by polyphosphate inhibition of CLK proteins. *Nucleic Acids Res* **54**; 10.1093/nar/gkag309 (2026).
93. Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein science : a publication of the Protein Society* **29,** 66–75; 10.1002/pro.3731 (2020).

Figures

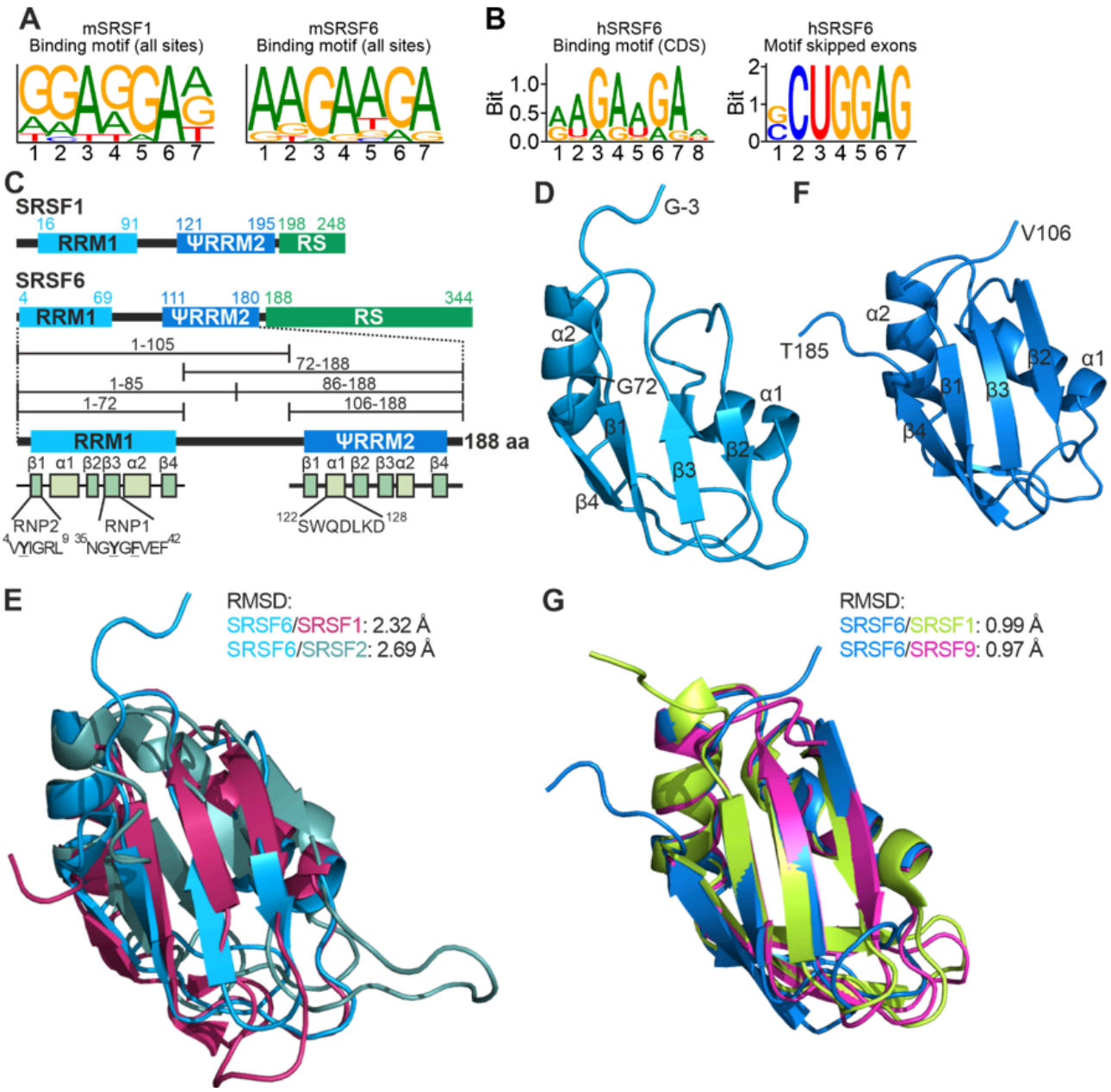


Figure 1

In vivo binding motifs, constructs and structures of SRSF6.

