

Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation

Ahmed Moursy, Frédéric H.-T. Allain* and Antoine Cléry*

Institute for Molecular Biology and Biophysics, Swiss Federal Institute of Technology (ETH), 8093 Zürich, Switzerland

Received July 26, 2013; Revised March 07, 2014; Accepted March 12, 2014

ABSTRACT

Regulation of SMN2 exon 7 splicing is crucial for the production of active SMN protein and the survival of Spinal Muscular Atrophy (SMA) patients. One of the most efficient activators of exon 7 inclusion is hnRNP G, which is recruited to the exon by Tra2- β 1. We report that in addition to the C-terminal region of hnRNP G, the RNA Recognition Motif (RRM) and the middle part of the protein containing the Arg-Gly-Gly (RGG) box are important for this function. To better understand the mode of action of hnRNP G in this context we determined the structure of its RRM bound to an SMN2 derived RNA. The RRM interacts with a 5'-AAN-3' motif and specifically recognizes the two consecutive adenines. By testing the effect of mutations in hnRNP G RRM and in its putative binding sites on the splicing of SMN2 exon 7, we show that it specifically binds to exon 7. This interaction is required for hnRNP G splicing activity and we propose its recruitment to a polyA tract located upstream of the Tra2- β 1 binding site. Finally, our data suggest that hnRNP G plays a major role in the recruitment of the Tra2- β 1/hnRNP G/SRSF9 trimeric complex to SMN2 exon 7.

INTRODUCTION

Spinal Muscular Atrophy (SMA) is an inherited disease characterized by degeneration of the spinal cord α -motor neurons, which results in a system-wide muscle wasting (1). SMA is considered one of the most frequent genetic causes of infantile death with an incidence rate of 1 in 6000 (2,3). It is caused by the genetic homozygous inactivation of Survival of Motor Neuron-1 (SMN1) gene (4). In humans, another copy of the SMN gene (SMN2) is present, which is nearly identical to SMN1 with only five nucleotide substitutions including a silent cytosine to thymine (C to T) mutation at position +6 of exon 7 (5). This point mutation inactivates an Exonic Splicing Enhancer (ESE) resulting in exon 7 skipping in most SMN2 transcripts (6,7). This splic-

ing isoform encodes an unstable truncated form of the SMN protein (8). Consequently, the amount of functional SMN proteins produced from SMN2 gene is not sufficient to compensate for the absence of SMN1 gene expression and maintain functional motor neurons (8,9). Importantly, all SMA patients have at least one intact copy of the SMN2 gene in their genome (3,5). Acting on the splicing regulation of SMN2 to favor the inclusion of exon 7 is therefore a promising strategy to increase the cellular level of functional SMN proteins and develop a treatment for SMA patients (1,10–11).

Several positive and negative regulators of SMN2 exon 7 splicing have been identified (1). One of the most efficient activators of exon 7 inclusion is the protein hnRNP G (12). It was proposed that hnRNP G is recruited to the SMN2 pre-mRNA by its interaction with another splicing factor named Tra2- β 1 (12–14). The structure of Tra2- β 1 RRM bound to RNA was recently determined by NMR showing that this factor recognizes specifically a 5'-AGAA-3' motif within an ESE located at position +21 of exon 7 (14,15). Both proteins act in synergy and can activate exon 7 inclusion to up to 80% when overexpressed simultaneously, a level that could not be reached when each protein was overexpressed separately (12,16). Understanding the mode of interaction of this heterodimer with RNA at the molecular level would then facilitate the development of therapeutic methods that stabilize its binding to exon 7.

HnRNP G belongs to the heterogeneous nuclear ribonucleoproteins family and is encoded by the RBMX gene located on the X chromosome (17). Although this protein is ubiquitously expressed (18,19) its expression level is variable and tissue-dependent (20). hnRNP G has multiple functions. In addition to SMN2, it regulates the splicing of dystrophin, α -tropomyosin (20), and Tau pre-mRNAs (21,22). Moreover, hnRNP G was shown to be involved in the transcription regulation of SREBP-1c (23,24) and GnRH1 (25). Several reports have also linked hnRNP G to cancer as it suppresses tumor growth at least in part by up-regulating the transcription of the tumor suppressor Txnip (26,27) or by modulating apoptosis (28). Other functions of hnRNP G also include the regulation of the neural development of frog (29) and zebrafish embryos (30) and the extracellular

*To whom correspondence should be addressed. Tel: +41 (0) 44 633 07 23; Fax: +41 (0) 44 633 12 94; Email: aclery@mol.biol.ethz.ch
Correspondence may also be addressed to Frédéric H.-T. Allain. Tel: +41 (0) 44 633 07 40; Fax: +41 (0) 44 633 12 94; Email: allain@mol.biol.ethz.ch

release of TNFR1, a receptor mediating the inflammatory actions of TNF (31). Finally, it was reported that hnRNP G is a regulator of sister chromatids cohesion (32) and that its expression is enhanced by p53 in response to DNA damage (33,34).

HnRNP G is composed of an N-terminal canonical RNA Recognition Motif (RRM) followed by a succession of motifs, namely an RGG box with three Arg–Gly–Gly repeats (17), a Nascent Targeting Domain (NTD) and a C-Terminal Domain (CTD) containing a SRGY box, Arg–Ser (RS) repeats (35) and a second RNA binding site (C-RBD) (36) (Figure 1A). Although the C-RBD part of the CTD was proposed to bind a 5'-GGAAA-3' capped stem-loop (36), the RRM is believed to be primarily responsible for the binding of hnRNP G to RNA (37). In the context of SMN2 exon 7 splicing regulation the CTD was reported to interact with Tra2- β 1 (12). To date, the specificity of RNA recognition by hnRNP G RRM remains elusive. This domain shares 88% sequence similarity with the RRM of its paralogue in testis RBMY. The structure of RBMY RRM bound to RNA was determined and showed that it interacts specifically with 5'-CAA-3' capped RNA stem-loops (38). Most residues involved in RNA recognition are conserved in hnRNP G suggesting that the protein could also recognize 5'-CAA-3' containing sequences. However, Systematic Evolution of Ligands by Exponential Enrichment (SELEX) experiments conducted with hnRNP G showed that its RRM binds single-stranded 5'-CCA-3' or 5'-CCC-3' containing sequences (37). Finally, another study suggested that hnRNP G could bind to a 5'-AAGU-3' motif (20). These contradictory data prevent the prediction of putative binding sites for this protein and the correct understanding of its functions in cells.

In this study, we determined the structure of hnRNP G RRM bound to RNA using NMR. The structure revealed that this domain specifically recognizes two consecutive adenines. This allowed us to identify a putative binding site for this protein on SMN2 and to propose a model in which hnRNP G binds specifically to exon 7 upstream of the Tra2- β 1 binding site. Finally, we identified the regions of hnRNP G that are required for its activity as a regulator of SMN2 exon 7 splicing.

MATERIALS AND METHODS

Expression and purification of the recombinant proteins

Escherichia coli BL21 (DE3) codon plus cells transformed with pET28a::hnRNP G RRM (residues 1–95), pET24b::GB1-hnRNP G RRM + RGG (residues 1–127) or pET28a::Tra2- β 1 RRM (residues 106–201) (14) were grown at 37°C in M9 minimal medium supplemented with 50 μ g/ml kanamycin, 50 μ g/ml chloramphenicol, 1 g/l $^{15}\text{NH}_4\text{Cl}$ and 4 g/l unlabeled or 2 g/l ^{13}C labeled glucose for ^{15}N or ^{15}N and ^{13}C labeled proteins, respectively. Proteins were purified by two successive nickel affinity chromatography (Qiagen[®]) steps, as previously described (14), dialyzed against the hnRNP G NMR buffer (50 mM NaCl, 20 mM NaH_2PO_4 , pH 5.5) or the Tra2- β 1 NMR buffer (50 mM L-Arg, 50 mM L-Glu, 0.05% β -mercaptoethanol, 20 mM NaH_2PO_4 pH 5.5) (14) for hnRNP G and Tra2- β 1 recombinant proteins, respectively. Concentration of recombinant

proteins was carried out using 10-kDa molecular mass cut-off Centricons (Vivascience[®]). The absence of RNases was confirmed using the RNase Alert Lab Kit (Ambion[®]).

ORF encoding full-length hnRNP G and Tra2- β 1 were cloned in pFX::SUMO vector (39) and expressed in MC1061 *E. coli* strain. The cells were grown in LB and induced at 18°C with 0.1 g of arabinose/l of culture. The insoluble fraction of lysed cells was dissolved in 6 M urea and the unfolded proteins were purified using Ni-NTA column (Qiagen[®]). The purified proteins were refolded by rapid dilution (20 \times) in the refolding buffer (880 mM L-arginine, 21 mM NaCl, 0.88 mM KCl, 55 mM Tris, pH 8.2) and subsequently dialyzed in the gel shift buffer (200 mM L-Arg, 200 mM L-Glu, 0.05% β -mercaptoethanol, 20 mM Na_2HPO_4 pH 7). The proteins were then concentrated and used for gel shift experiments.

