Isolated pseudo–RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition

Antoine Clérya, Rahul Sinha, Olga Anczukówb, Anna Corrioneroc,d, Ahmed Moursya, Gerrit M. Daubnera, Juan Valcárelc,d, Adrian R. Krainerb, and Frédéric H.-T. Allaina1

aInstitute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology, 8093 Zurich, Switzerland; bCold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; cInstitució Catalana de Recerca i Estudis Avançats, Universitat Pompeu Fabra 08003 Barcelona, Spain; and dCentre de Regulació Genòmica, 08003 Barcelona, Spain

Edited by Dinshaw J. Patel, Memorial Sloan-Kettering Cancer Center, New York, NY, and approved June 17, 2013 (received for review February 22, 2013)

Serine/arginine (SR) proteins, one of the major families of alternative-splicing regulators in Eukarya, have two types of RNA-recognition motifs (RRMs): a canonical RRM and a pseudo-RRM. Although pseudo-RRMs are crucial for activity of SR proteins, their mode of action was unknown. By solving the structure of the human SRSF1 pseudo-RRM bound to RNA, we discovered a very unusual and sequence-specific RNA-binding mode that is centered on one α-helix and does not involve the β-sheet surface, which typically mediates RNA binding by RRMs. Remarkably, this mode of binding is conserved in all pseudo-RRMs tested. Furthermore, the isolated pseudo-RRM is sufficient to regulate splicing of about half of the SRSF1 target genes tested, and the bound α-helix is a pivotal element for this function. Our results strongly suggest that SR proteins with a pseudo-RRM frequently regulate splicing by competing with, rather than recruiting, spliceosome components, using solely this unusual RRM.

NMR | protein–RNA complex | splicing factor

Serine/arginine (SR) proteins are highly conserved in Eukarya and are key regulators of gene expression. These proteins are required for pre-mRNA splicing, regulate alternative splicing events, and play crucial roles in genomic stability, mRNA transcription, nuclear export, nonsense-mediated mRNA decay, and translation (1–5). Several studies have revealed links between these proteins and diseases (6), making the proteins potential therapeutic targets (7). For the last two decades, SR proteins have been studied intensively as regulators of alternative splicing, a mechanism used to modulate the expression of more than 95% of human genes (8, 9). Importantly, it is estimated that 15–50% of human disease-causing mutations affect splicing (10). Alternative splicing consists of the alternative selection of splice sites present within pre-mRNA, leading to different versions of mature mRNAs from a single gene (11). As a result, in some cases, proteins with opposite functions can be generated. In the context of cancer, alternative splicing can generate pro- or antiapoptotic isoforms (12). As an example, alternatively spliced variants of FAS/CD95 can be generated by inclusion or skipping of exon 6, depending on competition between antagonistic splicing factors (13–16). In the absence of Fas exon 6, the apoptotic receptor lacks the transmembrane domain and becomes soluble and antiapoptotic (17).

Each SR protein has a modular structure with one or two RNA-recognition motifs (RRMs) at its N-terminal part, followed by a C-terminal arginine-serine–rich (RS) domain containing multiple RS dipeptide repeats. In many cases, SR proteins have been shown to interact with purine-rich exonic splicing enhancers (ESEs) and to promote inclusion of the targeted exon in mRNAs (18). Two main models have been proposed to explain the mode of action of ESE-bound SR proteins in splicing regulation. In the recruitment model SR proteins recruit U1-70K and/or U2AF35 to 5′ and 3′ splice sites, respectively, to promote spliceosome assembly (4). In the inhibitor model SR proteins activate splicing by preventing the binding of splicing repressors, such as hnRNP A1, around the regulated splice site (4). In some cases, SR proteins have been shown not to enhance splicing but rather to promote exon skipping (19–26). However, their mode of action in these cases is still unclear.

The RS domain of SR proteins was shown to be important for splicing, because it initiates spliceosome recruitment by interacting with the branchpoint sequence (27) and/or splicingosomal components (28–30). However, in some cases, this domain also can be dispensable for constitutive and enhancer-dependent splicing (31), indicating a key function of the RRMs in this process. Although previously it was shown that the RRMs of SRSF1 can contact the splicingosomal component U1-70K directly (32), the main role of the RRMs seems to be to provide RNA-binding specificity and to dictate the position of SR proteins on pre-mRNAs (18).

The oncoprotein SRSF1 (SF2/ASF) (33–36) is one of the best-studied SR proteins. Depletion of SRSF1 protein results in death in chicken cells (37) and embryonic lethality in Caenorhabditis elegans (38). Like all SR proteins containing two RRMs, it comprises an N-terminal canonical RRM followed by a pseudo-RRM (Fig. L4). Although pseudo-RRMs appear to be critical for the activity of SR proteins (39–42), their mode of action still

Significance

Serine/arginine (SR) proteins are key regulators of eukaryotic gene expression and have been associated with multiple human diseases including cancers. Several members of this protein family contain a noncanonical RNA recognition motif (RRM), the pseudo-RRM, for which the mode of RNA recognition is unknown. Here, we solved the structure of SRSF1 pseudo-RRM bound to RNA. It reveals the RNA motif recognized and a very unusual mode of interaction, which is conserved for all the SR proteins containing pseudo-RRMs. Finally, we show that the pseudo-RRM in isolation often is sufficient to regulate splicing, and we reveal its mechanism of action.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The chemical shifts of SRSF1 RRM2–UGAAGGAC have been deposited in the BioMagResBank (BMRB), www.bmrb.wisc.edu (accession no. 19048). The coordinates of the SRSF1 RRM2–UGAAGGAC structures have been deposited in the Protein Data Bank (PDB), www.pdb.org (PDB ID code 2MID).