A: Comparison of *in vivo* binding motifs of mouse SRSF1 (left) and SRSF6 (right) considering all sites using iCLIP data from Müller-McNicoll et al[19]. B: Comparison of *in vivo* binding motifs of human SRSF6 considering all exons (left) or only regulated exons (right: skipped after SRSF6 depletion) integrating RNA-seq data and iCLIP data from Oliveira Freitas Machado et al[13]. C: Domain architecture of SRSF1 and SRSF6. SRSF6 constructs used in this study are shown with respective secondary structures for each domain[21]. Domain boundaries are given by amino acid residues. D: Representative NMR structure of SRSF6 RRM1 with secondary structure elements labeled. E: SRSF6 RRM1 structure from D aligned and

overlaid with two other SRSF RRM1 structures. The RMSD of each structure to SRSF6 RRM1 is given. F: Crystal structure of SRSF6 ΨRRM2 with secondary structure elements labeled. G: SRSF6 ΨRRM2 structure from F aligned and overlaid with two other SRSF ΨRRM2 domain structures. The RMSD of each structure to SRSF6 ΨRRM2 is given. The PDB IDs for each structure are given in Supplementary Figure 2. For the full ensemble of RRM1 and electron density map of ΨRRM2 see Supplementary Figure 4.

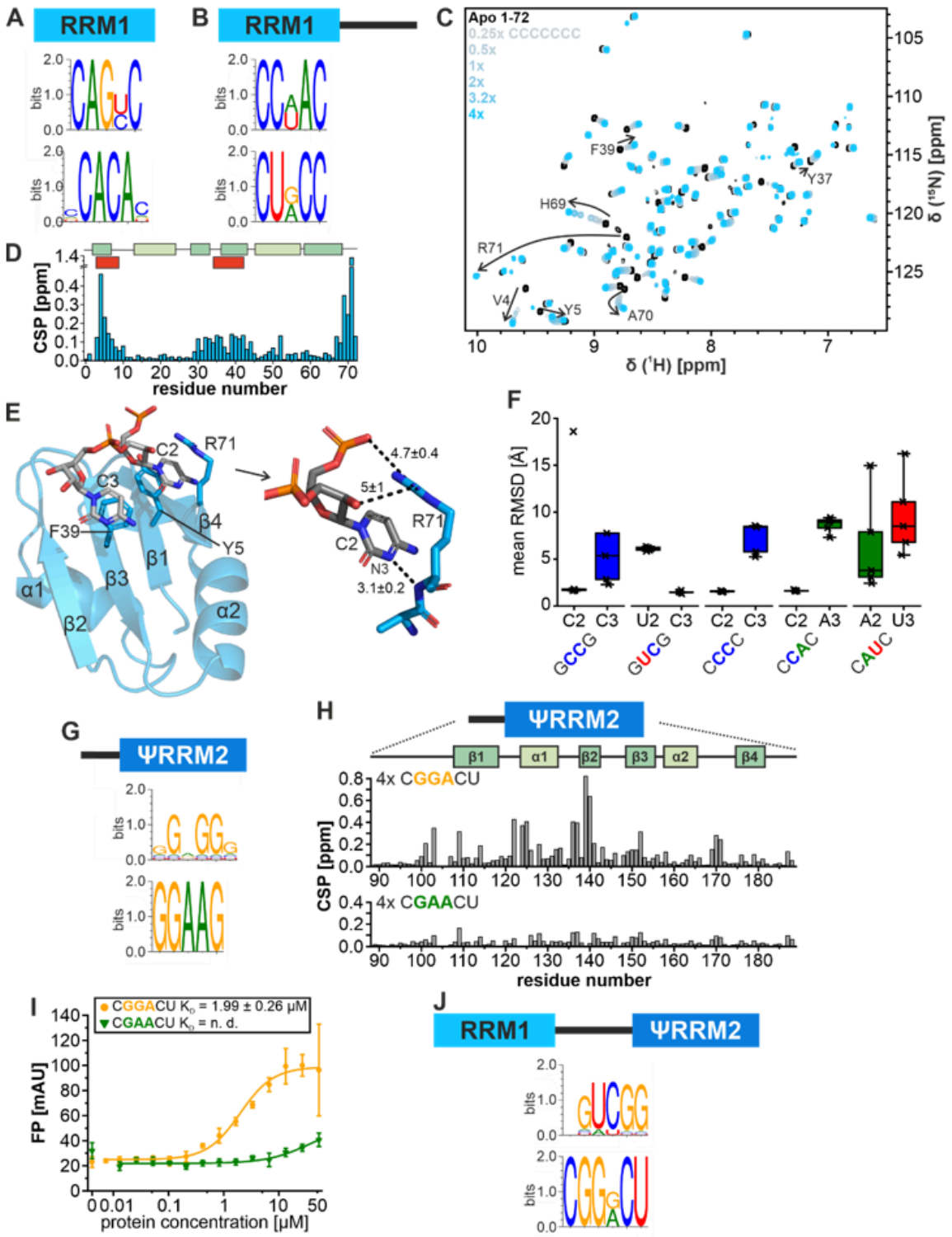


Figure 2

RNA binding by single RRMs.

