nature structural & molecular biology

Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1

Antoine Cléry[1], Sandrine Jayne[1], Natalya Benderska[2], Cyril Dominguez[1], Stefan Stamm[2,3] & Frédéric H-T Allain[1]

Tra2-β1 is a unique splicing factor as its single RNA recognition motif (RRM) is located between two RS (arginine-serine) domains. To understand how this protein recognizes its RNA target, we solved the structure of Tra2-β1 RRM in complex with RNA. The central 5′-AGAA-3′ motif is specifically recognized by residues from the β-sheet of the RRM and by residues from both extremities flanking the RRM. The structure suggests that RNA binding by Tra2-β1 induces positioning of the two RS domains relative to one another. By testing the effect of Tra2-β1 and RNA mutations on the splicing of *SMN2* exon 7, we validated the importance of the RNA-protein contacts observed in the structure for the function of Tra2-β1 and determined the functional sequence of Tra2-β1 in *SMN2* exon 7. Finally, we propose a model for the assembly of multiple RNA binding proteins on this exon.

Tra2-β is a crucial splicing factor that is expressed in all known metazoan genomes[1]. It is the homolog of the fly Transformer-2 protein, which regulates sex determination[2]. In humans, five isoforms[3] of the Tra2-β protein are produced by alternative splicing, with isoform 1, Tra2-β1, being the longest. This protein is ubiquitously expressed[4] and contains two RS domains separated by one RRM. The two RS domains sustain protein-protein interactions by interacting either with other RS-containing proteins or with each other[2,3,5]. RS domains also interact with RNA[6–9] and favor RNA-RNA base pairing[7]. This interaction seems to be reinforced by phosphorylation of the RS domain[8]. The RRM contains a functional protein phosphatase-1 (PP1) binding site and is responsible for the specific interaction with RNA[1]. *In vitro* selection revealed that this protein preferentially binds GA-rich sequences[10,11]. However, many splicing regulators, including some of the SR family, have also been shown to interact with similar purine-rich sequences[12,13]. Moreover, it has been suggested that Tra2-β1 acts *in vivo* through degenerate exonic splicing enhancer (ESE) sequences in the form GHVVGANR (H=A/C/U, V=G/A/C, R=A/G), which makes it difficult to precisely identify Tra2-β1 targets solely on the basis of RNA binding sequences[10].

The interaction of Tra2-β1 with an ESE located in exon 7 of the survival of motoneuron (*SMN*) gene is a well-documented and medically relevant example of the importance of this protein. Spinal muscular atrophy (SMA) is one of the leading genetic causes of death in children. It is caused by the loss of functional *SMN1*, but patients retain one or more copies of the related *SMN2* gene, which differs from *SMN1* by five nucleotide substitutions including a C-to-T conversion at position 6 of exon 7 (ref. 14). This substitution does not affect the reading frame, but it converts an ESE into a repressor and thereby prevents the inclusion of exon 7 (refs. 15,16). Consequently, most of the *SMN2* transcripts are translated into dysfunctional, truncated SMN proteins. There is a Tra2-β1 binding site in the middle of exon 7 in both *SMN1* and *SMN2* pre-mRNAs[17]. Current models propose that Tra2-β1 specifically interacts with this sequence and recruits two splicing factors, SRp30c and heterogeneous nuclear ribonucleoprotein (hnRNP) G (refs. 18,19), to activate the inclusion of exon 7 in both mRNAs[17,18]. In *SMN2*, this effect is thought to be counterbalanced by hnRNP A1[16], as well as by the presence of a weak 5′ splice site[20] and intronic silencers[21,22]. As a consequence, considerably less functional SMN proteins are produced from *SMN2* (ref. 14). This low production of functional SMN is insufficient to compensate for the homozygous loss of *SMN1* expression[23] and, depending on the number of copies of *SMN2*, patients suffer varying degrees of neuromuscular degeneration. As all SMA patients have at least one partially functional *SMN2* gene, improving the splicing regulation of *SMN2* offers a potential therapeutic strategy for increasing the level of functional SMN proteins[24]. Because Tra2-β1 is important for the inclusion of *SMN2* exon 7, a precise understanding of its mode of RNA recognition will aid with the development of rational therapies against SMA.

We now report on the solution structure of Tra2-β1 bound to an optimal GA-rich ESE and test the importance of the Tra2β1-RNA interaction for the splicing of exon 7 of *SMN2*.

RESULTS

Structure determination

To understand how the human Tra2-β1 protein specifically interacts with GA-rich ESEs, we expressed and purified a recombinant protein containing amino acids 106–201 of Tra2-β1, which includes the RRM (118–193) and an N-terminal His-tag (**Fig. 1a**). We used the 5′-AAGAAC-3′ RNA sequence as it is compatible with the

[1]Institute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology, Zürich, Switzerland. [2]Institute of Biochemistry, University of Erlangen-Nürnberg, Erlangen, Germany. [3]Department of Molecular and Cellular Biochemistry, Biomedical Biological Sciences Research Building, University of Kentucky, Lexington, Kentucky, USA. Correspondence should be addressed to F.H.-T.A. (allain@mol.biol.ethz.ch).

Received 26 February 2010; accepted 6 December 2010; published online 13 March 2011; corrected online 25 March 2011 (details online); doi:10.1038/nsmb.2001

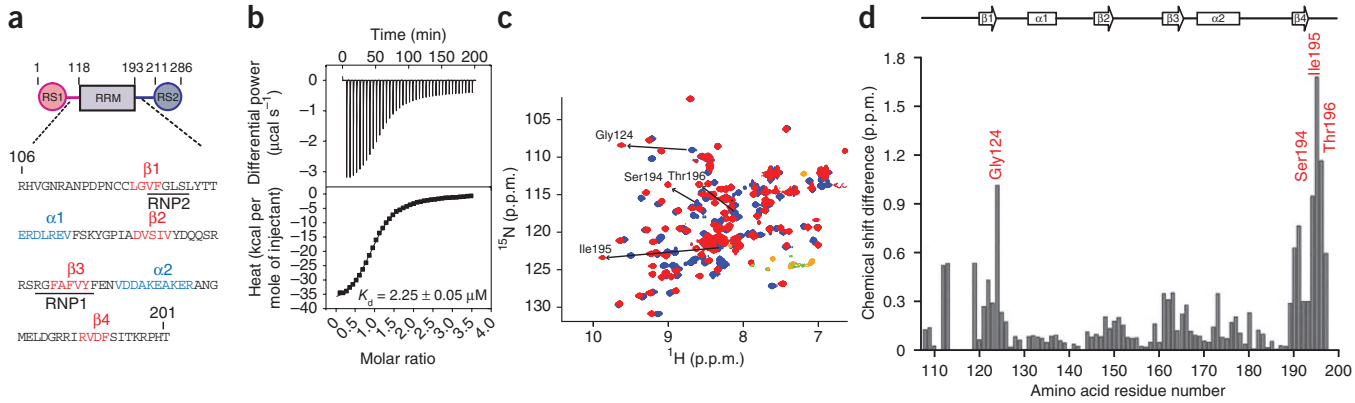

Figure 1 Study of the interaction between Tra2-β1 RRM and the 5′-AAGAAC-3′ RNA by ITC and NMR. (**a**) Schematic representation of the Tra2-β1 domains with linkers. The Tra2-β1 RRM is located between two RS domains. Amino acids delimitating each protein domain are indicated by their corresponding numbers according to the PDB sequence. The sequence of the recombinant protein used in this study is shown. Amino acids involved in the formation of β-strands and α-helices are red and blue, respectively. The conserved RNP1 and RNP2 motifs are indicated. (**b**) K_d determination of Tra2-β1 RRM in complex with the 5′-AAGAAC-3′ RNA by ITC. K_d is given ± s.d. (**c**) Superimposition of ^{1}H-^{15}N HSQC spectra representing NMR titration of the ^{15}N-labeled Tra2-β1 RRM with increasing amounts of unlabeled 5′-AAGAAC-3′ RNA. Titration was performed at 40 °C in NMR buffer. The peaks corresponding to the free and RNA-bound protein states (RNA:protein ratios of 0.3:1 and 1:1) are blue, orange and red, respectively. Negative peaks corresponding to the amides of arginine side chains in the free and RNA bound (1:1 ratio) states are green and orange, respectively. Black arrows indicate highest chemical shift perturbations (>0.9 p.p.m.) seen upon RNA binding. (**d**) Representation of the combined chemical shift perturbations ($\Delta\delta = [(\delta HN)^2 + (\delta N/6.51)^2]^{1/2}$) of Tra2-β1 RRM amides upon binding 5′-AAGAAC-3′ RNA as a function of the Tra2-β1 RRM amino acid sequence. The secondary structure elements of the RRM are shown at the top. The highest chemical shift perturbations (for Gly124, Ser194, Ile195 and Thr196) are indicated.