1To whom correspondence should be addressed. E-mail: allain@mol.biol.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1303445110/-/DCSupplemental.
needs to be clarified. Unlike canonical RRM s, pseudo-RRMs have a conserved, distinctive, invariant heptapeptide, SWQDLKD, on their α-helix 1 (43). In addition, their β-sheet surface lacks the set of conserved aromatic residues usually involved in RNA binding (44), suggesting that this domain uses an unusual mode of RNA recognition.

In this study, we solved the NMR structure of SRSF1 pseudo-RRM bound to RNA. The structure reveals that the domain specifically binds a GGA motif, primarily using the conserved residues located in α-helix 1. Remarkably, this unusual mode of RNA recognition is conserved for all pseudo-RRMs we tested from human, yeast, and fly. Moreover, we show that the isolated pseudo-RRM can regulate many different alternative splicing events as effectively as the full-length protein. Finally, our data suggest that the pseudo-RRM acts in splicing regulation by competing with binding of other splicing factors rather than by recruiting them to the cassette exon.
natural linker showed that chemical shifts are similar (SI Appendix, Fig. S1A). This analysis reveals that, upon RNA binding, the RRMAs behave similarly when they are separated or linked in their natural conformation. In addition, isothermal titration calorimetry (ITC) measurements indicated that only the pseudo-RRM interacts efficiently with the purine-rich sequence tested (SI Appendix, Fig. S1B), demonstrating that SRSF1 RRM2 is the primary domain for recognition of purine-rich ESEs. Therefore, we investigated the RNA-binding mode of this domain using NMR spectroscopy. To obtain NMR spectra of sufficient quality for structure determination, we tested several RNA sequences derived from SELEX (45) and CLIP consensus (46). High-quality NMR spectra were obtained with SRSF1 RRM2 (amino acids 107–203) fused to the solubility tag GB1 and bound to 5′-UGAAGGAC-3′ (Fig. 1B). Importantly, this sequence fits the consensus sequence determined by CLIP with SRSF1 (5′-UGAAGGAC-3′) (46). We measured a dissociation constant of 0.8 μM for this complex, using ITC (Fig. 1C). NMR titration revealed large chemical shift perturbations in the β2-strand and, more surprisingly in the α-helix 1 and N-terminal extremity of the domain (Fig. 1D). We calculated the structure of this complex using 1,487 nuclear Overhauser effect (NOE)-derived distance restraints, including 38 intermolecular ones (SI Appendix, Table S1 and Fig. S2). We obtained a very precise ensemble of 16 structures (Fig. 2A) with an r.m.s.d. of 0.92 Å for all of the heavy atoms (SI Appendix, Table S1).

**SRSF1 Pseudo-RRM Uses α-Helix 1 to Recognize GGA.** The pseudo-RRM in complex with RNA adopted a canonical β2α2β1-α2β2 fold that is unchanged compared with the non–ligand-bound (apo) form of the protein (47). The RNA structure adopted an extended conformation with all nucleotides in the C2'-endo sugar conformation, except for G7, which had a C3'-endo conformation (Fig. 2). The mode of binding is very unusual for an RRM, because the RNA-binding site is centered on α-helix 1, and the β-sheet surface does not contact the RNA (Fig. 2B and C). The pseudo-RRM recognizes only three of the eight nucleotides, G6, G7, and A7, which wrap the bottom of α-helix 1. The N-terminal part of this helix, comprising Ser133, Trp134, and Gln135, forms an ideal surface for binding three nucleotides, with G6 stacking on the Trp134 side-chain, G7 stacking on the main-chain, and A7 stacking over the surface formed by the Ser133 and Gln135 side-chains (Fig. 2C and D). Additionally, the side-chains of Trp134 and Gln135 interact with the sugar rings of G6 and G7, respectively. All these contacts provide affinity for the three nucleotides and organize the GGA triplet into an arch with two consecutive bases perpendicular to one another, but they do not confer sequence specificity. Sequence specificity is achieved by other features of the pseudo-RRM. The Watson–Crick edges of G6 and G7 are recognized by several hydrogen bonds formed with the main-chain of Ala150 (in the β2-strand) and the side-chains of Lys138 and Asp139 (in the α-helix 1), respectively (Fig. 2D). In addition, the Arg117 side-chain in the inter-RRM linker contacts the N7s of both guanines. Finally, A7’s N6 amino group is hydrogen-bonded with Asp136 from α-helix 1, and its H2 interacts with the His183 side-chain (in the α2-β4 loop) (Fig. 2D). No specific contact is observed for A4, U2, G3, A4, and C4. Overall, the structure revealed that the SRSF1 pseudo-RRM

**Fig. 2. Overview of the solution structure of SRSF1 pseudo-RRM in complex with 5′-UGAAGGAC-3′ RNA.** (A) Overlay of the 16 lowest-energy structures superimposed on the backbone of the structured parts of the protein and heavy atoms of RNA. The protein backbone is shown in gray, and heavy atoms are shown in orange (P atoms), yellow (C atoms for RNA), green (C atoms for protein), red (O atoms), and blue (N atoms). Only the ordered regions of the RRM (residues 116–193) and RNA (A4 to C2) are shown. (B) Structure of the complex is shown in ribbon (protein backbone) and stick (RNA) representation. The color scheme is the same as in A. Important protein side-chains involved in RNA interactions are represented as sticks. (C) Stereo view of the structure of the SRSF1 pseudo-RRM in complex with RNA. (D and E) Two views of the molecular recognition of the 5′-GGA-3′ RNA sequence by the pseudo-RRM. Protein–RNA interactions involving G6, G7, and A7 nucleotides are shown. Color schemes are as in A. Hydrogen bonds are represented by purple dashed lines.
recognizes a 5'-GGA-3' motif, with a majority of the contacts to the RNA involving the highly conserved S133WQDLKD139 sequence within α-helix 1.