Preparation of RNA–protein complexes

All RNA oligonucleotides were purchased from Dharmacon[®], de-protected according to manufacturer's instructions, lyophilized and resuspended in the corresponding NMR buffer. The 5'-GAGACAAAUCAAAAGAAG-3' RNA was transcribed *in vitro*, purified by HPLC and resuspended in the Tra2- β 1 NMR buffer.

The hnRNP G RRM–RNA complex was prepared in the hnRNP G NMR buffer at a protein:RNA stoichiometric ratio of 1:1 in a final volume of 250 μ l and at a final concentration of 1 mM.

The NMR titrations performed with hnRNP G RRM were done in the hnRNP G NMR buffer (Supplementary Figures S4 and S5), while those performed in the presence of hnRNP G and Tra2- β 1 RRM (Figure 4 and Supplementary Figure S6A) or hnRNP G RRM + RGG (Supplementary Figure S7) were done in the Tra2- β 1 NMR buffer with RNA and protein concentrations of 0.2 mM. In Figures 4 and S6A, increasing amounts of ^{15}N -labeled Tra2- β 1 or hnRNP G RRM were first added to the 5'-UCAAAAAGAAG-3' or 5'-GAGACAAAUCAAAAGAAG-3' RNA. After reaching saturation at stoichiometric ratio of 1:1, ^{15}N -labeled hnRNP G RRM was added in a stepwise manner to the Tra2- β 1/RRM complex until a final stoichiometric ratio of the three components of 1:1:1.

NMR experiments

All the NMR spectra were recorded in the hnRNP G NMR buffer except titrations with Tra2- β 1 RRM, which were recorded in the Tra2- β 1 NMR buffer. Experiments were recorded at 313 K using Bruker AVIII-500 MHz, 600 MHz, 700 MHz, Avance-900 MHz equipped with a cryoprobe, and AVIII-750 MHz spectrometers. Topspin 2.1 (Bruker[®]) was used for data processing and Sparky (<http://www.cgl.ucsf.edu/home/sparky/>) for data analysis.

Protein backbone assignment was achieved using 2D ^1H – ^{15}N HSQC, 3D HNCA, 3D CBCACONH and 3D HNCACB, while side chain assignments were achieved using 2D ^1H – ^{13}C HSQC, 3D HcccoNH TOCSY, 3D hCcoNH TOCSY, 3D NOESY ^1H – ^{15}N HSQC and 3D NOESY ^1H – ^{13}C HSQC aliphatic. Aromatic protons were assigned using

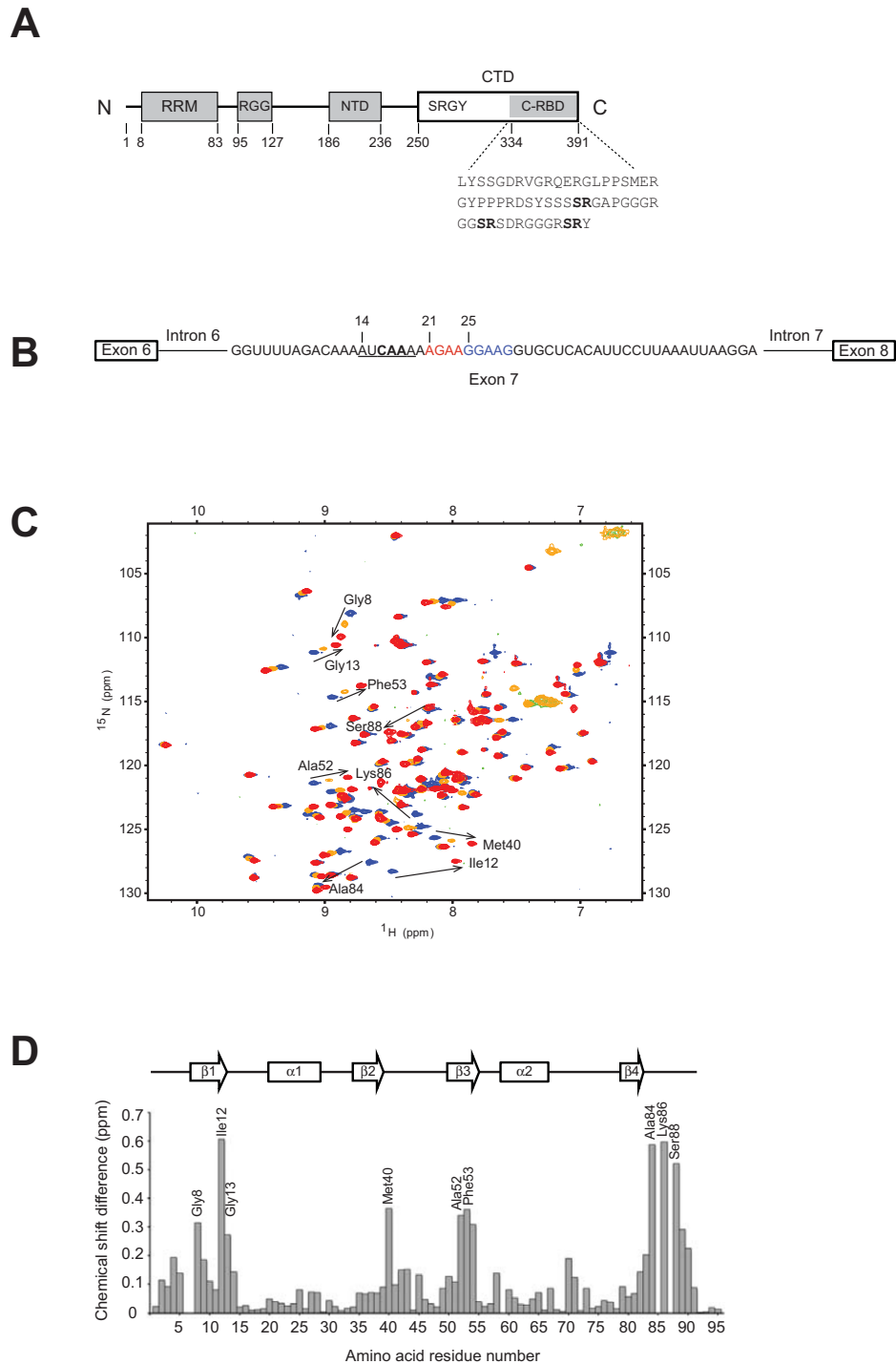


Figure 1. Study of the interaction of hnRNP G with its proposed binding site on SMN2 exon 7. **(A)** Schematic representation of hnRNP G domains composition. The RRM is located at the N-terminus followed by an RGG domain, a NTD, and a CTD. The CTD comprises a SRGY motif and a C-RBD. Amino acids located at the boundaries of each domain are numbered. The sequence of the C-RBD is shown with RS repeats marked in bold. **(B)** Schematic representation of the SMN2 minigene containing exon 6 to exon 8 with intermittent introns. The sequence of exon 7 is shown. The Tra2-β1 binding site is in red and the previously proposed hnRNP G binding site in bold (14). The RNA sequence tested for the interaction with hnRNP G RRM is underlined. The sequence in blue indicates another putative hnRNP G binding site. **(C)** Overlay of ^1H - ^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectra recorded during NMR titration of the ^{15}N labeled hnRNP G RRM with increasing amounts of the unlabeled 5'-AUCAAA-3' RNA. The titration was performed at 40°C in the hnRNP G NMR buffer. The peaks corresponding to 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 amides of arginine side chains in the free and RNA-bound (1:1 ratio) forms are green and orange, respectively. Black arrows indicate highest chemical shift perturbations observed upon RNA binding. **(D)** Representation of the combined chemical shift perturbations ($\Delta\delta = [(\delta\text{HN})^2 + (\delta\text{N}/6.51)^2]^{1/2}$) of hnRNP G amides upon binding to the 5'-AUCAAA-3' RNA at a ratio of 1:1 as a function of hnRNP G residue numbers. The corresponding secondary structure elements are represented at the top of the graph. The highest chemical shift perturbations annotated in (C) are indicated.

2D ^1H - ^1H TOCSY and 3D NOESY ^1H - ^{13}C HSQC aromatic (40).

RNA resonance assignments in complex with hnRNP G RRM were performed using 2D ^1H - ^1H TOCSY, natural abundance 2D ^1H - ^{13}C HSQC and 2D ^{13}C 1F-filtered 2F-filtered NOESY (41) in 100% D_2O . Intermolecular NOEs were obtained using 2D ^1H - ^1H NOESY and 3D ^{13}C 1F-edited 3F-filtered HSQC-NOESY (42) in the presence of unlabeled RNA and ^{15}N - and ^{15}N - ^{13}C -labeled proteins, respectively.

All NOESY spectra were recorded with a mixing time of 150 ms, the 3D TOCSY spectrum with a mixing time of 17.75 ms and the 2D TOCSY with a mixing time of 60 ms.