5′-GAARGARR-3′ consensus sequence found in all known Tra2-β1 targets[10], and provided high quality NMR spectra in complex with the protein. We verified that the His-tag did not interfere with the RNA interaction (**Supplementary Fig. 1**).

We achieved optimal NMR experimental conditions at 40 °C in a buffer containing 50 mM L-Glu, 50 mM L-Arg and 20 mM NaH_2PO_4 at pH 5.5. The use of arginine and glutamate salts increased the solubility and stability of RRMs for NMR studies[25]. Using isothermal

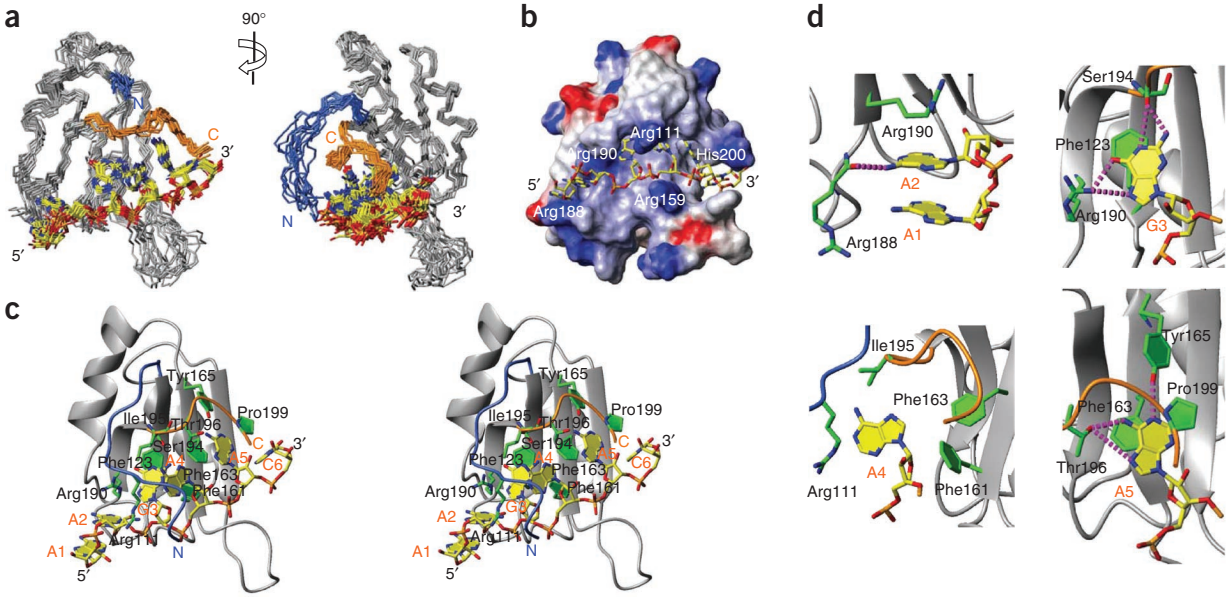

Figure 2 Overview of the solution structure of Tra2-β1 RRM in complex with the 5′-AAGAAC-3′ RNA. (**a**) Overlay of the 12 lowest-energy structures superimposed on the backbone of the protein-structured parts and heavy atoms of RNA. Protein backbone, gray; RNA heavy atoms, orange (phosphate atoms), yellow (C atoms), red (O atoms) and blue (N atoms). Only the ordered region of the RRM (residues 111–201) is shown. The N- and C-terminal regions of the RRM are shown in blue and orange, respectively. The N-terminal region of the protein has been truncated in the overlay on the left to avoid masking the RNA molecule. (**b**) Surface representation of the RRM (residues 106–201) and stick representation for the heavy atoms of the RNA molecule of the most representative structure of the complex. The protein surface is colored according to surface potential, with red and blue indicating negative and positive charges, respectively. The positively charged amino acids that are in contact with RNA are indicated. The RNA is colored as in **a**. (**c**) Structure of the complex represented in ribbon (protein backbone) and stick (RNA) representation. Color scheme as in **a**. Important protein side chains involved in RNA interactions are represented as green sticks. The N- and C-terminal extremities are blue and orange, respectively. (**d**) Molecular recognition of the 5′-AGAA-3′ RNA sequence by Tra2-β1 RRM. Single structures focus on the intermolecular interactions that are most commonly observed in the structures. Protein-RNA interactions involving A2, G3, A4 and A5 are represented. Color schemes as in **a**. Hydrogen bonds, purple dashed lines. Figures generated by MOLMOL[49].

Table 1 Structural statistics of the Tra2-β1 RRM in complex with AAGAAC RNA

	Protein		RNA
NMR distance and dihedral constraints			
Distance restraints			
Total NOE	1,107		112
Intra-residue	309		91
Inter-residue			
Sequential ($\|i-j\|=1$)	313		21
Nonsequential ($\|i-j\|>1$)	485		0
Hydrogen bonds	22		0
Protein–RNA intermolecular		93	
Total dihedral angle restraints			6
Protein			
ϕ	0		
ψ	0		
Nucleic acid			
Base pair			0
Sugar pucker			6
Backbone			0
Based on A-form geometry			0
Structure statistics			
Violations (mean and s.d.)			
Distance constraints (>0.3 Å) (Å)	0.365 ± 0.046		
Dihedral angle constraints (°)	0		
Max. dihedral angle violation (°)	0		
Max. distance constraint violation (Å)	0.45 ± 0.01		
Deviations from idealized geometry			
Bond lengths (Å)	0.004		
Bond angles (°)	1.67		
Average pairwise r.m.s. deviation[a] (Å)			
Protein			
Heavy	1.95 ± 0.28		
Backbone	1.31 ± 0.24		
RNA			
All RNA heavy			0.92 ± 0.47
Complex			
Protein and RNA heavy			1.87 ± 0.26

[a]Protein r.m.s. deviation was calculated using residues 111–201 for the ensemble of 12 refined structures. RNA r.m.s. deviation was calculated using residues 1–6 for the ensemble of 12 refined structures.

titration calorimetry (ITC) under these conditions, we determined a dissociation constant (K_d) of 2.25 μM for the RRM in complex with 5′-AAGAAC-3′ (**Fig. 1b**). NMR titration of the RNA into Tra2-β1 RRM showed that saturation was reached at a 1:1 ratio and that protein resonances experienced intermediate to fast exchange during the titration (**Fig. 1c**). We observed the largest chemical shift perturbations for residues from the β-sheet (**Fig. 1d**), which is the canonical surface of RRM-RNA interaction, and also for residues in the N- and C-terminal regions of the RRM.

We calculated the structure of Tra2-β1 RRM bound to RNA using 1,219 nuclear Overhauser effect (NOE)-derived distance restraints including 93 intermolecular NOEs. Twelve conformers represented the final ensemble of structures (**Fig. 2a**). The precision of the structure was high, with an r.m.s. deviation of 1.87 Å for all heavy atoms (**Table 1**).