In good agreement with the structure, the domain binds 5'-UGAAGGAC-3' and 5'-AGGACG-3' RNAs with almost the same affinity ($K_d$ of 0.8 and 1 μM, respectively) (Fig. 1 and SI Appendix, Fig. S3). Moreover, substitutions of either G3 or G6 by adenines led to no detectable binding measured by ITC, and substitution of A1 by a guanine or a cytosine decreased the affinity by a factor of two or three, respectively (SI Appendix, Fig. S3). The invariant SWQDLKD motif located in α-helix 1 of pseudo-RRMs represents a sequence signature of these domains (43). Our structure clearly explains this conservation, considering that, except for the leucine that anchors α-helix 1 into the domain hydrophobic core, each residue of this heptapeptide contacts the RNA (Fig. 2). We measured a five- to 10-fold decrease in the RNA-binding affinity when we mutated each of these residues individually (SI Appendix, Fig. S4).

Isolated SRSF1 Pseudo-RRM Regulates Splicing Events in Cells. We next investigated whether the RNA-binding specificity we determined for SRSF1 pseudo-RRM is relevant in vivo in the context of the full-length protein. We transfected human cells with several variants of SRSF1 and analyzed their effect on the splicing of insulin-like growth factor 1 (IGF1) exon 5, which has an SRSF1-dependent purine-rich ESE containing a GGA motif (Fig. 3A) (48). As previously reported (48), exon 5 inclusion increased in the presence of the SRSF1 WT protein (Fig. 3B–E). Each single amino acid substitution introduced into the full-length protein and expected to affect the binding of SRSF1 pseudo-RRM to RNA severely reduced the splicing enhancement of IGF1 exon 5 (Fig. 3B–E, lanes 8–13 and 15). Such large changes in splicing were surprising, considering that the mutations tested decreased RNA-binding affinity only by five- to 10-fold (SI Appendix, Fig. S4). This effect might originate from the overexpression of these protein variants resulting in nuclear concentrations close to their $K_d$. Conversely, mutation of Tyr149, which is not involved in the interaction with RNA, did not have a significant effect on splicing (Fig. 3B–E, lane 14). These results indicate that, in the context of the full-length protein, SRSF1 pseudo-RRM uses the mode of RNA interaction shown in our structure. In good agreement with the dominant function of the pseudo-RRM in the recognition of purine-rich ESEs by SRSF1, deletion of the pseudo-RRM (ΔRRM2) prevented this splicing enhancement, whereas deletion of the canonical RRM1 or the RS domain still allowed exon 5 inclusion, although less efficiently than in the full-length protein (Fig. 3B–E, lanes 1–5). However, the increase in the percentage of exon 6 inclusion observed in the absence of the RS domain shows that this part of the protein also is involved in the regulation of IGF1 alternative splicing (Fig. 3E, lane 5). More surprisingly, unlike RRM1, the isolated RRM2 could enhance exon 5 splicing to nearly the same extent as SRSF1 after deletion of the RS domain (ΔRS) and more than RRM1-deleted SRSF1 (ΔRRM1) (Fig. 3B–E, lanes 1–7), showing that the pseudo-RRM can activate splicing as a single domain in vivo.

In contrast to the effect observed with IGF1 exon 5, SRSF1 promotes FAS/CD95 exon 6 skipping, which gives rise to an antiapoptotic isoform (Fig. 4A) (49). To understand better the mode of action of SRSF1 pseudo-RRM in splicing, we investigated its role in splicing of FAS exon 6. As shown in Fig. 4, the isolated pseudo-RRM could also repress splicing of FAS exon 6, even more efficiently than the full-length protein (Fig. 4B–D). Together with our data for IGF1, these results demonstrate that the isolated pseudo-RRM can recapitulate the effects of full-length SRSF1 either as an activator or as a repressor of exon inclusion, depending on the pre-mRNA substrate. Similar to our observations for IGF1 exon 5, mutating Lys138 or Asp139 in the SWQDLKD motif of the domain impaired the ability of the isolated RRM to promote FAS exon 6 skipping (Fig. 4B–D). These data indicate that binding of the SRSF1 pseudo-RRM to RNA via α-helix 1 is required for its function as a regulator of alternative splicing.

To determine whether the pseudo-RRM can regulate splicing of additional SRSF1 target genes, we analyzed the splicing
Fig. 4. Effect of SRSF1 mutations on FAS exon 6 splicing in cells. (A) Schematic representation of FAS exon 6 minigene. Exon 6 inclusion or skipping generates mRNA isoforms encoding a pro- or an anti-apoptotic form of the Fas/CD95 receptor, respectively. The location of primers used to amplify transcripts derived from the FAS minigene used in cotransfection experiments is shown. (B) Western blots showing the relative expression levels of the different T7-tagged versions of SRSF1. (C) RT-PCR gel showing the levels of exon 6 mRNA isoforms upon overexpression of either full-length SRSF1 or WT and various mutant versions of the SRSF1 pseudo-RRM in HeLa cells. The black arrow indicates the junction between two initially distant lanes present in the same gel. (D) Graph showing the percentages of exon 6 inclusion in FAS mRNAs upon overexpression of various SRSF1 versions. Error bars indicate SD.