A and B: RBNS sequence logos for RRM1 (1-72) (A) and RRM1 with linker (1-105) (B). C: ^{1}H-^{15}N-HSQCs showing the titration of RRM1 with poly-C RNA. Measurements were performed at 295K in buffer containing 10 mM NaPi, 80 mM NaCl, 4 mM DTT, 0.02 % NaN_3 and 7 % (v/v) D_2O, pH 6.5. D: CSP-plot based on C for the 4-fold molar excess of poly-C RNA. RRM1 secondary structure elements and RNPs (red) are highlighted at the top. E: AF3 model of RRM1 with CCCC RNA. Only two central bound nucleotides are shown. Secondary structure elements and key residues are labeled. The R71-RNA interaction is highlighted on the right. F: Median RMSD of RNAs second and third residue from MD simulations of SRSF6 RRM1 with RNA tetramers. RMSDs were measured after alignment of protein and RNA heavy atoms. Mean RMSDs over time for each replica are displayed as black crosses. The corresponding RNA sequence used during simulations is given at the bottom. Bar colors correspond to the nucleotide. The median and error are calculated from five replicas. G: RBNS sequence logos for ΨRRM2 with partial linker (86-188). H: CSP-plots of ΨRRM2 in complex with 4-fold molar excess of CGGACU and CGAACU RNA, respectively (Supplementary Figure 12). I: FP data of ΨRRM2 with CGGACU and CGAACU RNAs. Data points are averages from triplicate measurements with the standard deviation as error. Fits have been performed on the averaged points. J: RBNS sequence logos for RRM1-ΨRRM2.