Structure calculation and refinement

AtnosCandid software (43,44) was used to generate preliminary structures and a list of automatically assigned NOE distance constraints for hnRNP G RRM in complex with RNA. Peak picking and NOE assignments were performed using 3D NOESY (^{15}N - and ^{13}C -edited) spectra. Additionally, intra-protein hydrogen bond constraints were based on hydrogen-deuterium exchange experiments on the amide protons. For these hydrogen bonds, the oxygen acceptors were identified based on preliminary structures calculated without hydrogen bond constraints. Protein dihedral angle constraints were generated by TALOS (45).

Seven iterations were performed and 80 independent structures were calculated at each iteration step. Structures of the protein-RNA complexes were calculated with CYANA (43) 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 9.0 (46) by simulated annealing in implicit water using the ff99SB protein force field (47). The 20 best structures based on energy and NOE violations were analyzed with PROCHECK (48). Figures were generated with MOLMOL (49).

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed on a VP-ITC instrument (Microcal®). The calorimeter was calibrated according to the manufacturer's instructions. Concentrations of proteins and RNAs were determined using optical density absorbance at 280 and 260 nm, respectively. Twenty micromolars of all the tested RNAs were titrated with 400 μM of hnRNP G RRM variants. Both protein and RNA were in the same NMR buffer. The injection protocol used was 40 injections of 6 μl every 5 min. All measurements were done at 40°C. Raw data were integrated and normalized for the molar concentration. After subtraction of the reference data recorded in the absence of RNA (Supplementary Figure S3B), the resulting integrated data were analyzed using the Origin 7.0 software according to a 1:1 RNA:protein ratio binding model.

Cell culture and plasmids

HEK293 (human embryonic kidney) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supple-

mented with 10% fetal bovine albumin (FBS). The pCI-SMN2 plasmid containing the SMN2 minigene was previously described (16). The human hnRNP G ORF was amplified by PCR and cloned in pcDNA3.1 mammalian expression vector with an N-terminal FLAG tag (pcFLAG) to generate the pcFLAG-hnRNP G plasmid. The hnRNP G ΔRRM (90–391), Δc57 (1–334), $\Delta\text{95–184}$ (1–95 fused to 184–391), $\Delta\text{95–235}$ (1–95 fused to 235–391), and $\Delta\text{95–250}$ (1–95 fused to 250–391) were amplified by PCR using pcFLAG-hnRNP G as a matrix and subsequently cloned in the pcFLAG vector. SMN2 and hnRNP G mutants were created by site-directed mutagenesis using specific primers.

In vivo splicing assay

One microgram of pCI-SMN2 [wild-type (WT) or mutant] was co-transfected with 1 μg of pcFLAG-hnRNP G (WT or mutant) in HEK293 cells plated in six-well plates using calcium phosphate method. Total RNA was extracted 48 h after transfection and 1 μg was then used for reverse transcription reaction using Oligo(dT) and M-MuLV Reverse Transcriptase RNaseH⁻ (Finnzyme®). One-tenth of the resulting cDNA was used for PCR amplification using a vector specific forward primer (pCI-fwd: 5'-GGTGTCCACTCCCAGTTCAA-3') and an SMN2 specific reverse primer (SMNex8-rev: 5'-GCCTACCAACCGTGCTGG-3'). PCR products were then resolved on a 2% agarose gel. Same results were obtained when using ^{32}P labeled SMNex8-rev primer and resolving the PCR products on a 4% polyacrylamide gel. The bands corresponding to the products of the splicing reaction were then quantified using ImageQuant. Experiments were repeated independently three times allowing the calculation of the mean and standard deviation for each assay.

Gel shift assay

The RNA was dephosphorylated using Antarctic phosphatase (NEB) and rephosphorylated with T4 polynucleotide kinase (NEB) in the presence of γ - ^{32}P ATP. The protein-RNA complexes were formed in the gel shift buffer. Tra2- β 1 protein amounts were similar to what was used previously (50). Five femtomoles of labeled RNA were mixed with increasing amounts of protein in a final volume of 6 μl and incubated for 30 min on ice. The resulting RNA-protein complexes were subsequently resolved on 6% native polyacrylamide gel at 4°C.

RESULTS

Structure determination

We previously proposed a model suggesting that Tra2- β 1 could recruit hnRNP G to SMN2 exon 7 upstream of its 5'-AGAA-3' binding site (14). Among all the RNA motifs proposed to interact specifically with hnRNP G (14,20,37), only a 5'-CAA-3' sequence located upstream of the Tra2- β 1 binding site could be identified as a putative hnRNP G binding site (Figure 1B). As hnRNP G RRM was previously shown to be primarily responsible for the binding of the protein to RNA (37), we purified a recombinant

protein containing the first 95 amino acids of human hnRNP G (Figure 1A) fused to an N-terminal His-tag. We then tested its interaction with the 5'-AUCAAA-3' RNA sequence, which corresponds to its putative binding site on SMN2 exon 7 (Figure 1B). NMR titration of hnRNP G RRM with increasing amounts of this RNA showed that they interact together (Figure 1C). Saturation was reached at a stoichiometric ratio of 1:1 and the protein resonances experienced fast to intermediate exchange throughout the titration steps. Mapping of the chemical shift perturbations observed during this titration revealed that residues from the β -sheet and the C-terminal region were primarily affected upon RNA binding (Figure 1D).

To characterize this interaction at the atomic level, we determined the structure of hnRNP G RRM bound to the 5'-AUCAAA-3' RNA using 2395 distance restraints derived from Nuclear Overhauser Effect (NOE) (Supplementary Table S1). The binding interface was characterized by 97 intermolecular NOEs (Supplementary Figure S1 and Table S1). The precision of the structure was high, with an Root-Mean-Square Deviation (RMSD) of 0.68 Å for all heavy atoms of the twenty lowest energy conformations represented in the final ensemble (Figure 2A).

Structure of hnRNP G RRM in complex with RNA

The structure shows that the RRM adopts a canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold (51) with the RNA lying as a single strand on the β -sheet surface (Figure 2B). In addition to the β -sheet surface, the C-terminal region of the RRM participates directly in the interaction with the RNA. The strong H1'-H2' correlations observed in a 2D total correlation spectroscopy (TOCSY) experiment indicate that all riboses of the RNA adopt a C2'-endo conformation.

The three adenines located at the 3' end of the RNA sequence are contacted by the RRM, but only adenines 4 and 5 are specifically recognized (Figure 2B). The base of A₄ stacks on the ring of Phe11 located within the RNP2 motif and is specifically recognized by three hydrogen bonds (Figure 2C). Two involve the side chains of Lys80 and Glu82 while the third one is formed with the backbone amide of Thr85 (Figure 2C). The base of A₄ adopts an unusual *syn* conformation, which is most likely induced by the presence of two aromatic groups located in the RNP1 motif, namely Phe51 and Phe53 (Figure 2B and C). In an *anti* conformation, the base would probably experience a steric clash with the rings of these aromatic residues. The base of A₅ is sandwiched between Phe53 located within the RNP1 motif and Pro87 from the C-terminal region and is specifically recognized by two hydrogen bonds formed with the side chain of Lys9 and the backbone carbonyl oxygen of Thr85 (Figure 2D). Additional contacts that are not involved in the specific recognition of the adenines but rather stabilize the protein-RNA interactions are also observed. A hydrogen bond is formed between the backbone amide of Ser88 and the phosphate group of A₅, and explains the downfield shift observed for the Ser88 amide proton upon RNA binding (Figures 1C and 2D). Moreover, Phe51 forms hydrophobic contacts with the riboses of A₄ and A₅ (Figure 2C and D). In contrast to A₄ and A₅, A₆ is not specifically recognized as the base forms only hydrophobic contacts with Phe89

located in the C-terminal region of the RRM. Finally, the side chain of Arg49 primarily forms a hydrogen bond with the phosphate group of A₆ (Figure 2B). However, this side chain seems to be flexible as in some structures of the ensemble it rather forms a hydrogen bond with the phosphate group of A₅.