patterns of 18 endogenous SRSF1 targets. Among these, MKNK2 and MSTR1/RON were shown previously to be regulated by SRSF1, whereas the remaining splicing events correspond to targets identified by RNA-seq in HeLa cells. We performed semiquantitative radioactive RT-PCR on RNA from HeLa cells upon two- to threefold overexpression of either full-length SRSF1 or the pseudo-RRM only (Fig. 5 A). To enrich for cells expressing SRSF1 or the pseudo-RRM, we cotransfected GFP to enable sorting of the cells by flow cytometry. We observed that seven of the 18 splicing events responded to expression of the pseudo-RRM or the full-length protein to a similar extent (Fig. 5 B–D and SI Appendix, Fig. S5). Moreover, the pseudo-RRM promoted exon skipping or exon inclusion, depending on the particular SRSF1 splicing target. These results suggest that SRSF1 pseudo-RRM is sufficient for the regulation of a substantial fraction of SRSF1 splicing targets.

SRSF1 Pseudo-RRM Competes Against hnRNP A1. We then tried to determine how the isolated SRSF1 pseudo-RRM regulates alternative splicing. Because of the relatively small size of the domain, it is unlikely that it would act on splicing by recruiting spliceosome components. Indeed, it was reported recently that both RRMs of SRSF1 are required to interact with U1-70K (32).

A more tempting hypothesis could be that the pseudo-RRM regulates splicing by competing for RNA binding with one or several splicing factors. Previous studies indicated that hnRNP A1 is an efficient competitor of SRSF1 full-length protein (50, 51). HnRNP A1 recognizes AGG motifs sequence specifically (52). Because we showed that SRSF1 pseudo-RRM recognizes GGA, a sequence containing AGGA in principle could be recognized by both proteins. To test our hypothesis, we performed in vivo competition assays using IGF1 and FAS minigenes with AGGA elements (Fig. 6). Increasing the expression of hnRNP A1 promoted IGF1 exon 5 skipping (Fig. 6 A, Upper, lanes 1–4) and, more surprisingly, FAS exon 6 inclusion (Fig. 6 B, Upper, lanes 1–4), thus exhibiting an effect on FAS and IGF1 splicing opposite to that of SRSF1 (Fig. 6 A and B, Lower, lanes 1–4). In addition, the effects of hnRNP A1 on FAS and IGF1 splicing were efficiently reversed by overexpression of the SRSF1 pseudo-RRM alone (Fig. 6 A and B, Lower, lanes 5–8). Similarly, increasing the expression of hnRNP A1 prevented SRSF1 pseudo-RRM2 from promoting IGF1 exon 5 inclusion and FAS exon 6 skipping (Fig. 6 A and B, Upper, lanes 5–8). These results suggest that the isolated SRSF1 pseudo-RRM can stimulate exon inclusion or exon skipping (depending on the substrate) by competing with hnRNP A1.

Unusual Mode of Interaction of Pseudo-RRMs with RNA Is Conserved. With the exception of R117 and H183, all of the residues of SRSF1 pseudo-RRM involved in sequence-specific recognition of GGA are fully conserved in the pseudo-RRMs of all known SR proteins containing two RRMs, including Drosophila B52 and yeast Npl3p (SI Appendix, Fig. S6). This conservation strongly suggests that all these domains have the same sequence specificity. To test this hypothesis, we performed affinity measurements by ITC and NMR titrations of various pseudo-RRMs in complex with 5′-AGGAC-3′ or 5′-UGAAGGAC-3′ RNA (SI Appendix, Figs. S7 and S8). Although all tested pseudo-RRMs interacted with these RNAs, their affinity was 2–15 times less than that of SRSF1 pseudo-RRM. The lower binding affinity of the pseudo-RRMs of SRSF4, -5, and -6 and Npl3p may result from the lack of conservation of Arg117 and His183 and/or from a different amino acid environment around the binding site (Fig. 2 and SI Appendix, Fig. S6). However, the NMR titration with the 5′-AGGAC-3′ RNA unambiguously showed that α-helix 1 remains the RNA-binding surface for all these pseudo-RRMs, as evidenced by the large chemical shift changes observed for the NMR signals corresponding to the tryptophan and glutamine residues in the conserved SWQDLKD motif and the alanine located in the β2-strand of these domains (Fig. 7 A and SI Appendix, Fig. S8). In addition, none of the pseudo-RRMs that we tested interacted with the mutant 5′-AGAAGC-3′ RNA, in which the central G6 was mutated to an adenine (underlined), supporting our observation above that the interaction is sequence specific (SI Appendix, Fig. S9). Taken together, these results indicate that this unusual mode of binding and the GGA sequence specificity can be generalized to all known pseudo-RRMs. In agreement with this conclusion, and similar to our observations with the corresponding full-length proteins, SRSF9’s pseudo-RRM also activated FAS exon 6 skipping, whereas SRSF5’s
Fig. 5. Effect of SRSF1 pseudo-RRM expression on alternative splicing of endogenous SRSF1 target genes. (A) SRSF1 levels in control (pCGT) and in SRSF1- and RRM2-overexpressing HeLa cells were assessed by radioactive RT-PCR, followed by native PAGE and autoradiography. SRSF1 levels were normalized to ACTB mRNA and expressed as the fold change compared with control cells (n = 3). ACTB mRNA was used as a loading control. Error bars indicate SEM. (B) Splicing of known SRSF1 target genes (MKMK2, RON) was analyzed upon SRSF1 and RRM2 overexpression in cells from A, as described above. Representative gels are shown in SI Appendix, Fig. S5A. The ratio of each splicing isoform was quantified, and the relative level of exon inclusion was plotted (n = 3; t test ***P < 0.0005, **P < 0.005). Error bars indicate SD. (C and D) RT-PCR analysis of SRSF1 target genes expressed in cells from A that respond to both RRM2 and full-length SRSF1 (D). Primers located in the exons flanking the alternatively spliced exon (SI Appendix, Table S1) allow amplification of either the full-length isoform or the skipped-exon isoform. For representative gels see SI Appendix, Fig. S5 B and C. The ratio of each isoform was quantified, and the relative level of exon inclusion was plotted (n ≥ 3; t test ***P < 0.0005, **P < 0.005, *P < 0.01). Error bars indicate SD.