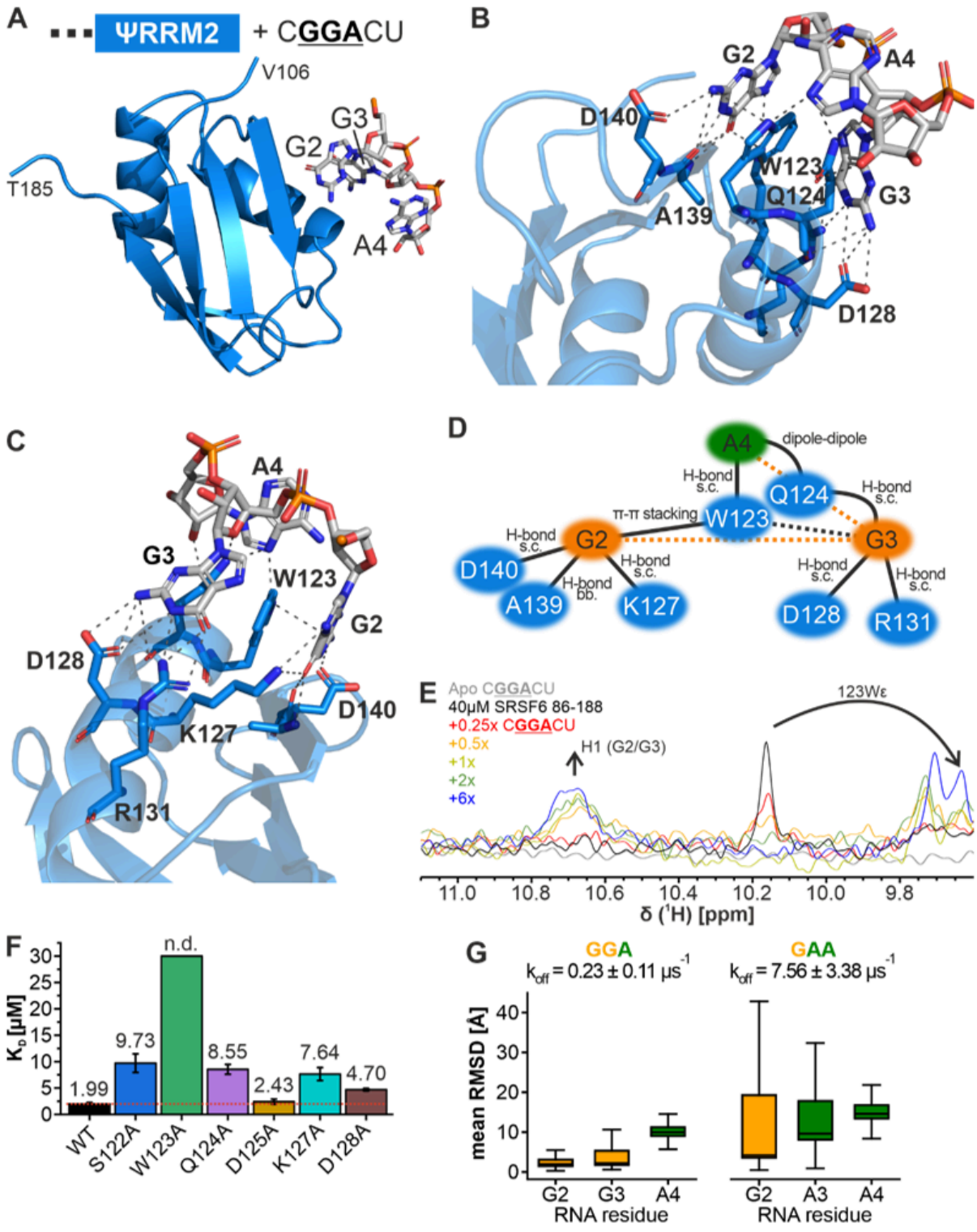


Figure 3

RNA-interactions of the single ΨRRM2 in high resolution.

A: Crystal structure of ΨRRM2 with CGGACU. Only the central “GGA” showed defined electron density (Supplementary Figure 4B). B and C: Zoom-ins on RNA-protein interaction of ΨRRM2 from crystal structure. D: Contact map of ΨRRM2 interacting with CGGACU RNA. E: ^{1}H-imino proton spectrum of

CGGACU RNA upon titration of ΨRRM2. F: Binding affinities of ΨRRM2 WT and mutants obtained from FP assays with CGGACU RNA. The red dotted line marks the affinity of the WT for comparison. For all data points and fits see Supplementary Figure 16C. G: Mean RMSD over replicas of SRSF6 ΨRRM2 simulations with GGA and GAA fragments. Error bars show the standard error of the mean. Bar colors correspond to the nucleotide. For details, see Supplementary Figure 17.