To verify the importance of these interactions, we mutated to alanine most of the residues involved in RNA binding individually and tested the effect of these mutations on the RNA binding affinity of the RRM using ITC and NMR titrations. We verified by NMR that none of these point mutations affects the global fold of the RRM (Supplementary Figure S2). In agreement with the structure, all the mutations tested significantly decreased the affinity of the domain for the 5'-AUCAAA-3' RNA (Supplementary Figures S3C and S4B). The ITC data could be fitted in the presence of the WT protein and an apparent equilibrium dissociation constant ' K_d ' of 18 μ M was calculated from the curve (Supplementary Figure S3A). In the presence of the protein mutants, the affinity becomes too low to allow the fitting of recorded ITC data and estimate a K_d (Supplementary Figure S3C). However, the decrease in affinity was unambiguously confirmed by NMR titrations performed with the K80A and F89A protein variants, as chemical shift perturbations were shorter and experienced fast instead of intermediate exchange (Supplementary Figure S4B). A similar effect was observed when one or both of the specifically recognized adenines were mutated to cytosine (Supplementary Figures S3D and S5) strongly supporting the specific recognition of these two consecutive nucleotides by the RRM of hnRNP G. The importance of the non-specific interaction of hnRNP G RRM with the third nucleotide of the 5'-AAN-3' motif was also validated by the decrease in affinity observed with the 5'-AUCAA-3' RNA, which still retains the two consecutive adenines but misses the last nucleotide (Supplementary Figures S3D and S5). Finally, we tested the binding affinity of hnRNP G RRM for the 5'-UAAGAC-3' RNA, which contains flanking U and G nucleotides (underlined) instead of C and A in the WT sequence. In agreement with our structure, replacing the unbound cytosine by a uracil and the non-specifically recognized adenine by a guanine did not affect the RNA binding affinity of hnRNP G (Supplementary Figures S3D and S4A). Altogether, these data strongly support the intermolecular contacts observed in our structure and validate 5'-AAN-3' (N is for any nucleotide) as the minimal motif required for a sequence-specific interaction of hnRNP G RRM with RNA.

HnRNP G interacts specifically with SMN2 exon 7 using its RRM

HnRNP G was previously shown to be recruited by Tra2- β 1 to SMN2 exon 7 (12). As several 5'-AAN-3' putative binding sites are present around the Tra2- β 1 binding site (Figure 1B), we investigated whether hnRNP G could interact specifically with these motifs. We co-transfected HEK 293 cells with plasmids that encode different versions of the hnRNP G protein and an SMN2 minigene containing exons 6-8 of SMN2 with the intermittent introns (16). After RNA isolation, we monitored the levels of skipping or inclusion of exon 7 by RT-PCR. As previously reported (12),

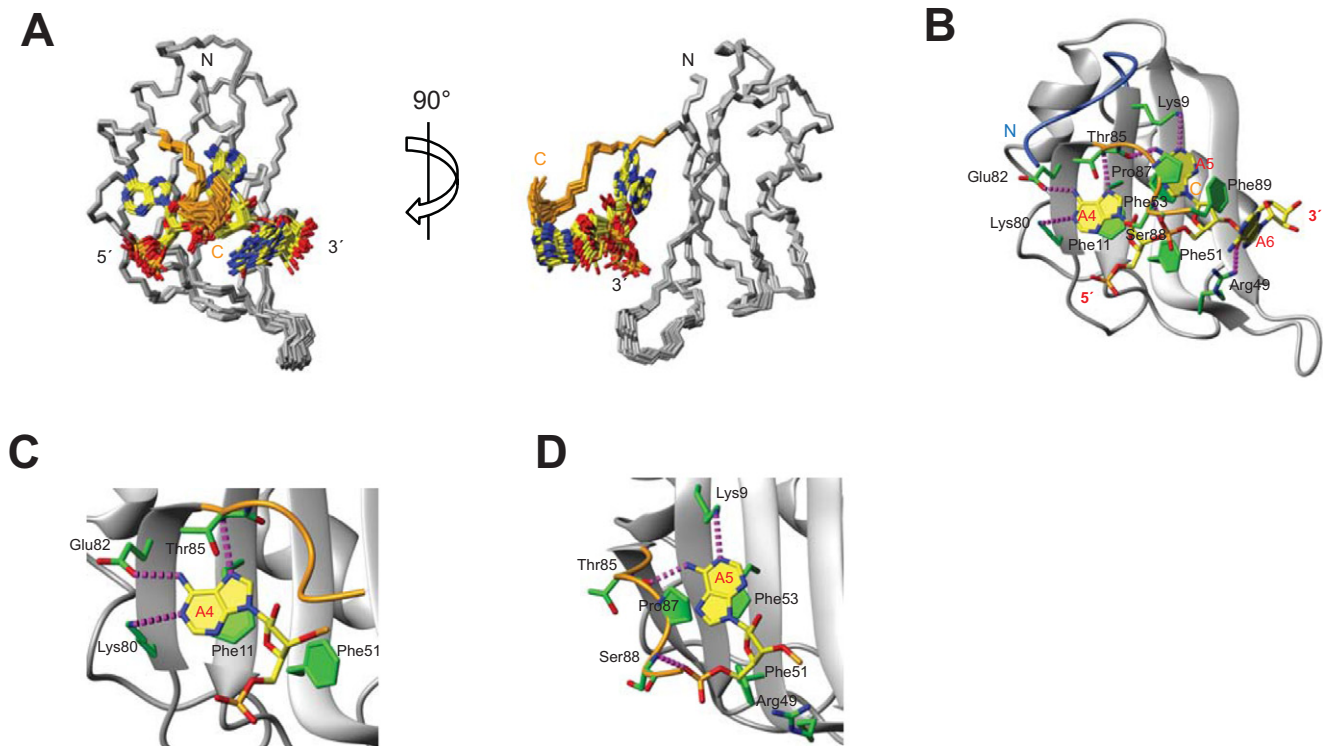


Figure 2. Overview of the solution structure of hnRNP G RRM bound to the 5'-AUCAAA-3' RNA. (A) Ensemble of the 20 lowest energy calculated structures fitted on the protein backbone and heavy atoms of the RNA. The protein backbone is shown in gray, with the C-terminus in orange. The heavy atoms of the RNA are colored yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus. The unstructured first three nucleotides and the N-terminus (amino acids 1–7) are hidden. (B) A representative structure from the ensemble showing the RNA bound to hnRNP G RRM. The protein and the RNA are represented in ribbons and sticks, respectively. The side chains of amino acids involved in the interaction with the RNA are represented in green sticks. The color scheme is the same as in (A). The N and C termini are colored in blue and orange, respectively. Hydrogen bonds are represented by purple dashed lines. (C) Molecular recognition of A₄ by hnRNP G RRM. Side chains of amino acids involved in the interaction are shown. Color scheme is as in (B). (D) Molecular recognition of A₅ by hnRNP G RRM. Representation is similar to (C). All figures were generated with MOLMOL (49).

overexpression of hnRNP G full-length activated the inclusion of exon 7 from around 20% in the absence of ectopic hnRNP G expression to around 65% (Figure 3A, lanes 1 and 2). Cell transfection with a truncated version of hnRNP G lacking the RRM induced a strong decrease in exon 7 inclusion to only 35% (Figure 3A, lane 3), showing that this domain is required for the function of hnRNP G as a splicing regulator of SMN2. It strongly suggested that hnRNP G interacts directly with the polyA tract located in exon 7 (Figure 1B). To determine whether this interaction was specific, we tested the effect of mutations in hnRNP G that affect its ability to recognize adenines on SMN2 exon 7 splicing. As shown in Figure 3A, all the mutations tested significantly decreased the level of exon 7 inclusion except the replacement of Phe53 with alanine, which had an effect only when combined with the Lys9 mutation to alanine (Figure 3A, lanes 4–7). The absence of effect observed with the F53A mutation can be explained by the fact that hnRNP G is recruited to SMN2 by Tra2- β 1 (12). Indeed, this inter-protein interaction could compensate for the decrease in RNA affinity induced by the single mutation. Finally, the effects observed for the F11A mutation and the K9A + F53A double mutation were very close to a full truncation of the RRM (Figure 3A, lanes 3, 5 and 7), indicating that the RRM of hnRNP G interacts with SMN2 specifically by recognizing two consecutive adenines.

HnRNP G binds an A-tract located upstream of the Tra2- β 1 binding site on SMN2 exon 7

As hnRNP G was shown to be recruited to SMN2 exon 7 by Tra2- β 1 (12), we examined sequences located on both sides of the Tra2- β 1 binding site and identified three regions containing at least two consecutive adenines, two upstream (A₁₁–A₁₄ and A₁₇–A₂₀) and one downstream (A₂₇A₂₈) (Figure 1B). Mutations of adenines to uracils in the regions A₁₁–A₁₂ or A₁₇–A₁₈ were previously tested and had a strong effect on SMN2 exon 7 splicing (14). We then tested the effect of the A₂₇ to U mutation, but could not detect any significant effect on exon 7 splicing when hnRNP G was overexpressed (Figure 3B). These results strongly suggest that hnRNP G binds to the adenine-tracts located upstream of the Tra2- β 1 binding site rather than downstream (Figure 1B).