Discussion

Interaction of Pseudo-RRM with RNA, a Conserved Mode of Binding.

In this paper, we show that the pseudo-RRM is primarily responsible for the interaction of SRSF1 with purine-rich sequences, which are found in the ESEs targeted by SR proteins in vivo (18). In addition, our data reveal that all pseudo-RRMs we tested interact specifically with the 5′-GGA-3′ motif using the same mode of binding. In particular, the presence of two consecutive guanines is crucial, because their substitution abolished the interaction of the pseudo-RRM with RNA (SI Appendix, Figs. S3 and S9). Consistent with our results, all the consensus sequences obtained by SELEX for pseudo-RRM-containing SR proteins comprise at least one GG dinucleotide (18), and RNA sequences with the highest affinity for Npl3p also comprise this motif (53). This very similar RNA sequence specificity shared by all pseudo-RRMs could explain the functional redundancy observed in some cases for SR proteins (18). However, although the pseudo-RRM plays a crucial role in RNA recognition and alternative splicing regulation, it is likely that in some cases the specificity of SR proteins is modulated by the canonical RRM1 to accommodate longer or slightly different sequences.

α-Helix 1 of Pseudo-RRMs, a Binding Platform for Both Protein and RNA.

Although previous studies suggested the potential involvement of the α-helix 1 of SRSF1 pseudo-RRM in protein (42, 54) and RNA (40, 47) interactions, there was no rational as to how such a small surface could accommodate both types of polymers. These observations now can be explained by our structure and by the structure of SRPK1 (54) bound to the SRSF1 pseudo-RRM. To our surprise, we found that the mode of binding of the α-helix 1 is very similar for both the RNA and the protein. Remarkably, the side-chains of SRPK1, Trp88, and His90 occupy the same sites as the bases of G5 and G6, respectively. The backbones of both polymers follow the same path in both structures with the Gln135 side-chain being hydrogen bonded with the main-chain of both ligands (Fig. 7C). Note that the directionality of the polymers is different. In addition, three nucleotides are needed to replace two amino acids (Figs. 2D and 7C). This double involvement of α-helix 1 also implies that SRSF1 phosphorylation by SRPK1 and RNA binding must be mutually exclusive events. This information is important to understand better the mode of action of SRSF1 in cells, because the use of the same binding platform to interact with either RNA or protein partners most likely dictates sequential ligand recognition at different regulatory steps in the gene-expression pathway.

An Additional Mode of RNA Recognition for the Family of RRM.

Although three subclasses of RRM were identified in Eukaryota...
Fig. 6. Effect on IGF1 and FAS splicing of SRSF1 pseudo-RRM and hnRNPA1 coexpression in cells. (A) RT-PCR gels showing the levels of IGF1 exon 5 and exon 6 mRNA isoforms upon coexpression of hnRNPA1 and SRSF1 pseudo-RRM in HeLa cells. (Upper) The SRSF1 pseudo-RRM was coexpressed with increasing amounts of hnRNPA1. (Lower) hnRNPA1 was coexpressed with increasing amounts of the SRSF1 pseudo-RRM. GAPDH was used as a reference for loading. The black arrow indicates the junction between two initially distant lanes present in the same gel. (B) As in A, RT-PCR gels show the levels of FAS exon 6 mRNA isoforms upon coexpression of hnRNPA1 and the SRSF1 pseudo-RRM.

Pseudo-RRM, a Key Domain of SRSF1 for Splicing Regulation. SR proteins are well known for promoting splice-site selection via their interaction with ESEs. However, their mode of action appears to be more diverse than anticipated, with a variety of possible mechanisms influencing splice-site selection. Recent genome-wide analysis performed in mammalian cells confirmed the idea that SR proteins can either promote or inhibit exon inclusion depending on their position on pre-mRNAs (57). Their binding to intronic sequences was reported primarily to promote exon inclusion (57), whereas their interaction with exonic sequences seems to be more complex. It was proposed that the binding of SR proteins to alternative exons promotes their inclusion, whereas their interaction with a flanking exon causes skipping of the internal alternative exon (22, 59).