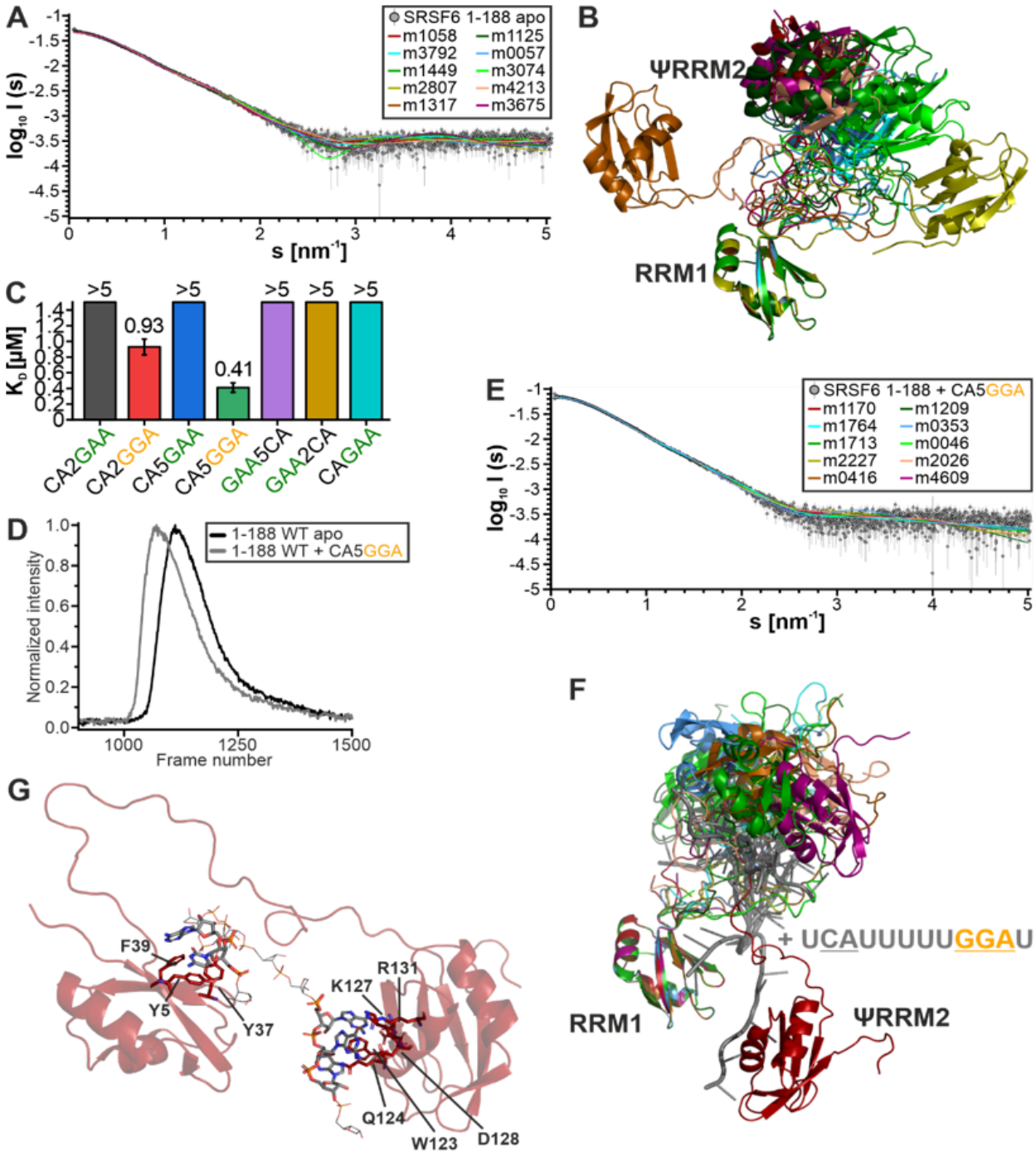


Figure 4

The SRSF6 tandem RRM arrangement and combined RNA recognition.

A: Crysol fits of best 10 selected models (out of 5000) to SEC-SAXS data of the tandem RRM. Chi^2-values were between 2.26 and 2.84. B: Best 10 apo models based on filtering against SEC-SAXS data for the tandem RRM. Models are structurally aligned on RRM1 and shown in cartoon representation. C: FP data of SRSF6 tandem RRM (1-188) with different RNAs. Experiments were performed as triplicates. D: Selected frames from SEC-SAXS runs of the tandem RRM alone or in complex with CA5GGA RNA. Intensity is normalized for easier comparison. E: Crysol fits of best 10 (out of 5000) selected models to SEC-SAXS data of the tandem RRM in complex with CA5GGA RNA. Chi^2-values were between 2.32 and 2.72. F: Best 10 models for the tandem RRM in complex with CA5GGA RNA as derived from filtering against SEC-SAXS data. Models are structurally aligned on RRM1. The RNA is depicted in grey. Both protein and RNA are shown in cartoon representation. G: Detailed view of m1170 (from E/F) in semi-transparent cartoon representation with selected RNA-interacting residues depicted as sticks and labeled. Bound nucleotides are also depicted as sticks with the side chain present, while the rest of the RNA is shown as lines. For further details, see Supplementary Table 20.