Overview of the Tra2-β1 RRM structure in complex with RNA

Tra2-β1 RRM bound to RNA adopts a canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold (**Fig. 2**). The RNA is single-stranded with all sugar puckers adopting a C2′-*endo* conformation (evidence for which came from the strong H1′-H2′ correlation in a 2D total correlation spectroscopy (TOCSY) spectrum). Although the RNA is mainly bound by the β-sheet, amino acids from the N- and C-terminal regions of the RRM are also involved in RNA binding. The domain forms a positively charged groove that accommodates the RNA (**Fig. 2b**). Five positively charged residues—Arg111, Arg159, Arg188, Arg190 and His200—stabilize the complex through electrostatic interactions with the negatively charged phosphates of the RNA. In addition, several hydrophobic RNA-protein contacts contribute to the binding affinity. We found intermolecular stacking interactions between Arg190 and A2; Phe123 (RNP2 motif) and G3; and both Phe163 (RNP1 motif) and Pro199 and both sides of A5 (**Fig. 2c**). Finally, Phe161 (RNP1 motif) inserts between the sugars of G3 and A4, and Ile195 contacts A4 (**Fig. 2c**). To verify the importance of these interactions, we tested the effect of single alanine substitutions of the involved residues on the affinity of the RRM for RNA (**Table 2** and **Supplementary Fig. 2**). We verified by NMR that none of these point mutations affected the global fold of the RRM (**Supplementary Fig. 3**). In agreement with the structure, these substitutions decreased the affinity relative to the wild-type protein (**Table 2**).

The Tra2-β1 RRM specifically recognizes the 5′-AGAA-3′ sequence

Although six nucleotides are imbedded in the RRM, only the central AGAA motif is specifically recognized. At the 5′ end, A1 stacks with A2 to stabilize the 5′ end of the RNA (**Fig. 2d**). A2 is also specifically recognized by a hydrogen bond between the N6 amino proton and the backbone carbonyl oxygen of Arg188 (**Fig. 2d**). G3 is specifically recognized by four hydrogen bonds that involve the side chain of Arg190 and the backbone carbonyl oxygen of Ser194 (**Fig. 2d**). A guanine is rarely found at this position in RRM-RNA structures, especially in the context of RNA containing AG sequences[26]. The A4 base adopts an unusual *syn* conformation that results from the presence of the two aromatic residues (Phe161 and Phe163) on β3 (**Fig. 2d**) rather than an *anti* conformation in which this base would sterically interfere with these amino acids. The *syn* conformation of A4 is further stabilized by hydrophobic contacts with the side chains of Arg111 and Ile195 (**Fig. 2d**) and by an intramolecular hydrogen bond between the N7 atom of A4 and the amino group of G3 (**Fig. 2c**). A5 is the last nucleotide to be specifically recognized by the RRM with three intermolecular hydrogen bonds involving the hydroxyls of the side chains of Tyr165 and Thr196 (**Fig. 2d**).

We used ITC to test the effect of each nucleotide substitution in the 5′-AGAA-3′ motif on the affinity of Tra2-β1 RRM. The specific recognition of these four nucleotides was well supported by the decreased affinity of Tra2-β1 RRM for these RNA variants (**Table 2**). The structure indicates that whereas G3 and A5 are 'positively' selected by the specific interaction of functional groups in the bases with the protein backbone and side chains, A4 appears to be 'negatively' selected because the other bases cannot be accommodated in this binding pocket. Indeed, the 2-amino group of a guanine would interact unfavorably with the Arg111 side-chain and a *syn* conformation for pyrimidine is not energetically favored (**Fig. 2c**).

The ESEs that are targeted by Tra2-β1 usually contain several 5′-GA-3′ dinucleotides (**Supplementary Fig. 4**). Our results show that only one 5′-GA-3′ dinucleotide is specifically recognized by Tra2-β1. Moreover, on the basis of the structure and affinity measurements, we conclude that a 5′-NAGAAN-3′ (N can be any nucleotide) sequence is the minimal and optimal binding partner for the RRM of Tra2-β1.

Table 2 Binding affinity of wild-type and variants of Tra2-β1 RRM and 5′-AAGAAC-3′ RNA

Protein	K_d (μM)	Affinity factor
Wild type	2.25 ± 0.05	1
R111A	3.65 ± 0.05	1.65
F123A	> 30	> 13
V137A	2.25 ± 0.05	1
F161A	> 30	> 13
Y165F	12.85 ± 1.15	5.7
R190A	> 30	> 13
S194A	3.25 ± 0.05	1.45
I195A	12.3 ± 0.4	5.5
T196A	15 ± 1	6.7
P199A	7.1 ± 0.1	3.15
Y165F + T196A	22 ± 1	9.8
ΔN-Ter (116-201)	5.1 ± 0.3	2.3
ΔC-Ter (106-193)	> 30	> 13
RNA	**K_d (μM)**	**Affinity factor**
AAGAAC (WT)	2.25 ± 0.05	1
A**G**GAAC (A2G)	4.5	2
A**C**GAAC (A2C)	4.4 ± 0.1	2
A**U**GAAC (A2U)	4.5 ± 0.2	2
AA**A**AAC (G3A)	> 30	> 13
AA**C**AAC (G3C)	> 30	> 13
AA**U**AAC (G3U)	> 30	> 13
AAG**G**AC (A4G)	> 30	> 13
AAG**C**AC (A4C)	> 30	> 13
AAG**U**AC (A4U)	> 30	> 13
AAGA**G**C (A5G)	12.25 ± 0.45	5.5
AAGA**C**C (A5C)	> 30	> 13
AAGA**U**C (A5U)	> 30	> 13

K_d values determined by ITC at 40 °C in the buffer used for the NMR structure determination. The s.d. are indicated.

Mutations in the Tra2-β1 RRM affect splicing of the *SMN2* exon 7

The structure of the RRM bound to 5′-AAGAAC-3′ is the first example of a splicing factor specifically interacting with a purine-rich ESE, a type of sequence recognized by most SR or SR-like proteins. To test the functional importance of these protein-RNA interactions, we performed *in vivo* splicing assays using an *SMN2* reporter system[17] (**Fig. 3a**). Using β-actin as a loading control, we performed western blot analysis of transfected cell lysates and found that the protein expression reflected the amounts of transfected protein (**Fig. 3b**). After RNA isolation, we monitored skipping or inclusion of exon 7 by RT-PCR (**Fig. 3c**,**d** and **Supplementary Fig. 5**). As previously described[17], in the absence of ectopic Tra2-β1 expression, only about 25% of the spliced *SMN2* mRNAs included exon 7 (**Fig. 3c**,**d**), whereas overexpression of Tra2-β1 induces an increase in the inclusion of exon 7 of *SMN2* between 40% (1 μg) and 60% (3 μg; **Fig. 3c**,**d**).

We then tested five single amino acid mutants of Tra2-β1 that strongly affect the affinity of Tra2-β1 for RNA (**Table 2**) and found a strong decrease in inclusion of exon 7 (**Fig. 3d**). The percentage of inclusion was reduced (2–9% inclusion of exon 7) over that seen in the absence of ectopic protein expression (25% inclusion of exon 7; **Fig. 3d**). SRp30c and hnRNP G promote the inclusion of exon 7[18,19]. As both hnRNP G[27] and SRp30c[5] bind directly to Tra2-β1, it is likely that they form an exon enhancer complex with Tra2-β1 on *SMN2* pre-mRNA[18]. To test whether SRp30c and hnRNP G still interact with Tra2-β1 mutants, we performed coimmunoprecipitation experiments. The immunoprecipitations contained benzonase to prevent RNA-mediated interactions. All Tra2-β1 variants still bound to hnRNP G and SRp30c (**Supplementary Fig. 6**). These results suggest that overexpressed Tra2-β1 variants sequester hnRNP G and SRp30c, leading to the observed dominant-negative effect on the inclusion of exon 7. In addition, although hnRNP G and SRp30c probably stabilize the interaction between Tra2-β1 and exon 7, they cannot compensate for the decreased RNA binding affinity of the Tra2-β1 mutants. As a positive control, we also expressed the Val137A Tra2-β1 variant. The valine Val137 is located in the α1 helix of the RRM and has no contact with RNA. As expected, this point mutation does not affect RNA binding (**Table 2**) and exon 7 inclusion (**Fig. 3d**).