Here, we found that the isolated pseudo-RRM of SRSF1 is sufficient to modulate splicing of nearly half of the tested SRSF1 target genes (Fig. 5 and SI Appendix, Fig. S5). We propose that the domain either activates or inhibits splicing by competing with binding of the splicing repressor hnRNPA1 (50, 51) to a 5′-AGGA-3′ motif, which contains overlapping binding sites for both proteins (5′-AGG-3′ and 5′-GGA-3′ for hnRNPA1 and SRSF1 pseudo-RRM, respectively) (SI Appendix, Fig. S10). In the context of IGF1, the motif is located in alternative exon 5, and the pseudo-RRM activates inclusion of the targeted exon, whereas hnRNPA1 represses splicing (Figs. 6A and 6B). An opposite effect was observed with FAS, because SRSF1 pseudo-RRM promotes exon 6 skipping, and hnRNPA1 activates its inclusion (Fig. 6B). The effect is not observed with hnRNPA1 alone, because hnRNPs were shown primarily to mediate splicing repression from exonic positions (60–62). Nevertheless, a study performed in parallel with our work recently reported the same effect of hnRNPA1 on FAS exon 6 splicing (63). The authors identified the binding site of this splicing factor on exon 5, close to the 5′ splice site. Remarkably, the binding sequence comprises a 5′-AGGA-3′ motif, which could explain the competition we observed between SRSF1 pseudo-RRM and hnRNPA1 on FAS alternative splicing (Fig. 8B).

In this context, the function of hnRNPA1 in 3′ splice-site proofreading and 3′ splice-site recognition (64) also could help explain the antagonism between SRSF1 and hnRNPA1 in the regulation of FAS alternative splicing. Finally, our data confirm that the effect of SR and hnRNPs proteins on splicing is position dependent (65).

Interestingly, our data obtained with the SRSF1 pseudo-RRM alone fit perfectly with the above-mentioned mode of action of SR proteins in splicing regulation (22). Indeed, the binding of SRSF1 pseudo-RRM to the alternative exon of IGF1 itself activates its inclusion (Fig. 8A), whereas interaction of the domain with the flanking exon 5 and/or with the 3′ splice site of intron 5 promotes skipping of FAS alternative exon 6 (Fig. 8B).

We still need to understand the link between SRSF1 pseudo-RRM and the spliceosome machinery, but our data indicate that this domain can be crucial and even sufficient to regulate these types of splicing events. Finally, the activity of pseudo-RRMs in splicing is likely to be influenced by multiple additional parameters, including formation of RNA secondary structures or competition with other proteins targeting overlapping binding sites (57, 59), which may limit the accessibility to these domains of 5′-AGGA-3′ motifs that act as potential splicing modifiers.

Taken together, our results suggest that the pseudo-RRM can activate or repress splicing, depending on the position of the sequence bound by SRSF1 and hnRNPA1 (an AGGA motif located in an ESE or close to a splice site, respectively) (Fig. 8). Remarkably, and contrary to the usually invoked mechanism of action, in this competitive model, the RS domain of the SR proteins is dispensable. Finally, our detailed insights into the unusual mode of RNA binding and mode of action of SR proteins should help facilitate the design of molecules that can shift the balance of alternative splicing isoforms for therapeutic purposes.

Accession Codes. We deposited the chemical shifts of SRSF1 RRM2–UGAAGGAC in the BioMagResBank under accession number 19248. We have deposited the coordinates of the SRSF1 RRM2–UGAAGGAC structures in the Protein Data Bank (PDB) under the PDB ID 2MSD.

Materials and Methods

Preparation of RNA–Protein Complexes. We cloned in the pET28 expression vector all pseudo-RRM ORFs corresponding to amino acids 107–183 of SRSF5, 103–193 of B52, 193–282 of Nip1sp, 97–178 of SRSF4, 101–183 of SRSF5, 103–189 of SRSF6, and 103–185 of SRSF9, as well as SRSF1 RRMM1 (residues 1–97) and SRSF1 RRMM1–2 (residues 1–203). These recombinant proteins were fused to an N-terminal 6×His tag and were overexpressed at 37 °C in Escherichia coli BL21 (DE3) codon plus cells in minimal M9 medium containing 1 g/L 3[15]NH₄Cl and 4 g/L glucose (for 15N-labeled protein) or 1 g/L 13C-glucose and 2 g/L 13C-glucose (for 15N- and 13C-labeled protein). A GB1 tag was fused at the
N-terminal extremity of the SRSF1 pseudo-RRM to increase the solubility and stability of the protein. As seen in Fig. 1 and SI Appendix, Fig. S8, we observed the same chemical shift perturbations upon RNA binding in the presence or absence of the GB1 tag, demonstrating that the tag does not influence the pseudo-RRM interaction with RNA.

Proteins were purified by two successive nickel affinity chromatography (Qiagen) steps, as previously described (66), dialyzed against NMR buffer (50 mM L-Glu, 50 mM L-Arg, 0.05% β-mercaptoethanol, and 20 mM NaH2PO4 at pH 5.5) and concentrated to 0.8 mM with a 10-kDa molecular mass cutoff Centricon device (Vivascience). The SRSF1 RRM1, RRM2, and RRM1+2 proteins used in this experiment were deprotected according to the manufacturer’s instructions, desalted using a G-15–size exclusion column (Amersham), lyophilized, and resuspended in NMR buffer.

RNA-protein complexes used to solve structures were formed in NMR buffer at an RNA:protein ratio of 1:1, at a 0.8 mM concentration. All NMR measurements were performed in NMR buffer at 298 K using Bruker AVIII-500 MHz and 700 MHz equipped with a cryoprobe and AVIII-600 MHz and Avance-900 MHz spectrometers. Data were processed using Topspin 2.0 (Bruker) and analyzed with Sparky (www.cgl.ucsf.edu/home/sparky/).

NMR Measurement All NMR measurements were performed in NMR buffer at 313 K using Bruker AVIII-500 MHz and 700 MHz equipped with a cryoprobe and AVIII-600 MHz and Avance-900 MHz spectrometers. Data were processed using Topspin 2.0 (Bruker) and analyzed with Sparky (www.cgl.ucsf.edu/home/sparky/).