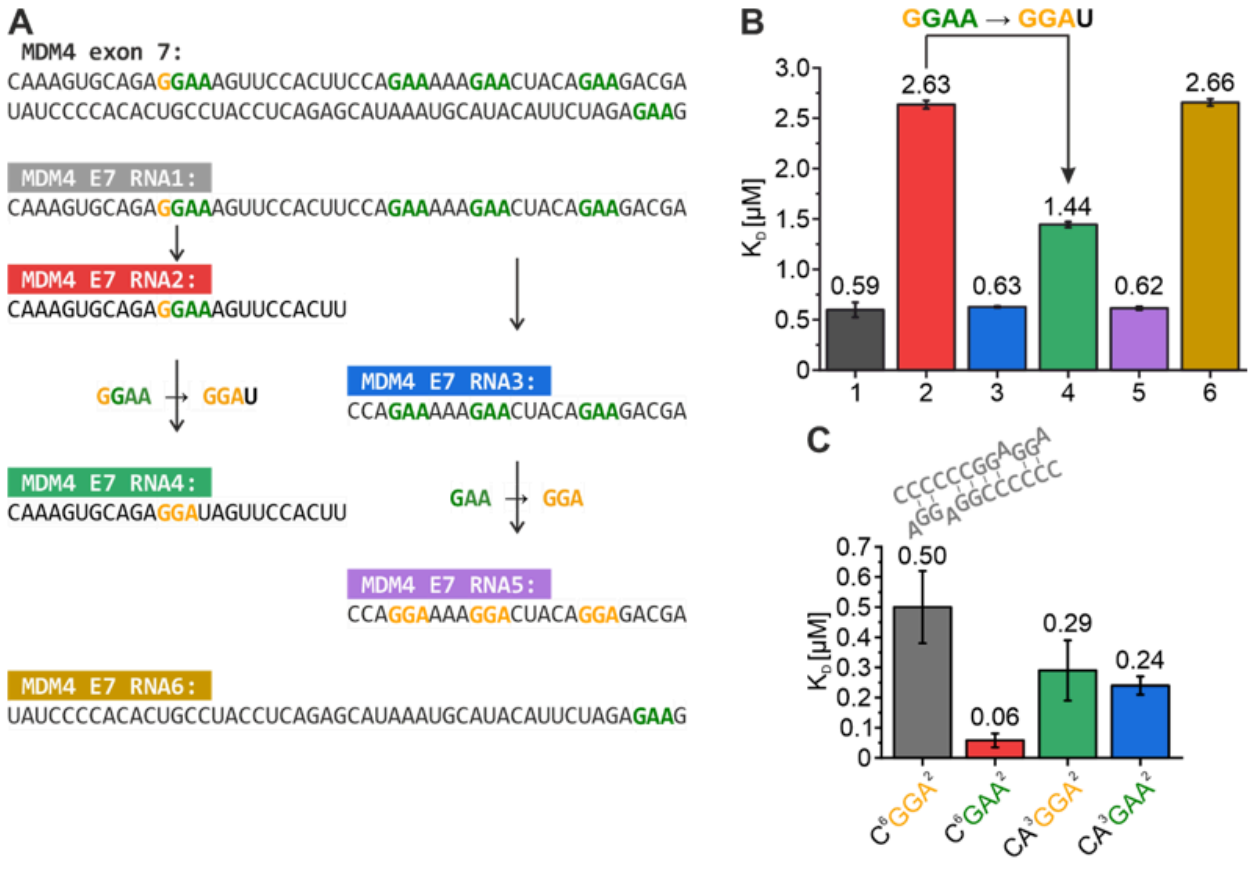


Figure 5

SRSF6 binding to clustered GAA triplets rescues lower binding affinity.

A: Sequence of *MDM4* exon 7 and the six respective RNA variants generated from this. For more details, see Supplementary Figure 24B. B: Barplot of K_D values derived from dual emission spectral shift assay

(see Supplementary Figure 24C) with SRSF6 tandem RRM and *MDM4* exon 7 variants. C: Barplot of K_D values derived from FP assays (see Supplementary Figure 24D) with SRSF6 tandem RRM and four different designed RNAs as indicated below the x-axis. Note that for C^6GGA^2 we observed secondary structure formation by NMR (see scheme on top and Supplementary Figure 25).