To investigate whether hnRNP G RRM could bind to the A₁₇–A₂₀ site without inducing any steric hindrance with the RNA bound Tra2- β 1 RRM, we first investigated by NMR the interaction of each of these two RRM with the SMN2 exon 7 derived RNA 5'-U₁₅CAAAAAGAAG₂₅-3' (Figure 1B). This RNA sequence contains the motif 5'-AGAA-3', which was previously identified as the Tra2- β 1 binding site (14), and the closest upstream putative hnRNP G binding site 5'-AAA-3' (Figure 1B). As illustrated in Figure 4, the chemical shift perturbations observed upon bind-

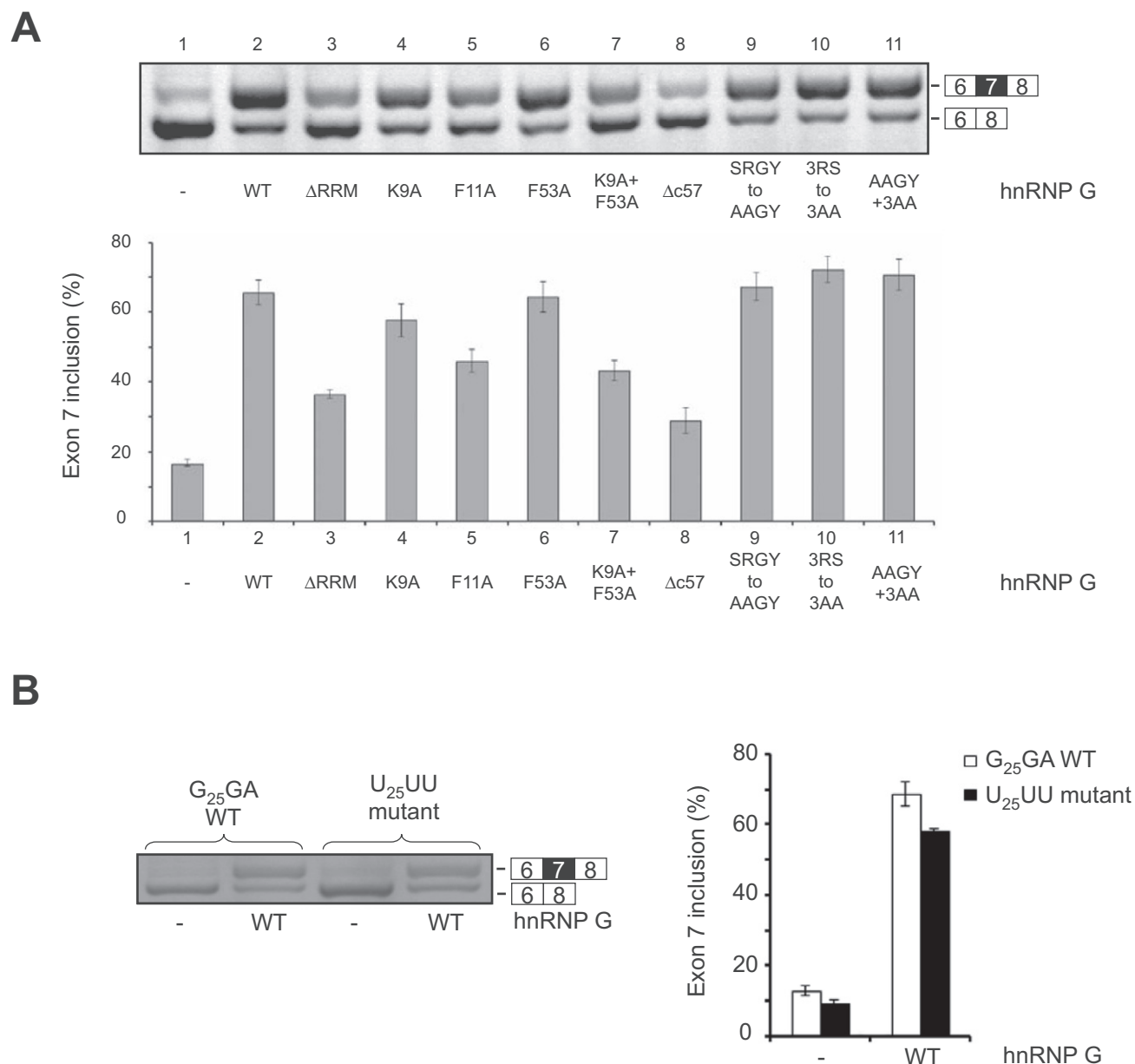


Figure 3. Effect of mutations in hnRNP G and the pre-mRNA on SMN2 exon 7 splicing. (A) RT-PCR gel shows the levels of exon 7 inclusion in SMN2 mRNAs upon overexpression of WT or mutated versions of hnRNP G in HEK 293 cells. The positions of PCR products corresponding to SMN2 mRNA with or without exon 7 are indicated on the right of the gel. The graph is the result of at least three independent experiments. Error bars represent standard deviations. The negative control corresponds to the percentage of exon 7 inclusion in the absence of ectopic hnRNP G expression. (B) RT-PCR gel shows the effect of the GGA to UUU mutation in the potential hnRNP G binding site located downstream of the Tra2- β 1 binding site (Figure 1B) on SMN2 exon 7 splicing. Conditions are similar to (A).

ing of Tra2- β 1 or hnRNP G RRM to the long RNA (5'-UCAAAAAGAAG-3') or to RNAs containing their single binding sites (5'-AAGAAC-3' and 5'-AUCAAA-3' for Tra2- β 1 and hnRNP G, respectively) were similar (Figure 4A and B). This showed that both Tra2- β 1 and hnRNP G RRMs could interact with their expected binding sites (5'-AGAA-3' and 5'-AAN-3', respectively) in the context of the long RNA. Surprisingly, our data also show that hnRNP G RRM binds to the long RNA with a higher affinity than the short RNA tested (K_d of 0.6 μ M instead of 18 μ M) (Supplementary Figures S3A and S6B). In agreement, some

chemical shift perturbations had a larger magnitude or experienced an intermediate instead of fast exchange regime in the presence of the long RNA (Figure 4A and B). This difference in affinity is most likely the result of an avidity effect due to the presence of five overlapping 5'-AAN-3' motifs in the long RNA 5'-UCAAAAAGAAG-3' (Figure 4B), which could all be bound by hnRNP G in the absence of Tra2- β 1.

We then tested the binding of hnRNP G RRM to the complex formed by Tra2- β 1 RRM with the long 5'-UCAAAAAGAAG-3' RNA. At a Tra2- β 1:hnRNP G:RNA ratio of 1:1:1, chemical shifts correspond to the