Protein sequence-specific backbone and side-chain assignments were achieved using 2D 1H-15N HSQC, 2D 1H-13C HSQC, 3D HNCA, 3D HNCO, 3D CBCACONH, 3D HcCH total correlation spectroscopy (TOCSY), 3D nuclear Overhauser effect spectroscopy (NOESY), 1H-15N HSQC, and 3D NOESY 1H-13C HSQC aliphatic for review, see ref. 67). Aromatic proton assignments were performed using 2D 1H-1H TOCSY and 3D NOESY 1H-13C HSQC aromatic.

Resonance assignments of RNA in complex with SRSF1 pseudo-RRM were performed using 2D 1H-1H TOCSY, 2D 1H-1H NOESY, 2D 13C 1F-filtered 2F-filtered NOESY (68), and natural abundance 2D 13C-1H HSQC in 100% D2O.

Intermolecular NOE were obtained using 2D 1H-1H NOESY and 3D 13C-1F-edited 3F-filtered HSQC-NOESY (69) using unlabeled RNA and 15N-labeled and 13C-labeled proteins, respectively. Intermolecular NOE between imino protons of the RNA G3 and G4 nucleotides and SRSF1 protons were obtained using 2D 1H-1H NOESY at 298 K in H2O.

All NOESY spectra were recorded with a mixing time of 150 ms, the 3D TOCSY spectrum with a mixing time of 23 ms, and the 2D TOCSY with a mixing time of 50 ms.

Structure Calculation and Refinement. AtmosCandid software (70, 71) was used to generate preliminary structures and a list of automatically assigned NOE distance constraints for SRSF1 pseudo-RRM in complex with RNA. Peak picking and NOE assignments were performed using 3D NOESY (1H- and 13C-edited) spectra. Additionally, intraprotein hydrogen-bond constraints were added based on hydrogen–deuterium exchange experiments on the amide protons. For these hydrogen bonds, the oxygen acceptors were

![Fig. 7. Pseudo-RRMs use a conserved mode of interaction with RNA, centered on α-helix 1. (A) Selection of chemical-shift perturbations observed for all the tested pseudo-RRMs in the presence of S′-AGGAC-3′ RNA. Close view of chemical-shift perturbations observed in SI Appendix, Fig. S8 and corresponding to residues for which the interaction with RNA is characteristic of the binding mode observed for SRSF1 pseudo-RRM. Titrations were performed at 40 °C in the NMR buffer. The peaks corresponding to the free and RNA-bound states (RNA:protein ratios of 0.3:1 and 1:1, respectively) are colored blue, orange, and red, respectively. All the recombinant proteins used in this figure were purified in the absence of the GB1 tag. Ky values determined by ITC measurements are shown for each protein. (B) SRSF9’s but not SRSF5’s pseudo-RRM induces FAS exon 6 skipping. RT-PCR gels (Top), Western blot analyses (Middle), and graph (Bottom) showing percentages of exon 6 inclusion in FAS mRNAs upon overexpression of various pseudo-RRM versions are shown. Error bars indicate SD. (C) Structures of the SRSF1 pseudo-RRM bound to the GG dinucleotide of the S′-UGAAGGAC-3′ RNA and to the Trp-Gly-His tripeptide of SRPK1 in green and black, respectively. The protein backbone is shown in gray, with the side-chains of SRSF1 and SRPK1 in green and red, respectively. Heavy atoms of RNA are in orange (P atoms), yellow (C atoms), red (O atoms), and blue (N atoms).](https://www.pnas.org/content/early/2018/07/12/1807460115.short)
Fig. 8. Models for splicing regulation by the SRSF1 pseudo-RRM. (A) Binding of the SRSF1 pseudo-RRM to IGF1 exon 5 promotes its inclusion. The domain interacts with an AGGA-containing ESE and prevents the recruitment of hnRNP A1 to the targeted exon. (B) The SRSF1 pseudo-RRM binds to FAS exon 5 close to the 3′ splice site and promotes the skipping of exon 6. HnRNP A1 competes with the binding of the SRSF1 pseudo-RRM to FAS exon 5 and activates exon 6 inclusion (63). Alternatively, the SRSF1 pseudo-RRM also could compete with the proofreading activity of hnRNP A1 at the 3′ splice site of FAS introns 5 (64).

identified on preliminary structures calculated without hydrogen-bond constraints.

Seven iterations were performed, and 100 independent structures were calculated at each iteration step. Structures of the protein–RNA complexes were calculated with CYANA (71) by adding the manually assigned intra-molecular RNA and RNA–protein intermolecular distance restraints. In addition, hydrogen-bond constraints between protein and RNA were used based on characteristic chemical-shift perturbations observed upon RNA binding for amide and carbonyl of the Ala150 backbone (using 20 \(^{1}H\), 15\(^{N}\) HSQC and 3D HNCO spectra, respectively) and of carboxyl groups of the Asp136 and Asp139 side-chains (using the 3D HNCA spectrum). Oxygen acceptors were identified as described above. For each CYANA run, 50 independent structures were calculated. These 50 structures were refined with the SANDER module of AMBER 9.0 (72) by simulated annealing run in independent structures were calculated. These 50 structures were refined with the SANDER module of AMBER 9.0 (72) by simulated annealing run in implicit water using the ff99 force-field (73).