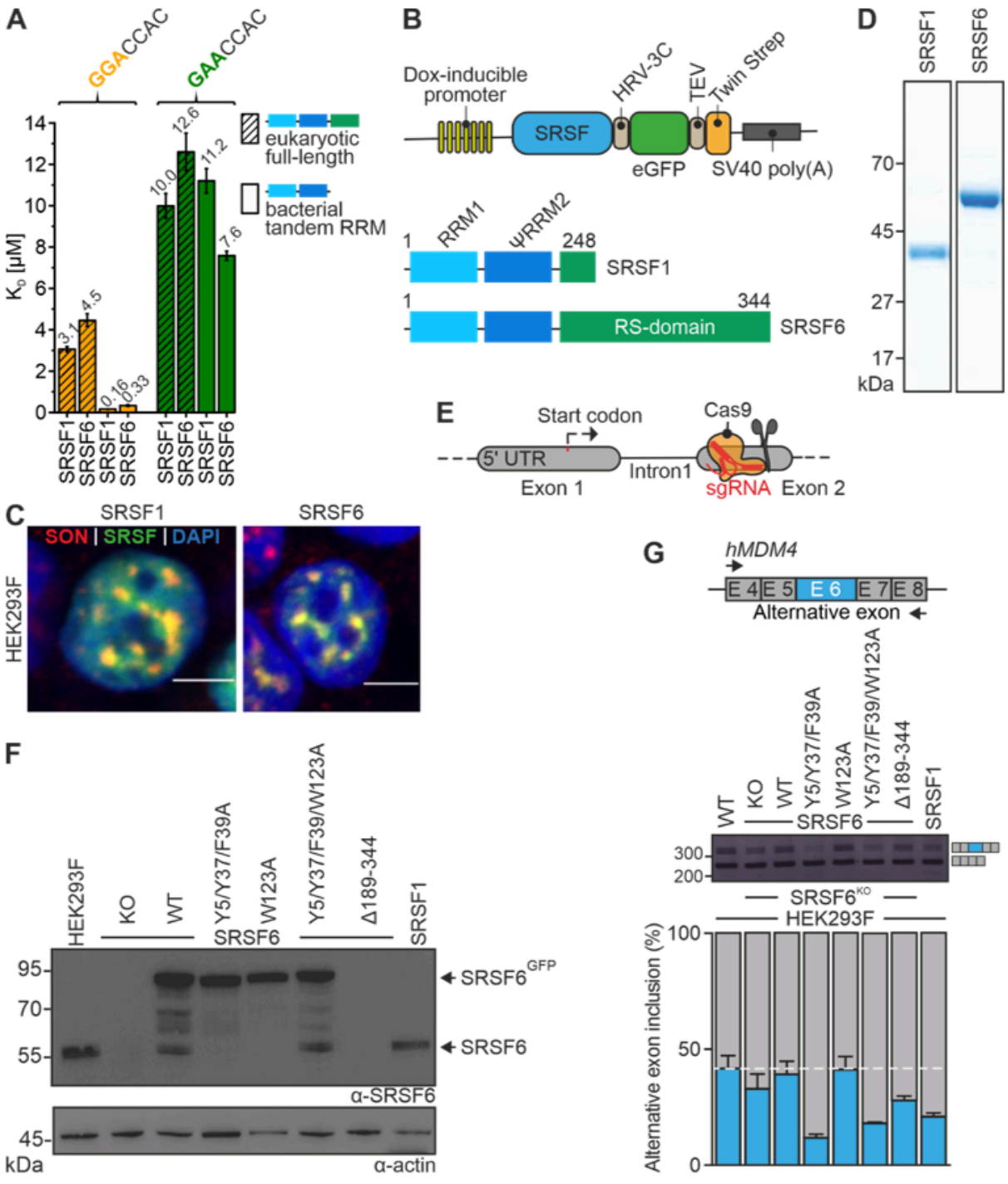


Figure 6

The RS domain equalizes the binding affinities of SRSF6 and SRSF1 to GGA.

A: Comparison of SRSF1 and SRSF6 RNA binding selectivity for the GGACCAC and GAACCAC motifs, determined by FP. Tandem RRMs were derived from *E. coli*, FL proteins were purified from HEK293F cells (Supplementary Figure 26). B: Schematic representation of the doxycycline (Dox)-inducible, stable expression construct for FL SRSF1 and SRSF6. The domain architecture is given for both FL proteins. C: Confocal immunofluorescence images showing the intracellular localization of eGFP-tagged SRSF1 and SRSF6 in nuclear speckles. Nuclear speckle marker SON (red) and nucleus (DAPI, blue). Scale bars = 5 µm. D: SDS-PAGE showing purified FL SRSF1 and SRSF6 from HEK293F cells. E: CRISPR-Cas9 genome editing strategy to knock out (KO) endogenous SRSF6. A single-guide RNA (sgRNA, red) targets the region between the start codon and the Cas9 cut site (scissors) near exon 2 of the *SRSF6* gene. F: Western blot analysis confirming the KO of endogenous SRSF6 in HEK293F cells. $SRSF6^{KO}$ cells were complemented with wild-type SRSF6 or respective RRM1-, RRM2- or RS domain mutants. Protein expression was induced for 48 h and assessed by western blotting using a-SRSF6 antibodies. a-actin antibodies were used as loading control. G: Model of human MDM4 (*hMDM4*) showing exons 4 to 8. Alternative exon 6 is highlighted in blue. Semi-quantitative RT-PCR of *hMDM4* splice isoforms either including (top) or skipping (bottom) alternative exon 6. Bar graph showing the percentage of exon 6 inclusion (blue) based on densitometric analysis of the RT-PCR products (n=3). Error bars represent standard deviations. The dotted line indicates the inclusion level in parental HEK293F cells.