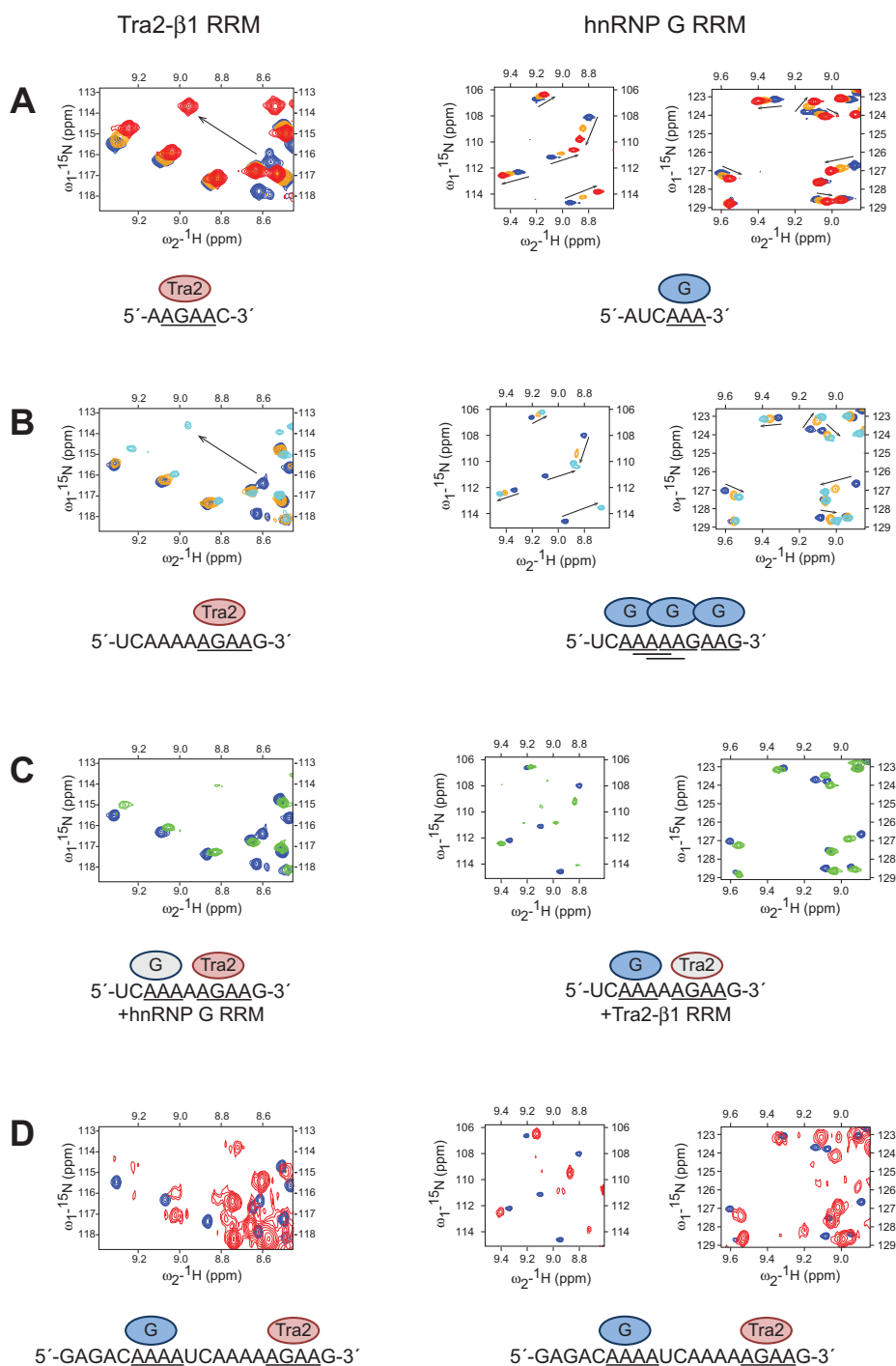


Figure 4. Selection of chemical shift perturbations observed upon titrations of Tra2-β1 and hnRNP G RRM with SMN2 exon 7 derived RNAs. (A) Close-ups on the titration of Tra2-β1 and hnRNP G RRM with the unlabeled 5'-AAGAAC-3' and 5'-AUCAAA-3' RNAs, respectively. The peaks corresponding to free and RNA-bound protein states (RNA:protein ratios of 0.3:1 and 1:1) are blue, orange and red. The underlined sequences represent the nucleotides that are bound by each protein. On this diagram, red and blue ovals represent Tra2-β1 and hnRNP G RRM, respectively. (B) Close-ups on the titration of Tra2-β1 and hnRNP G RRM with the unlabeled SMN2 derived 5'-UCAAAAAGAAG-3' RNA containing binding sites of both proteins (underlined sequences). The color code is similar to (A) except that the 1:1 bound state is in cyan. The code for the diagrams is similar to (A). (C) Close-ups on overlay of ^1H - ^{15}N HSQC spectra corresponding to the binding of ^{15}N labeled hnRNP G RRM and ^{15}N labeled Tra2-β1 RRM to the unlabeled SMN2 derived 5'-UCAAAAAGAAG-3' RNA containing the binding sites of both proteins (underlined sequences). Blue peaks represent the free proteins and green peaks represent the complex formed in the presence of Tra2-β1 and hnRNP G RRM with the RNA at a ratio of 1:1:1. To simplify the spectra, peaks corresponding to the protein represented by a gray oval on the diagrams are hidden. Full spectra are in Supplementary Figure S6A. (D) Close-ups on overlay of ^1H - ^{15}N HSQC spectra corresponding to the binding of ^{15}N labeled hnRNP G RRM and ^{15}N labeled Tra2-β1 RRM to the long unlabeled SMN2 derived 5'-GAGACAAAUCAAAAGAAG-3' RNA containing the binding sites of both proteins (underlined sequences). Blue peaks represent the free proteins and red peaks represent the complex formed in the presence of Tra2-β1 and hnRNP G RRM with the RNA at a ratio of 1:1:1. Full spectra are shown in Supplementary Figure S6A.

RNA-bound states of both proteins (Figure 4C and Supplementary Figure S6A), showing that Tra2- β 1 and hnRNP G RRMs could both be accommodated on a single RNA molecule containing their adjacent binding sites. In addition, the absence of additional chemical shift perturbations between spectra recorded in the presence of RNA bound to either each protein alone or to both proteins together indicates that the two RRMs do not interact together when they bind two adjacent binding sites (Figure 4B and C).

Surprisingly, the comparison of the bound state of hnRNP G on the 5'-UCAAAAAGAAG-3' RNA in the presence and absence of Tra2- β 1 reveals that the chemical shift perturbations of hnRNP G are smaller when Tra2- β 1 is already bound to RNA (Figure 4B and C). This decrease in the RNA binding affinity of hnRNP G can be explained by the reduction of the number of its binding sites available when Tra2- β 1 is bound to the 5'-AGAA-3' motif. Indeed, Tra2- β 1 occupies this motif with around 8-fold higher affinity ($K_d = 2.25 \mu\text{M}$) (14) than hnRNP G leaving only the first 5'-AAN-3' binding site available instead of five. However, upon binding to the 5'-UCAAAAAGAAG-3' RNA/Tra2- β 1 complex, the chemical shift perturbations of hnRNP G become shorter than the ones observed with the 5'-AUCAAA-3' RNA (Figure 4A,C). This effect is most likely due to the close proximity between the binding sites of hnRNP G and Tra2- β 1, making the last 3' adenine inaccessible for hnRNP G.

To investigate whether this decrease in affinity could be compensated by hnRNP G/Tra2- β 1 interactions (12–14), we produced full-length versions of both proteins and tested their binding to the 5'-UCAAAAAGAAG-3' RNA using gel shift assays. As expected, we observe a shift corresponding to the binding of each single protein to RNA (Figure 5A). However, no additional shift was observed in the presence of the two proteins (Figure 5A). It indicates that the A₁₇–A₂₀ tract is not the binding site used by hnRNP G in the presence of Tra2- β 1.

We then investigated the importance of the most upstream A-tract (A₁₁–A₁₄) by testing the binding of the full-length proteins to an extended SMN2 derived RNA 5'-GAGACA₁₁AAA₁₄UCAAAAAGAAG₂₅-3'. As illustrated in the Figure 5B, a shift was observed in the presence of individual proteins and an upper-band appeared when the two proteins were added together, indicating that both proteins could be accommodated on the RNA (Figure 5B). In addition, an NMR titration performed with this RNA shows that in the presence of bound Tra2- β 1 RRM, the affinity of hnRNP G RRM is similar to what was observed with the short 5'-AUCAAA-3' RNA (Figure 4A and D) and not reduced as in the presence of the shorter 5'-UCAAAAAGAAG-3' sequence (Figure 4C). All together, these results suggest that hnRNP G binds to the A₁₁–A₁₄ site on SMN2 exon 7, when Tra2- β 1 is bound to the A₂₁GAA₂₄ motif.

Several domains of hnRNP G are important for the activation of SMN2 exon 7 inclusion

Our results reveal that hnRNP G has a weak RNA binding specificity and affinity, emphasizing the importance of its recruitment by Tra2- β 1 to SMN2 pre-mRNA. It was previ-

ously reported that the CTD part of hnRNP G is responsible for its interaction with Tra2- β 1 (12) (Figure 1A). In particular, the C-RBD seemed to be important for this interaction as truncation of the last 41 amino acids significantly decreases hnRNP G affinity for Tra2- β 1 without affecting its SMN2 binding capacity (12). In agreement with the functional importance of this domain, full truncation of hnRNP G C-RBD (Δ C57) strongly decreased the percentage of exon 7 inclusion in SMN2 mRNAs from 65 to 30% (Figure 3A, lane 8). Previous reports suggested that hnRNP G/Tra2- β 1 interaction could be mediated by the SRGY motif and several RS repeats located in the CTD of hnRNP G with one of the RS domains of Tra2- β 1 (14,35). However, mutation of all RS repeats located in the C-RBD and/or the SRGY motif (Figure 1A) to alanines did not induce significant reduction of exon 7 inclusion indicating that these motifs are not crucial for hnRNP G recruitment to SMN2 (Figure 3A, lanes 9–11).

To investigate whether additional parts of hnRNP G were involved in SMN2 exon 7 splicing, we tested the importance of the central region separating its RNA binding site (RRM) from its Tra2- β 1 binding site (the CTD). We progressively truncated the central region of hnRNP G keeping the RRM fused to the last 207 (Δ 95–184), 156 (Δ 95–235) or 141 (Δ 95–250) amino acids and tested the effect of these truncated proteins on SMN2 exon 7 splicing in cells. Unexpectedly, truncation of only 89 amino acids in this region of hnRNP G resulted in a strong decrease of SMN2 exon 7 inclusion (Figure 6). Interestingly, this region contains an RGG box (Figure 1A), which, to our knowledge, was not reported before to be important for the function of hnRNP G in SMN2 exon 7 splicing.

To investigate whether the RGG box could be involved in SMN2 exon 7 recognition, we produced a version of hnRNP G that contains both the RRM and the RGG box (residues 1–127) and tested its interaction with the 5'-AUCAAA-3' RNA. As illustrated in Supplementary Figure S7, the same residues of the RRM are involved in RNA binding with and without the RGG box (Figure 1C and Supplementary Figure S7) indicating that the mode of RNA recognition of the RRM is the same. In addition, we tested whether the presence of the RGG box could modulate the specificity of RNA recognition of the RRM. As observed with the RRM alone, very small chemical shift perturbations were observed with the suboptimal 5'-AUCCCC-3' RNA sequence in the presence of the RGG box (Supplementary Figure S7). Finally, no chemical shift perturbation was detected from the RGG box in the presence of these two RNAs. These data show that the RGG box does not bind to the RNAs tested and does not change the specificity of interaction of hnRNP G RRM with RNA. In addition to the function of the RRM binding to RNA and the CTD interacting with Tra2- β 1, the central part of the protein may then play a role either as a spacer between the protein domains or as a functional entity of hnRNP G.