Together, these results show that the amino acid residues responsible for the specific recognition of 5′-AGAA-3′ *in vitro* are also required *in vivo* in the context of the full length Tra2-β1 for interacting with exon 7 of *SMN2*.

Characterization of the Tra2-β1 binding site in the SMN exon 7

Next, we investigated the effect of mutations in the putative Tra2-β1 binding site of *SMN2* exon 7 to identify the RNA sequence responsible for its function in activating exon 7 inclusion. Tra2-β1 interacts directly with the 5′-$A_{19}A_{20}A_{21}G_{22}A_{23}A_{24}G_{25}G_{26}A_{27}A_{28}G_{29}$-3′ sequence of *SMN* exon 7 (ref. 17; underlined in **Fig. 3a**). On the basis of our structural and biochemical data, we identified two putative binding sites in this sequence, namely $A_{21}G_{22}A_{23}A_{24}$ and $G_{25}G_{26}A_{27}A_{28}$. To evaluate the importance of these motifs for the regulation of the splicing of *SMN* exon 7 and to identify precisely the Tra2-β1 binding site, we introduced substitutions into the *SMN2* minigene[1]. Plasmids containing the wild-type and mutant genes were cotransfected into HEK293 cells with an expression vector encoding Tra2-β1 (+Tra2-β1; **Fig. 4**) or a plasmid without insert (–Tra2-β1; **Fig. 4**).

The substitution of each AGAA or GGAA motif by AAAA (**Fig. 4**; AGAAAAAA and AAAAGGAA) almost completely abolished exon 7 inclusion in the absence of Tra2-β1 overexpression. However, whereas overexpressed Tra2-β1 could not stimulate exon 7 inclusion in absence of both motifs (**Fig. 4**; AAAAAAAA variant), under the same conditions there was a significant increase in exon 7 inclusion in the presence of AGAA or GGAA (**Fig. 4**). These results indicate that both motifs can interact with Tra2-β1. The higher activation of exon 7 inclusion by AGAA probably reflects the twofold difference in Tra2-β1 affinity (**Table 2**). Accordingly, in the absence of the upstream AGAA motif, the GGAA to AGAA substitution (AAAAAGAA) induced an increase in exon 7 inclusion from 15% to 51% with overexpressed Tra2-β1 (**Fig. 4**).

As the AGAA motif was preferred for activation of exon 7 inclusion by Tra2-β1, we tested the functional importance of the GGAA motif in the presence of the upstream AGAA sequence. We introduced two substitutions in this motif that were expected to strongly decrease protein binding (GGAA to GUAA or GGGA; **Table 2**). Although these variants strongly affected the inclusion of exon 7 (**Fig. 4**; AGAAGUAA and AGAAGGGA), overexpressed Tra2-β1 could still activate exon 7 inclusion at a level similar to that seen with the wild-type RNA (**Fig. 4**). These results suggest that in the presence of the upstream AGAA sequence, the GGAA motif has a Tra2-β1-independent role in the inclusion of exon 7.

The AGGA motif in the CE9 element of the mammalian hnRNP A1 pre-mRNA has been shown to interact with SRp30c (ref. 28). As Tra2-β1 recruits SRp30c on exon 7 of *SMN*[19], the G_{25}GA motif downstream of the Tra2-β1 binding site might interact with SRp30c rather than with Tra2-β1. Although SRp30c was shown to interact inefficiently with *SMN* exon 7 (ref. 19), this contact could be stabilized by the Tra2-β1-RNA interaction.

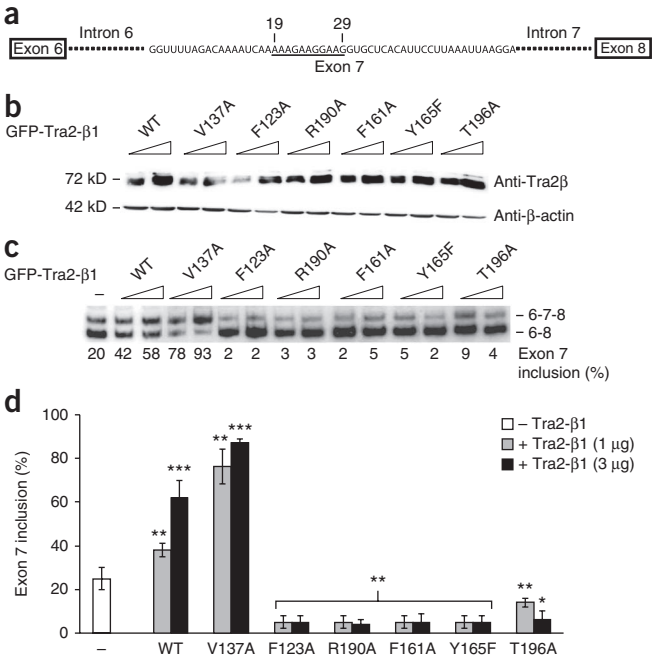

Figure 3 Effect of Tra2-β1 mutations on *SMN2* exon 7 splicing. The *SMN2* reporter minigene was cotransfected in HEK293 cells with increasing amounts (1 μg and 3 μg) of wild-type or mutant Tra2-β1 expression constructs. (**a**) Schematic representation of the *SMN2* minigene. The exon 7 sequence is shown and the previously reported Tra2-β1 interacting sequence is underlined[17]. (**b**) Western blot experiment showing the expression of the GFP-tagged proteins after cell transfection with 1 and 3 μg of Tra2-β1 (wild type and mutant) expression constructs. β-actin was used as a loading control. (**c**) RT-PCR amplification of mRNAs produced from the *SMN2* minigenes (wild type and mutants). The positions of PCR products with or without exon 7 are indicated on the right. The percentage of exon 7 inclusion in *SMN2* transcripts was determined using ImageQuant. The negative control corresponds to the percentage of exon 7 inclusion in the absence of ectopic Tra2-β1 expression. (**d**) Statistical representation of at least three independent experiments showing the strong negative effect of all the tested Tra2-β1 substitutions on *SMN2* exon 7 inclusion. Student's test comparing Tra2-β1 constructs with the negative control; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Error bars are s.d.

Tra2-β1 binding to SMN might change position of RS domains

A distinct feature of the Tra2-β1–RNA complex is the unusual manner in which the protein wraps around the RNA. Upon RNA binding, the two extremities of the protein interact with each other and cross each other above the AGAA sequence (**Fig. 2b**,**c**). Five residues from the extremities contact the RNA—Arg111 from the N-terminal region and Ser194, Ile195, Thr196 and Pro199 from the C-terminal region. Consistent with this unusual structural feature, deletion of the C-terminal extremity strongly decreases the affinity (from 2.25 μM to more than 30 μM) and deletion of the N-terminal extremity also diminishes binding affinity (from 2.25 μM to 5.1 μM) (**Table 2**).

To determine whether the positions of the N- and C-terminal regions outside the RRM in the structure (**Fig. 2c**) were induced by these RNA-protein interactions or preformed in the apo form of the protein we measured ^{1}H-^{15}N heteronuclear NOE values. These data show that both terminal regions are flexible in the free form, whereas they become rigid in presence of RNA (**Supplementary Fig. 7**). Accordingly, we found no interproton NOE involving residues from these two terminal regions in the free form of the protein, whereas we found unambiguous NOEs between residues from both extremities (between R111 and I195) in the complex. These data indicate that RNA binding influences the positions of protein regions outside the RRM.