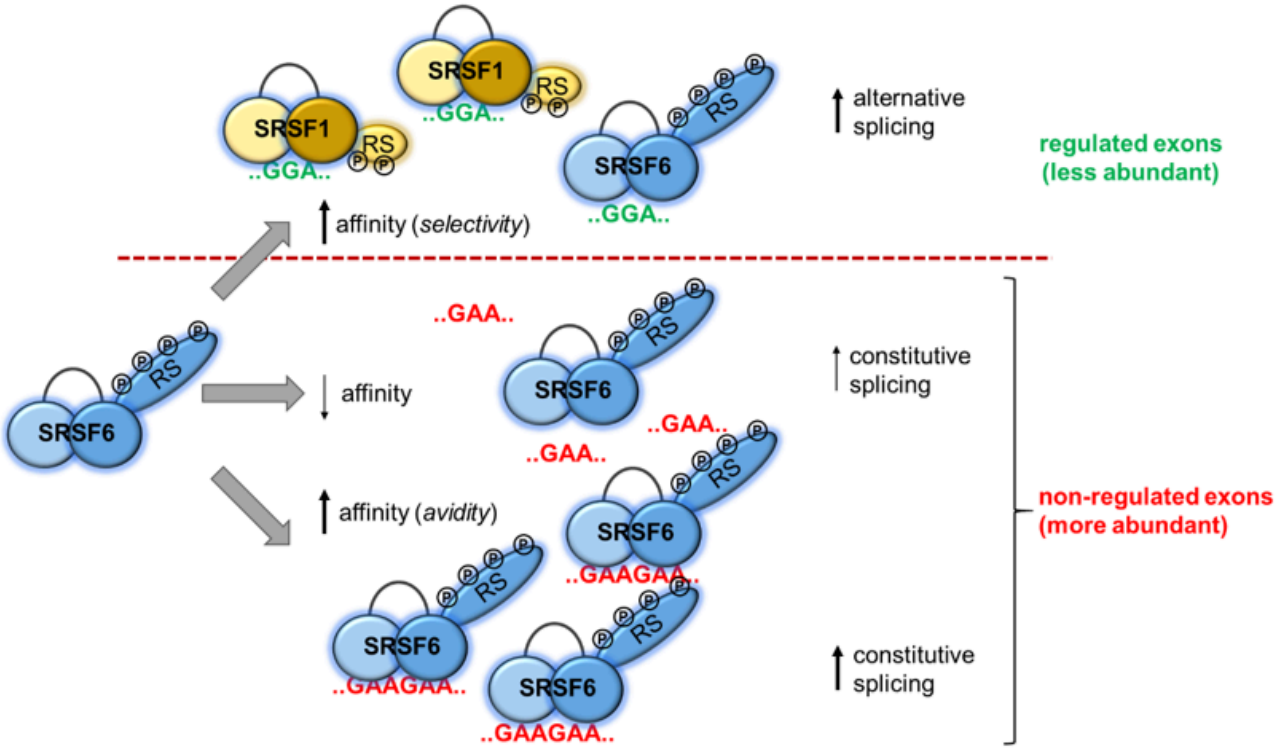


Figure 7

Summary model of SRSF6 recognizing alternative and constitutive exons.

The SRSF6 protein uses its tandem RRM domain for combined recognition of GGA-centered composite motifs in regulated exons or GAA motifs in non-regulated exons leading to alternative and constitutive splicing decisions, respectively. A higher intrinsic affinity for single GGA motifs is mediated by ΨRRM2 supported by contributions of RRM1 in flanking regions. GAA-repeats found in non-regulated exons can equalize binding to sufficiently high affinity, while the selectivity is provided through a filter function by the RS domain including its potential phosphorylation states. Selectively bound regulated GGA exons are competed for with SRSF1, which shows a higher abundance and mobility in cells providing an explanation for its dominant occupation of GGA sites in splicing decisions, which yet remain the selectively targeted exons for SRSF6, but with its occupancy hidden among the predominant localization to more abundant GAA-rich constitutively spliced exons.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- vonEhrSupplement.docx