DISCUSSION

hnRNP G interacts specifically with two consecutive adenines

Our structure shows that hnRNP G RRM binds to a 5'-AAN-3' motif in which the two consecutive adenines are

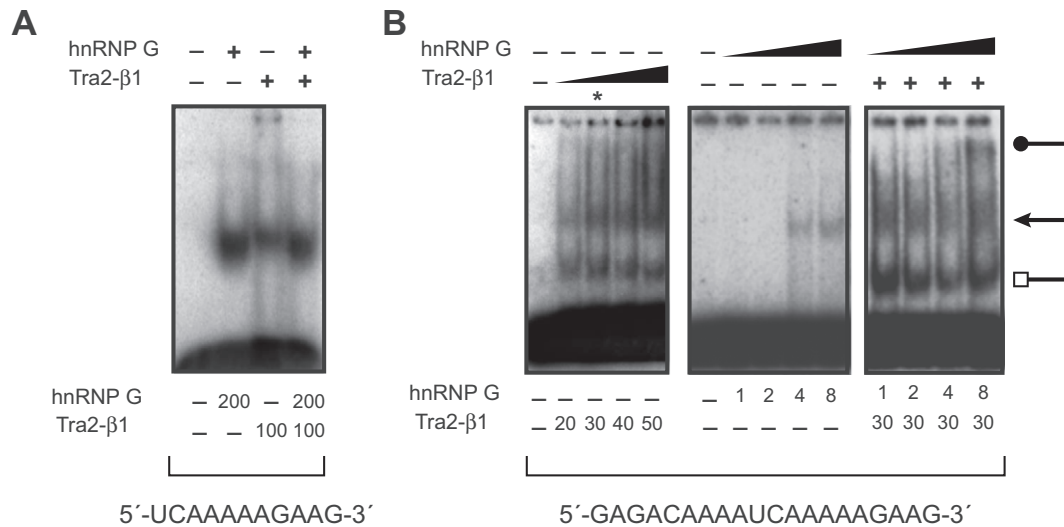


Figure 5. Gel shift experiments showing binding of full-length hnRNP G and Tra2-β1 proteins to SMN2 derived RNA sequences. **(A)** Gel shift experiment showing the binding of full-length hnRNP G and Tra2-β1 to the 5'-UCAAAAAGAAG-3' RNA. The amount of proteins in nanograms is indicated below each lane. Both proteins could bind the RNA separately but did not bind together. **(B)** Gel shift experiment showing the binding of increasing amount of full-length Tra2-β1 (left) or hnRNP G (middle) to the 5'-GAGACAAAUAUCAAAGAAG-3' RNA. The right panel shows the binding of both proteins to the RNA in the presence of 30 ng of Tra2-β1 (indicated by an asterisk on the left panel) and increasing amounts of hnRNP G. The amount of proteins in nanograms is indicated below each lane. The round head arrow represents the shift corresponding to the binding of the two proteins to RNA, the pointed head arrow represents binding of single protein molecules to RNA and the square head arrow represents binding of a co-purified truncated version of Tra2-β1 to the RNA.

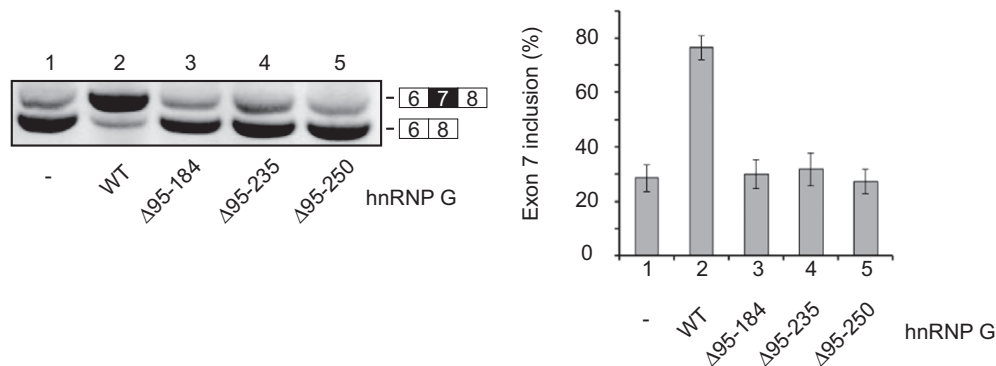


Figure 6. Effect of truncations in the central part of hnRNP G on the splicing of SMN2 exon 7. Three truncations were performed in the region of hnRNP G located between its RRM and the CTD. The effect of these protein variants on SMN2 exon 7 splicing was then tested in cells using the same conditions as described in Figure 3.

specifically recognized (Figure 2). This RNA binding specificity was validated *in vitro* using ITC and NMR measurements (Supplementary Figures S3, S4 and S5) and in cells by splicing assays (Figure 3). In agreement with our results, hnRNP G was proposed to bind specifically to the 5'-AAGU-3' RNA motif (20). Our data can also explain the absence of specificity reported by Hofmann and Wirth (12) when they investigated the interaction of hnRNP G with SMN2 as they tested the binding of hnRNP G to long RNA sequences containing two consecutive adenines (12,52). In addition, we show that hnRNP G can bind the 5'-CCC-3' or 5'-CCA-3' motifs selected by SELEX (37) (Supplementary Figures S3D and S5). However, our structure suggests that they would not be accommodated by the RRM as well as the 5'-AAN-3' motif (37). Indeed, a cytosine replacing the first recognized adenine would probably fail to

form the same hydrogen bond network, as the smaller cytosine base would then be more distant from the side chains of Lys80 and Glu82 (Figure 2C). Similarly, a cytosine located in the second pocket would be further away from the Lys9 side chain preventing the formation of the hydrogen bond observed in the presence of an adenine (Figure 2D). In agreement with these predictions, our ITC and NMR data showed that the affinity of hnRNP G RRM for the 5'-AUCCCC-3' and 5'-AUCCCA-3' RNA sequences is lower than for 5'-AAN-3' containing RNAs (Supplementary Figures S3D and S5). Moreover, having the SELEX motif 5' to the 5'-AAN-3' sequence (5'-ACCAAA-3' RNA) does not improve the RNA binding affinity (Supplementary Figure S5). In conclusion, we show that hnRNP G binds preferentially 5'-AAN-3' motifs, but can also accommodate 5'-CCC-3' and 5'-CCA-3' motifs.

hnRNP G and its paralogues have different RNA specificities

The mode of RNA recognition of hnRNP G RRM is reminiscent of what was observed in the structure of its paralogue in testis, RBMY (38) as the two consecutive adenines are recognized in a fairly similar manner (Supplementary Figure S8A). However, their modes of interaction are not identical. RBMY was shown to interact with a stem-loop RNA by insertion of its β_2 - β_3 loop in the major groove of the RNA helix, whereas hnRNP G binds only single-stranded RNA (38). hnRNP G was previously shown not to interact with a stem because the sequence of its β_2 - β_3 loop is different. An additional glutamate is inserted between Arg43 and Thr44, and Ser45 is replaced with an asparagine (38).

The binding of RBMY to a stem-loop RNA partially explains some of the other differences observed between the two RNA-protein complexes. The involvement of G₁₄ in the first base pair of the stem prevents the base from being available to contact Phe88 located in the C-terminal region of RBMY RRM (Supplementary Figure S8A and B). Formation of this corresponding interaction in hnRNP G, between A₆ and Phe89 maintains the C-terminal region of the protein in a position that allows the formation of a hydrogen bond between the backbone amide of Ser88 and the phosphate of A₅, which could not be observed in the case of RBMY (Supplementary Figure S8A). Although these interactions are unspecific, they contribute to the RNA binding affinity of hnRNP G. This could compensate, even partially, for the inability of this protein to interact with the stem of a stem-loop RNA.

Another major difference is the absence of interaction between hnRNP G and the cytosine located 5' to the two recognized adenines, contrary to what was observed with RBMY (Supplementary Figure S8A). Indeed, no intermolecular NOEs were observed between the cytosine and any of the hnRNP G RRM residues, suggesting that this nucleotide stays flexible (Supplementary Figure S8A). This was surprising because most residues involved in the recognition of the cytosine are conserved in both proteins (Supplementary Figure S9). This difference may originate from two interactions observed in RBMY involving Arg17 and Arg48, which form hydrogen bonds with the phosphate group of C₁₁ and A₁₂, respectively (Supplementary Figure S8A). These two hydrogen bonds possibly stabilize the cytosine by paying the entropic cost required for positioning the nucleotide. In hnRNP G, Arg17 is replaced by a threonine and Arg49 (the equivalent of Arg48 in RBMY) is not restrained by the interaction with the stem and therefore is oriented differently to interact with the phosphate groups of A₆ or A₅ (Supplementary Figure S8A). As a consequence of the flexibility of C₃ base, the side chain of Lys80 becomes available in hnRNP G to form a hydrogen bond with the base of A₄ pulling the base downwards and preventing the formation of the hydrogen bond observed in RBMY with the backbone carbonyl oxygen of Gln83 (Supplementary Figure S8A).