Tra2-β1 is unique among SR proteins in that it contains two RS domains attached to the RRM. Our construct includes the linker between RS1 and the RRM and most of the amino acids that separate RS2 from the RRM (**Fig. 1a**). Our data indicate that the positions of the linkers and probably the positions of the adjacent RS domains are regulated by RNA binding. RS domains have been shown to interact with RNA[6–9,29] or with other RS-containing proteins[13]. We therefore tested the influence of truncated versions of Tra2-β1 without either the first (ΔRS1) or second (ΔRS2) RS domain on the inclusion of exon 7. These two protein variants are correctly localized in the nucleus[3]. These mutants caused an increase in exon 7 usage that was much lower than that induced by the wild-type protein (**Fig. 5a**). These data suggest that both RS domains contribute to *SMN2* splicing either by increasing the RNA binding affinity of Tra2-β1 or by interacting with other proteins that assemble on exon 7, or both. These results are consistent with earlier studies that showed that ΔRS1, a naturally occurring splice variant of Tra2-β1, had no effect on the splicing of Tra2-β1 exon 2 (ref. 10).

HnRNP G interacts with Tra2-β1 (ref. 27) and promotes the inclusion of exon 7. However, its RNA binding site has not been determined[18]. The structure of RBMY[30], a homologous protein to hnRNP G, has been solved in complex with RNA. This structure predicts that RBMY and members of the hnRNP G family bind sequence-specifically to 5′-CAA-3′ motifs. Systematic evolution of ligands by exponential enrichment (SELEX) and *in vivo* splicing analyses of hnRNP G suggested that it binds preferentially to CCA motifs[27]. There is no CCA motif and in exon 7 and only two CAA motifs, located upstream of the Tra2-β1 binding site. We therefore investigated

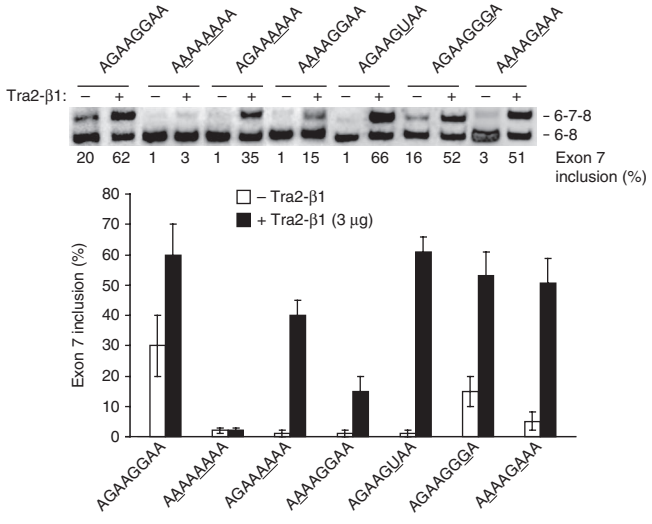

Figure 4 Effect of nucleotide substitutions in the Tra2-β1 binding site on *SMN2* exon 7 splicing. Wild-type and mutant *SMN2* minigene plasmids (1 μg) were cotransfected with 3 μg of empty vector (–) or a vector expressing human Tra2-β1 (+). RT-PCR amplification of mRNAs produced from the *SMN2* minigenes (wild type and mutants) is shown. The percentage of exon 7 inclusion in *SMN2* transcripts was determined using ImageQuant. The positions of PCR products with or without exon 7 are indicated on the right. Lower panel, data from at least three independent experiments. Nucleotide substitutions are underlined. Error bars are s.d.

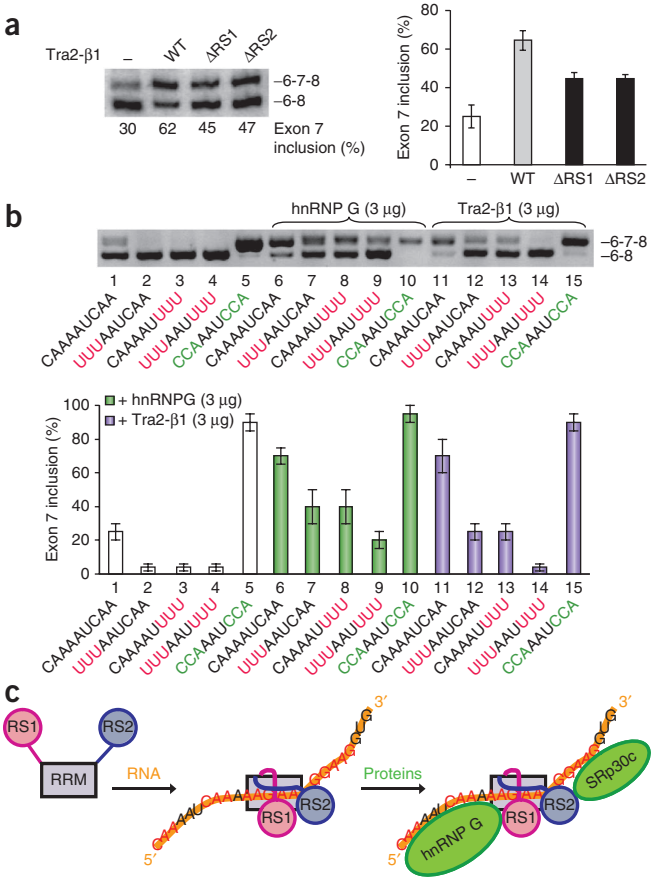

Figure 5 Binding of Tra2-β1 RRM might induce the positioning of the two RS domains on SMN exon 7 or the recruitment of hnRNP G and SRp30c. (**a**) Effect of deletion of RS1 or RS2 on *SMN2* exon 7 splicing. Wild-type *SMN2* minigene plasmid (1 μg) was cotransfected with 3 μg of empty vector (–) or a vector expressing wild-type, ΔRS1 or ΔRS2 Tra2-β1. RT-PCR amplification of mRNAs produced from the *SMN2* minigene is shown. The positions of PCR products with or without exon 7 are indicated on the right. Right panel, results of at least three independent experiments. (**b**) Effect on *SMN2* exon 7 splicing of UUU or CCA substitution (colored red and green, respectively) of the two CAA motifs located upstream of the Tra2-β1 *SMN* exon 7 binding site in the absence or presence of hnRNP G or Tra2-β1 co-expression. Wild-type and mutant *SMN2* minigene plasmids (1 μg) were cotransfected with 3 μg of empty vector (lanes 1–5) or a vector expressing wild-type hnRNP G (lanes 6–10) or Tra2-β1 (lanes 11–15). RT-PCR amplification of mRNAs produced from the *SMN2* minigenes is shown. (**c**) Model of the assembly of Tra2-β1, hnRNP G and SRp30c on SMN exon 7. Once the RRM of Tra2-β1 is bound to the RNA, the cross formed between the N and C termini might place RS domains near the RNA molecule, one upstream and the other downstream of the Tra2-β1 binding site. Our model proposes that the RS domains act as a platform to recruit hnRNP G and SRp30c on each side of the *SMN* exon 7 sequence bound by Tra2-β1. N-terminal (RS1) and C-terminal (RS2) RS domains are magenta and blue, respectively. The RNA is orange. The Tra2-β1 binding site and putative hnRNP G and SRp30c interacting sequences are red.

whether these CAA motifs could be binding sites for hnRNP G. When we mutated one or both CAA motifs, the inclusion of *SMN2* exon 7 was severely decreased (**Fig. 5b**, lanes 2–4). Moreover, when the CAA motifs were changed into CCA motifs, which are high-affinity sites for hnRNP G[27], the inclusion of exon 7 was markedly increased (**Fig. 5b**, lane 5). As expected, overexpression of hnRNP G increased the inclusion of exon 7 (**Fig. 5b**, lane 6); consistent with the importance of the CAA motifs for binding hnRNP G, this activator effect was progressively decreased when one or both CAA motifs were mutated (**Fig. 5b**, lanes 7–9). By contrast, the effect was increased by CCA mutations (**Fig. 5b**, lane 10). Finally, the fact that CAA mutations also affected the level of exon 7 inclusion upon overexpression of Tra2-β1 (**Fig. 5b**, lanes 11–15) confirmed the link between the upstream CAA motifs and the Tra2-β1 binding sequence. Both CAA motifs are in the first reading frame and encode a glutamine residue. Changing them into CCA would change the amino acid to proline and probably change the structure of the protein. It is therefore likely that Tra2-β1 stabilizes the binding of hnRNP G to a suboptimal binding site through one of the RS domains.