The best structures based on energy and NOE violations were analyzed with PROCHECK (74). The Ramachandran plot of the SRSF1 pseudo-RRM in complex with RNA indicates that 72.7% of the residues are in the most favored regions, 22.5% in the additional allowed regions, 3.8% in the generously allowed regions, and 1% in the disallowed regions. All of the figures showing structure were generated with MOLMOL (75).

ITC. ITC experiments were performed on a VP-ITC instrument (Microcal), calibrated according to the manufacturer’s instructions. Protein and RNA samples were dialyzed against the NMR buffer. Concentrations of proteins and RNAs were determined using optical-density absorbance at 280 and 260 nm, respectively. Then 20 \(\mu\)M of all of the tested RNAs were titrated with 400 \(\mu\)M of recombinant protein by 40 injections of 6 \(\mu\)L every 5 min at 40 °C. Raw data were integrated, normalized for the molar concentration, and analyzed using Origin 7.0 software according to a 1:1 RNA:protein ratio binding model.

Cell Culture and Transfections. We cultured HeLa cells in DMEM (Invitrogen) supplemented with 10% (vol/vol) FBS (Invitrogen). We used Fugene 6 transfection reagent (Invitrogen) to cotransfect 2.1 \(\mu\)g of pEGFP-IGF1 minigene (48) with 0.7 \(\mu\)g of pCG-T7::SRSF1 or its mutants, in 40 – 50% confluent HeLa cells grown on six-well plates, according to the manufacturer’s recommend-ations. For in vivo splicing assays performed with the FAS minigene, cells were cotransfected with 200 ng of FAS minigene plasmid and 1.5 \(\mu\)g of SRSF1 cDNA expression vectors (full-length or RRM2 derivatives). For the titration in Fig. 6A the following plasmid amounts were used: Upper, SRSF1 RRM2: 1 \(\mu\)g; hnRNP A1: 1, 2, and 4 \(\mu\)g; Lower, hnRNP A1: 1 \(\mu\)g; SRSF1 RRM2: 1, 2, and 4 \(\mu\)g. In Fig. 6B, Upper, SRSF1 RRM2: 2 \(\mu\)g; hnRNP A1: 0.5, 1, and 2 \(\mu\)g; Lower, hnRNP A1: 1 \(\mu\)g; SRSF1 RRM2: 1, 2, and 4 \(\mu\)g.

For splicing of endogenous SRSF1 targets, HeLa cells were transfected using FUGENE 6 (Promega) with 10 \(\mu\)g of pCG-T7-control, -SRSF1 or –RRM2 plasmids (76) together with 1 \(\mu\)g of pEGFP per 10 cm plate. Cells were trypsinized 48 h after transfection, and GFP-positive cells were sorted on a Becton Dickinson FACSAria IIU Cell Sorter.

In Vivo Splicing Assays. After 48 h, total RNA was isolated using TRIzol reagent (Invitrogen) and treated with DNase I (Promega). Reverse transcription was performed using ImProm-II reverse transcriptase (Promega) and oligo dT primers.

For assays performed with the IGF1 minigene, we used the cDNA templates to perform radioactive PCRs using vector-specific (pEGFP) primers, as described (77). For assays involving the FAS minigene, we performed RT-PCRs as described (78). To calculate the IGF1 exon 5 or exon 6 inclusion levels, the specific amplified PCR-amplicons were separated by native PAGE and quantified by phosphorimage analysis on a Fujifilm FLA-5100 instrument (Fuji Medical Systems USA, Inc.). We quantified the band intensities using MultiGauge software version 2.3 (Fujifilm) and normalized the values for the G+C content according to the DNA sequence. We further normalized the G+C corrected band intensities of individual isoforms in each lane to the corresponding intensities of GAPDH bands, as a control for loading. We then set the levels of exon 5 and exon 6 expressions from the empty vector lane to 100 and represented the expression levels of these isoforms upon overexpression of WT-SRSF1 or its mutant forms as fold changes.

For splicing of endogenous SRSF1 targets, radioactive touchdown PCR with \(\alpha\)-\(\[^{32}P\]\)-dCTP was used to amplify endogenous transcripts with primers described in Si Appendix Table S2. PCR products were separated by 8% native PAGE, and bands were quantified with a phosphorimager (Fujix Image Reader FLA-5100). The signal intensity was normalized first to the GC content of the amplified transcript and then to the sum of different isoform mRNAs or to \(\beta\)-actin if only one band was detected. The changes were expressed as percentages of exon inclusion, using empty vector as a control.

ACKNOWLEDGMENTS. We thank Dr. Lenka Skrisonoska for providing the Np3 clone; Prof. Stuart Wilson and Dr. Gaillaume Huitberg for providing the SRSF1 clone; Prof. Jamal Tazi and Dr. Francois Juge for providing the BS2 clone; Prof. Shern Chew for the IGF1 minigene; Markus Blatter for help with structure calculations; Dr. Mario Schubert, Dr. Christophe Maris, and Dr. Fred Damberger for help with NMR spectroscopy settings; Dr. Julien Boudet for advice during ITC measurements; and Prof. Giuseppe Biamonti and Dr. James Stevenin for useful discussions. We also thank the Schweizerischer Nationalen-Fonds-Nationalen Centers of Competence in Research Structural Biology and the European Alternative Splicing Network (EURASNET) for financial support of F-H.T.-A., EURASNET, Consolider RRNAREG, and Fundación Marcelino Botín for financial support of J.V., and the European Molecular Biology Organization for financial support of A.C., who also received a postdoctoral fellowship from the Fondation Suisse de Recherche sur les Maladies Musculaires, R.O.A., and A.R.K. were supported by National Institutes of Health Grant R37-GM24699.