Together, these data support a model in which binding of RNA by Tra2-β1 would coordinate the assembly of multiple RNA-binding proteins along the RNA; for example, the assembly of Tra2-β1, SRp30c and hnRNP G on *SMN* exon 7 (refs 18,19; **Fig. 5c**) with hnRNP G binding upstream and SRp30c binding downstream of the Tra2-β1 binding site. This assembly process would contribute to an increase in both affinity and specificity for the exonic splicing enhancer of *SMN* exon 7 compared to the binding of the individual protein and ensure tight regulation of this key splicing event.

DISCUSSION

In solving the structure of the Tra2-β1 RRM in complex with 5′-AAGAAC-3′, we have shown how a human alternative-splicing factor specifically interacts with a purine-rich ESE at the molecular level. Only a handful of structures of RRM-containing proteins bound to purine-rich sequences have been solved[31,32] such as, for example, the poly-A binding protein (PABP). The Tra2-β1 RRM binds 5′-AAGAAC-3′ with molecular features that are similar to poly-A sequence recognition by RRM1 and RRM2 of PABP[33] (**Supplementary Fig. 8**). The positioning of the first three nucleotides (AAG) in Tra2-β1 resembles the positioning of the first three adenines in RRM2 of PABP. In addition, the protuberant nucleotide A4 resembles the similarly bulged adenine (A7) in RRM1 of PABP (**Supplementary Fig. 8**). As might be expected, there are also marked differences between these structures. In PABP, A7 is in contact with amino acids from the β2–β3 loop, whereas in Tra2-β1, A4 is in contact with residues Arg111 and Ile195 from the N- and C-terminal regions of the Tra2-β1 RRM. In addition, Tra2-β1 recognizes a specific guanine (G3), whereas PABP recognizes two adenines (A3 and A6). In this context, the presence of an arginine (Arg190) in β4 of Tra2-β1, compared to a methionine (Met85) in RRM1 or a phenylalanine (Phe169) in RRM2 of PABP, accounts for the differences in binding specificity (**Supplementary Fig. 8**). Finally, the positioning of the N- and C-terminal regions of the RRMs is different in each protein. Although both extremities contact the RNA in PABP RRM2 and in Tra2-β1 RRM, they go in opposite directions in PABP whereas they cross each other in Tra2-β1 (**Supplementary Fig. 8**).

Another noncanonical feature of the Tra2-β1–RNA complex that is not found in the PABP–RNA complex is the stabilization of the A4 *syn* conformation (**Supplementary Fig. 8**). Although this conformation is not often adopted by RNA nucleotides in complexes with RRMs, *syn* conformations have been reported for A2 of the 5′-CAUC-3′ RNA sequence bound by the SRp20 RRM[34], for A12 from the 5′-CAA-3′ triplet that specifically interacts with RBMY RRM[30] and for guanines bound by several RRMs[26]. It is unclear whether this purine in a *syn* conformation is part of a protein-RNA interaction signature that has a functional implication or rather whether this conformation is adopted to avoid a steric interaction with aromatic amino acid side chains on the β-sheet.

Several structures of RRMs complexed with nucleic acids have been solved[31,32]. The involvement of the N-terminal (CUG-BP1, CBP20)[35,36] or C-terminal (hnRNP A1, PABP, PTB) regions of the RRM[33,37,38] in RNA binding is typical and allows both affinity and specificity to be increased[31,32]. However, involvement of both extremities from a single RRM as seen here is unusual, especially with the positioning of the two extremities relative to one another. In the structure of the PTB RRM4 bound to RNA, both extremities interact and cross each other in a similar fashion to that seen in Tra2-β1 (ref. 38). However, in the PTB RRM4, the extremities are already positioned in the free state[39] for Tra2-β1, this positioning is induced by RNA binding as shown by the ^{1}H-^{15}N NOE data (**Fig. 2c** and **Supplementary Fig. 7**).

Tra2-β1 recognizes the AGAA motif *in vitro* and *in vivo*

The structure of the Tra2-β1 RRM–RNA complex and its affinity measurements showed that 5′-AGAA-3′ is the optimal recognition sequence for Tra2-β1 binding. Accordingly, this motif is present in most of the pre-mRNAs targeted by this protein[1] (**Supplementary Fig. 4**) and RNAs that contained several 5′-GAA repetitions were selected *in vitro* by SELEX experiments with the full-length protein[11]. Our results show that only one GAA motif is bound by one Tra2-β1 RRM. The repetition of this motif seen in most of the ESEs targeted by this protein (**Supplementary Fig. 4**) probably increases the efficiency of Tra2-β1 recruitment by RNA either by allowing multiple binding registers for Tra2-β1 or by preventing the formation of secondary structures. In addition, the flanking purine residues might help to avoid binding around the targeted motif by other protein competitors that otherwise interact with different sequences.

Previous studies suggested that there was a binding site for Tra2-β1 in *SMN* exon 7 at 5′-$G_{25}G_{26}A_{27}$-3′ rather than 5′-$G_{22}A_{23}A_{24}$-3′ on the basis of UUU substitutions but not of direct binding examination by cross-link[17]. Our structure determination and structural work based on single nucleotide substitutions clearly show that the 5′-$A_{21}G_{22}A_{23}A_{24}$-3′ sequence is the principal Tra2-β1 binding site that is responsible for the inclusion of exon 7 of *SMN2* (**Fig. 4**). We interpret the previously published results on the basis of the expectation that the introduction of a UUU sequence in place of the Tra2-β1 binding site is likely to create an accessible new binding site for other splicing activators that would compensate for the loss of Tra2-β1 interaction. For example, U-rich sequences were required for the interaction of hnRNP C1-C2 with the *SMN1* intron 6–exon 7 boundary and this protein can activate the inclusion of *SMN1* exon 7 (ref. 40).

In agreement with the structure, the amino acids in the RRM of Tra2-β1 that are involved in the RNA interaction are conserved in protein homologs (**Supplementary Fig. 9**). Our results show the important role of the N- and C-terminal regions in RNA binding both in the structure of the complex (**Fig. 2c**) and in affinity measurements (**Table 2**) and explains why certain residues outside the RRM are strongly conserved. This amino acid conservation is also consistent with the finding that these protein homologs of Tra2-β1 bind RNA with the same specificity. In agreement with this notion, the *Drosophila* Tra2 protein interacts specifically with a GA-rich sequence, the purine-rich enhancer (PRE) in exon 4 of the *dsx* pre-mRNA[41]. Although this PRE does not contain the 5′-AGAA sequence, it includes two consecutive 5′-AAGGAC-3′ motifs separated by a single A. This sequence differs from the one used in our structure by only one A4-to-G substitution. As all the residues involved in the recognition of A4 (Arg111 and Ile195; **Fig. 2c**) are conserved in human and *D. melanogaster* proteins (**Supplementary Fig. 9**), we predict that, as determined for human Tra2-β1 (**Table 2**), the *Drosophila* Tra2 protein binds the 5′-AAGGAC-3′ sequence with low affinity. However, the repetition of this motif in the *dsx* PRE probably compensates for this apparently suboptimal RNA binding sequence.

Implications of the structure for genetic diseases

Despite the considerable amount of data published over the past 10 years it is still unclear how splicing factors involved in the splicing regulation of *SMN* exon 7 act together. We still need to identify their SMN RNA targets, to determine how many bind the exon at the same time, and to find out whether they interact together, bind in a timely coordinated manner or compete with one another. We have studied this regulation from a structural point of view and characterized, at the atomic level, the RNA and protein elements that are required for the recruitment of the principal activator of the inclusion of *SMN2* exon 7. This level of understanding is essential because single substitutions (C6 T or G25C) in exon 7 give raise to totally different splicing regulation of the corresponding exon 7 and influence the severity of SMA[42–45]. It will now be possible to predict the effects of mutations in other ESEs bound by this protein perhaps to better understand the origins of other diseases. For example, one can now explain how a 3-nucleotide deletion, found in exon 10 of *MAPT* (which encodes Tau) in patients with FTDP-17 (an autosomal dominant hereditary neurodegenerative disorder)[46,47] can affect the splicing of this exon; it converts a 5′-AAUAAGAAG-3′ ESE recognized by Tra2-β1 into a suboptimal 5′-AAUAAG-3′ binding site[48] in which the specific binding sequence has been deleted. Finally, our structure and functional data allowed us to propose a precise model of how and where Tra2-β1, SRp30c and hnRNP G assemble on the *SMN* exon 7 (**Fig. 5c**), a model that can now be tested by new structural and functional studies.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. We deposited the chemical shifts of Tra2β1 RRM–AAGAAC to the BioMagResBank under accession number 16920. We have deposited the coordinates of the RRM–AAGAAC structures to the Protein Data Bank under accession number 2KXN.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank J. Stevenin and D. Watt for reading the manuscript; J. Hall and M. Zimmermann for labeled RNA; the SNF-NCCR Structural Biology and EURASNET for financial support to F.H.-T.A.; the Muscular Dystrophy Association for support of S.S. and the European Molecular Biology Organization for a postdoctoral fellowship to A.C.

AUTHOR CONTRIBUTIONS

F.H.-T.A. and S.S. designed the project; A.C. prepared protein and RNA samples for structural studies; A.C., C.D. and F.H.-T.A. analyzed NMR data; A.C. and C.D. did structure calculations; S.J., N.B. and A.C. did *in vivo* splicing assays; A.C. and C.D. did ITC measurements; N.B. conducted the coimmunoprecipitation experiments; all authors discussed the results and wrote and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Novoyatleva, T. *et al.* Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing. *Hum. Mol. Genet.* **17**, 52–70 (2008).
2. Dauwalder, B., Amaya-Manzanares, F. & Mattox, W. A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl. Acad. Sci. USA* **93**, 9004–9009 (1996).
3. Nayler, O., Cap, C. & Stamm, S. Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. *Genomics* **53**, 191–202 (1998).
4. Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hubener, M. & Stamm, S. Activity-dependent regulation of alternative splicing patterns in the rat brain. *Eur. J. Neurosci.* **11**, 788–802 (1999).
5. Beil, B., Screaton, G. & Stamm, S. Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol.* **16**, 679–690 (1997).
6. Hertel, K.J. & Graveley, B.R. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem. Sci.* **30**, 115–118 (2005).
7. Shen, H. & Green, M.R. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol. Cell* **16**, 363–373 (2004).
8. Shen, H. & Green, M.R. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev.* **20**, 1755–1765 (2006).
9. Shen, H. & Green, M.R. RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat. Struct. Mol. Biol.* **14**, 597–603 (2007).
10. Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* **13**, 509–524 (2004).
11. Tacke, R., Tohyama, M., Ogawa, S. & Manley, J.L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93**, 139–148 (1998).
12. Tacke, R. & Manley, J.L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14**, 3540–3551 (1995).
13. Bourgeois, C.F., Lejeune, F. & Stevenin, J. Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **78**, 37–88 (2004).
14. Burghes, A.H. & Beattie, C.E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci.* **10**, 597–609 (2009).
15. Cartegni, L. & Krainer, A.R. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* **30**, 377–384 (2002).
16. Kashima, T. & Manley, J.L. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* **34**, 460–463 (2003).
17. Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J. & Wirth, B. Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc. Natl. Acad. Sci. USA* **97**, 9618–9623 (2000).
18. Hofmann, Y. & Wirth, B. hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum. Mol. Genet.* **11**, 2037–2049 (2002).
19. Young, P.J. *et al.* SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Hum. Mol. Genet.* **11**, 577–587 (2002).
20. Singh, N.N., Androphy, E.J. & Singh, R.N. *In vivo* selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA* **10**, 1291–1305 (2004).
21. Miyajima, H., Miyaso, H., Okumura, M., Kurisu, J. & Imaizumi, K. Identification of a cis-acting element for the regulation of SMN exon 7 splicing. *J. Biol. Chem.* **277**, 23271–23277 (2002).
22. Singh, N.N., Shishimorova, M., Cao, L.C., Gangwani, L. & Singh, R.N. A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy. *RNA Biol.* **6**, 341–350 (2009).
23. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
24. Horne, C. & Young, P.J. Is RNA manipulation a viable therapy for spinal muscular atrophy? *J. Neurol. Sci.* **287**, 27–31 (2009).
25. Golovanov, A.P., Hautbergue, G.M., Wilson, S.A. & Lian, L.Y. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126**, 8933–8939 (2004).
26. Auweter, S.D., Oberstrass, F.C. & Allain, F.H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943–4959 (2006).
27. Heinrich, B. *et al.* Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA. *J. Biol. Chem.* **284**, 14303–14315 (2009).
28. Simard, M.J. & Chabot, B. SRp30c is a repressor of 3′ splice site utilization. *Mol. Cell. Biol.* **22**, 4001–4010 (2002).
29. Shen, H., Kan, J.L. & Green, M.R. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol. Cell* **13**, 367–376 (2004).
30. Skrisovska, L. *et al.* The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.* **8**, 372–379 (2007).
31. Cléry, A., Blatter, M. & Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
32. Maris, C., Dominguez, C. & Allain, F.H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
33. Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845 (1999).
34. Hargous, Y. *et al.* Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.* **25**, 5126–5137 (2006).
35. Mazza, C., Segref, A., Mattaj, I.W. & Cusack, S. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* **21**, 5548–5557 (2002).
36. Tsuda, K. *et al.* Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3. *Nucleic Acids Res.* **37**, 5151–5166 (2009).
37. Ding, J. *et al.* Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115 (1999).
38. Oberstrass, F.C. *et al.* Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054–2057 (2005).
39. Vitali, F. *et al.* Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **25**, 150–162 (2006).
40. Irimura, S. *et al.* HnRNP C1/C2 may regulate exon 7 splicing in the spinal muscular atrophy gene SMN1. *Kobe J. Med. Sci.* **54**, E227–E236 (2009).
41. Lynch, K.W. & Maniatis, T. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* **9**, 284–293 (1995).
42. Prior, T.W. *et al.* A positive modifier of spinal muscular atrophy in the SMN2 gene. *Am. J. Hum. Genet.* **85**, 408–413 (2009).
43. Vezain, M. *et al.* A rare SMN2 variant in a previously unrecognized composite splicing regulatory element induces exon 7 inclusion and reduces the clinical severity of spinal muscular atrophy. *Hum. Mutat.* **31**, E1110–E1125 (2010).
44. Lorson, C.L., Hahnen, E., Androphy, E.J. & Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* **96**, 6307–6311 (1999).
45. Monani, U.R. *et al.* A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* **8**, 1177–1183 (1999).
46. D'Souza, I. *et al.* Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc. Natl. Acad. Sci. USA* **96**, 5598–5603 (1999).
47. Rizzu, P. *et al.* High prevalence of mutations in the microtubule-associated protein tau in a population study of frontotemporal dementia in the Netherlands. *Am. J. Hum. Genet.* **64**, 414–421 (1999).
48. Jiang, Z. *et al.* Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2 beta. *J. Biol. Chem.* **278**, 18997–19007 (2003).
49. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).

ONLINE METHODS

Preparation of the RNA–protein complexes. The Tra2-β1 RRM recombinant protein, fused to an N-terminal 6×His tag, was overexpressed at 37 °C in *Escherichia coli* BL21 (DE3) codon plus cells in minimal M9 medium containing 1 g l^{-1} $^{15}NH_4Cl$ and 4 g l^{-1} glucose (for ^{15}N-labeled protein) or 1 g l^{-1} $^{15}NH_4Cl$ and 2 g l^{-1} ^{13}C-glucose (for ^{15}N- and ^{13}C-labeled protein). The protein was purified by two successive nickel affinity chromatography (QIAGEN) steps (see **Supplementary Methods**), then dialyzed against the NMR buffer (50 mM L-Glu, 50 mM L-Arg, 0.05% β-mercaptoethanol and 20 mM NaH_2PO_4 at pH 5.5), and concentrated to 0.8 mM with a 10 kDa molecular mass cutoff Centricon device (Vivascience).

Wild-type and mutant RNA oligonucleotides were purchased from Dharmacon, deprotected according to the manufacturer's instructions, desalted using a G-15 size exclusion column (Amersham), lyophilized and resuspended in the NMR buffer.

Tra2-β1–RNA complexes were formed in the NMR buffer at a RNA:protein ratio of 1:1, at a 0.8 mM concentration and in a final volume of 250 µl.

NMR measurement. All NMR measurements were performed in the 50 mM L-Glu, 50 mM L-Arg, 0.05% β-mercaptoethanol and 20 mM NaH_2PO_4 at pH 5.5 buffer at 313 K using Bruker AVIII-500 MHz and 700 MHz equipped with a cryoprobe, AVIII-600 MHz and Avance-900 MHz spectrometers. Data were processed using Topspin 2.0 (Bruker) and analyzed with Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Protein sequence-specific backbone and side-chain assignments were achieved using 2D ^1H-^{15}N HSQC, 2D ^1H-^{13}C HSQC, 3D HNCA, 3D CBCACONH, 3D HcCH TOCSY, 3D NOESY ^1H-^{15}N HSQC and 3D NOESY ^1H-^{13}C HSQC aliphatic[50]. Aromatic proton assignments were performed using 2D ^1H-1 H TOCSY and 3D NOESY ^1H-^{13}C HSQC aromatic.

RNA resonance assignments in complex with the Tra2-β1 RRM were performed using 2D ^1H-1 H TOCSY, 2D ^1H-1 H NOESY and 2D ^{13}C 1F-filtered 2F-filtered NOESY[51] in 100% D_2O. These assignments were later confirmed using 5′-AAGAAC-3′ RNAs with chemically synthesized ^{13}C-labeled sugars of A_2, G_3, A_4 and A_5 (ref. 52).

Intermolecular NOEs were obtained using 2D ^1H-1 H NOESY and 3D ^{13}C 1F-edited 3F-filtered HSQC-NOESY[53] using unlabeled RNA and ^{15}N- and ^{15}N-^{13}C-labeled proteins, respectively. Intermolecular NOEs between the imino proton of the RNA G3 nucleotide and protein protons were obtained using 2D ^1H-1 H NOESY at 275 K in H_2O.

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.

Relaxation experiments were performed on a Bruker AVIII-500 MHz equipped with a cryoprobe, using a ^{15}N-labeled Tra2-β1 RRM sample at a concentration of 0.5 mM. The ^1H-^{15}N heteronuclear NOE was recorded in an interleaved fashion, recording alternatively one increment for the reference and one for the NOE spectrum. A relaxation delay of 2 s and a 1 H presaturation delay of 3 s were used in the NOE experiment while a 5 s relaxation delay was used in the reference experiment.

Structure calculation and refinement. AtnosCandid software[54,55] was used to generate preliminary structures and a list of automatically assigned NOE distance constraints for the Tra2-β1 RRM in complex with RNA. Peak peaking and NOE assignments were performed using 3D NOESY (^{15}N- and ^{13}C-edited) spectra and 2D ^1H-1 H NOESY spectrum recorded in 100% D_2O. Additionally, hydrogen bond constraints were based on hydrogen-deuterium exchange experiments on the amide protons. The oxygen acceptors were identified on the basis of 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[54] by adding the manually assigned intramolecular RNA and RNA-protein intermolecular distance restraints. For each cyana run, 50 independent structures were calculated. These 50 structures were refined with the SANDER module of AMBER 7.0 (ref. 56) using the simulated annealing protocol described previously[57]. The 12 best structures based on energy and NOE violations were analyzed with PROCHECK[58]. The Ramachandran plot of the Tra2-β1 RRM in complex with RNA indicates that 63.4% of the residues are in the most favored regions, 30.6% in the additional allowed regions, 4.5% in the generously allowed regions and 1.5% in the disallowed regions.

Isothermal titration calorimetry. ITC experiments were performed on a VP-ITC instrument (Microcal). The calorimeter was 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. Tested RNAs (20 µM) were titrated with 400 µM of wild-type and mutated versions of the Tra2-β1 RRM by 50 injections of 5 µl every 5 min at 40 °C. Raw data were integrated, normalized for the molar concentration and analyzed using Origin 7.0 according to a 1:1 RNA:protein ratio binding model.

Cell culture and plasmids. Human embryonic kidney HEK293 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; GibcoBRL) supplemented with 10% FCS (GibcoBRL). The pCI-*SMN2* recombinant plasmid containing the *SMN2* minigene was as described[17]. The mammalian vectors expressing GFP-tagged Tra2-β1 or FLAG-tagged Tra2-β1 were obtained by cloning the Tra2-β1 ORF, amplified by PCR, in frame into pEGFP-C2 or pcDNA3.1, respectively[10]. The FLAG-tagged Tra2-β1 ΔRS1 and ΔRS2 variants were PCR amplified from the Tra2-β1 ORF and cloned in frame into the pcDNA3.1 plasmid to express proteins containing amino acids 106–286 and 1–211, respectively. The mammalian vector expressing FLAG-tagged hnRNP G was obtained using the same procedure as for the wild-type Tra2-β1 clones. Tra2-β1 and *SMN2* mutants were created by site-directed mutagenesis using specific primers.

***In vivo* splicing assay.** *In vivo* splicing assays were carried out as described in **Supplementary Methods**. The ratio of exon inclusion to exon skipping was determined by using ImageQuant. Means and s.e. were calculated from three independent experiments.

Coimmunoprecipitation and immunoblots. Coimmunoprecipitation and immunoblotting were performed as described[59] (see **Supplementary Methods** for more details). Proteins were detected by immunoblotting using antibodies to Tra2β (pan Tra2β)[4], SRp30c (O-21) (sc-134036, Santa-Cruz Biotechnology) and hnRNP G[27].

50. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158 (1999).
51. Peterson, R.D., Theimer, C.A., Wu, H. & Feigon, J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR* **28**, 59–67 (2004).
52. Wenter, P., Reymond, L., Auweter, S.D., Allain, F.H. & Pitsch, S. Short, synthetic and selectively 13C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes. *Nucleic Acids Res.* **34**, e79 (2006).
53. Lee, W., Revington, M.J., Arrowsmith, C. & Kay, L.E. A pulsed field gradient isotope-filtered 3D 13C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.* **350**, 87–90 (1994).
54. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002).
55. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227 (2002).
56. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
57. Auweter, S.D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
58. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
59. Hartmann, A.M., Nayler, O., Schwaiger, F.W., Obermeier, A. & Stamm, S. The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell* **10**, 3909–3926 (1999).