In conclusion, our data show that hnRNP G and RBMY use a similar but distinct mode of interaction with RNA. hnRNP G-T, another paralogue of hnRNP G found in testis, was reported to interact with 5'-GUU-3' containing RNAs (53), a sequence that is different from what was reported

for RBMY (38) and what we report here for hnRNP G. All these results strongly suggest that in testis hnRNP G, hnRNP G-T and RBMY may regulate splicing by interacting with different RNA sequences and structures.

The RRM, an important domain for the function of hnRNP G

Although the contribution of hnRNP G RRM was in some cases reported to be negligible for its activity as a splicing regulator (35,53), our results show that RNA recognition by hnRNP G RRM is important for the activation of SMN2 exon 7 inclusion (Figure 3A). RRM truncation and mutation of residues involved in the recognition of two consecutive adenines significantly reduced the splicing activity of hnRNP G (Figure 3A). It strongly suggests that the specific interaction of hnRNP G with SMN2 exon 7 contributes to the selective recruitment of the Tra2- β 1/hnRNP G heterodimer to RNA. It then extends the initial recognition of a short 5'-AGAA-3' motif by Tra2- β 1 to the longer sequence 5'-AAAANNNNNNAGAA-3' (the binding sites of hnRNP G and Tra2- β 1 are underlined). In addition, the weak specificity of interaction of hnRNP G with RNA allows its binding to different registers on the A-rich sequences present upstream of Tra2- β 1 binding site. It increases its RNA binding affinity locally (Figure 4B) and probably facilitates the recruitment of the hnRNP G/Tra2- β 1 heterodimer at this position before to restrict hnRNP G interaction with the A₁₁-A₁₄ tract. The enhancement of weak protein-RNA interactions by the repetition of several consecutive binding motifs was previously reported (54) and was observed in the context of PTB binding to pyrimidine tracts (55).

Interestingly, the RRM was also shown to be important for the function of hnRNP G as a tumor suppressor. Two point mutations of residues located in this domain, namely K22R and G29D were reported to affect the tumor suppressive activity of hnRNP G (26,56). Surprisingly, these residues are not located on the RNA binding surface of the domain. They are both in the α_1 helix, an element that could rather be involved in intra or inter protein-protein interactions as reported for the human SR protein SRSF1 (57,58). It shows that the RRM is an important domain of the protein that contributes to hnRNP G functions using different modes of action.

Tra2- β 1 and hnRNP G, a simultaneous versus competitive binding to RNA

It was previously reported that hnRNP G and Tra2- β 1 sometimes have opposite effects on splicing (20). hnRNP G can sequester Tra2- β 1 and other SR proteins preventing their access to RNA and thereby their activity on splicing (35,53) (Figure 7A). Alternatively, hnRNP G and Tra2- β 1 could also antagonize the binding of each other to RNA (20) (Figure 7A). This later mode of action would explain the opposite effects observed for these two proteins on endometrial cancer progression (59). As illustrated in Figure 4C, we also observed a competition between these two proteins when both Tra2- β 1 and hnRNP G RRMs are bound to the SMN2 derived 5'-UCAAAAAGAAG-3' RNA. In that case, the binding of Tra2- β 1 restricted

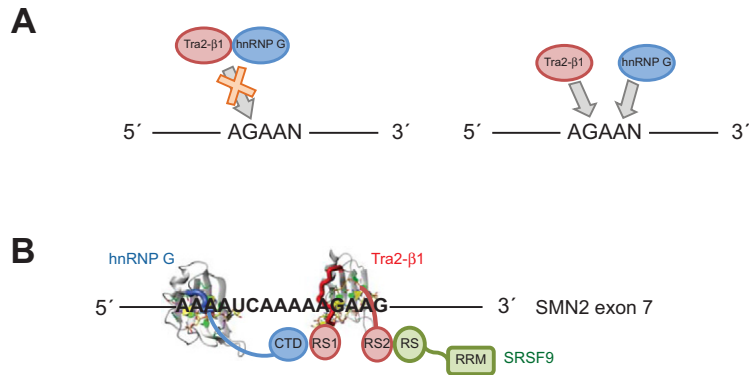


Figure 7. Models representing the competitive and cooperative binding modes of hnRNP G and Tra2-β1 to RNA. (A) Two competitive binding modes of hnRNP G and Tra2-β1 to RNA. hnRNP G either sequesters Tra2-β1 through protein–protein interactions preventing its binding to RNA (54) (model on the left) or the two proteins compete together to bind the same overlapping binding site (20) (model on the right). (B) Model showing the assembly of hnRNP G, Tra2-β1 and SRSF9 on exon 7 of SMN2 pre-mRNA (14). hnRNP G, Tra2-β1 and SRSF9 are represented in blue, red and green, respectively. The model shows interaction of hnRNP G with the binding site identified in this study, using its RRM and positioning the CTD toward the 3' end of the bound RNA. Tra2-β1 RRM binds to its previously identified binding site located downstream positioning its N-terminal RS1 domain toward the 5' end of the bound RNA (14). This RNA binding induced proximity between the RS1 domain of Tra2-β1 and the CTD of hnRNP G probably favors their interaction. The RS2 domain of Tra2-β1 is positioned toward the 3' end of the bound RNA and may interact with the RS domain of SRSF9 (14).

the binding of hnRNP G to the first 5'-AAN-3' motif instead of the five overlapping registers it could utilize in the absence of Tra2-β1. Our data reveal that this effect is due to overlapping of the motif 5'-AAN-3' bound by hnRNP G with the previously identified motif 5'-AGAA-3' recognized by Tra2-β1 (14,15) (Figure 7A) and by the 8-fold higher affinity of Tra2-β1 for its binding site on SMN2 (14). Finally, we demonstrate here that Tra2-β1 and hnRNP G can bind together a single RNA molecule that contains their recognition motifs adjacent to each other (Figure 7B) as shown with the long SMN2 derived 5'-GAGACAAAUAUCAAAGAAG-3' RNA sequence (Figures 4D and 5B). We also show that a spacing between the two binding sites is necessary, most likely to allow protein–protein interactions between these two partners. Altogether, these data indicate that Tra2-β1 and hnRNP G can either compete for binding to RNA or interact simultaneously depending on the targeted RNA sequence.

The mode of action of Tra2-β1 and hnRNP G as activators of SMN2 exon 7 splicing

Our data support a specific recognition of the SMN2 pre-mRNA by hnRNP G and Tra2-β1 (Figures 3–5). We propose that the presence of several 5'-AAN-3' motifs upstream of the Tra2-β1 binding site facilitates the recruitment of hnRNP G to SMN2 by Tra2-β1 and therefore increases the specificity and affinity of exon 7 recognition by the heterodimer (Figure 7B). Our structure shows that the C-terminal region of hnRNP G RRM positions the CTD downstream of the hnRNP G binding site (Figures 2B and 7B). In agreement with our data, the CTD was previously shown to be responsible for hnRNP G interaction with Tra2-β1 (12). The mode of interaction of hnRNP G with Tra2-β1 still needs to be characterized but seems not to occur via the RS repeats located in the CTD of hnRNP G (Figure 3A). In addition, our data suggest that this Tra2-β1/hnRNP G interaction could partially compensate for the loss of hnRNP G binding to SMN2. Indeed, mutations

in hnRNP G RRM that significantly decrease its binding to RNA or even a full truncation of the RRM had only a moderate effect on SMN2 exon 7 splicing (Figure 3A). In conclusion, we propose that Tra2-β1 anchors the binding of the heterodimer by positioning and stabilizing hnRNP G binding to RNA, which in turn increases the affinity and specificity of recognition of SMN2 exon 7 by interacting with the A-tracts located upstream of the Tra2-β1 binding site (Figure 7B). Interestingly, the 5'-AAN-3' motif was found within eight nucleotides upstream and/or downstream of 85% of potential Tra2-β1 binding sites selected by CLIP (60), strongly suggesting that this mode of Tra2-β1 and hnRNP G assembly on RNA could be more widely used.

Although we better understand the assembly of these factors on SMN2 exon 7, their modes of regulation of exon 7 splicing still need to be characterized. Interestingly, the two proteins, Tra2-β1 and hnRNP G were identified as part of the supraspliceosome (61). One or both of these factors could then play a role in the spliceosome recruitment or assembly by inducing protein–protein interactions. At least two domains of hnRNP G would still be available for spliceosome recruitment. The NTD located in the central region of hnRNP G was shown to be responsible for the interaction of hnRNP G with other proteins (36). In addition, we showed that the region of hnRNP G located between the RRM and the NTD (Figure 1A) was important for the splicing of SMN2 exon 7 (Figure 6). This part of the protein contains an RGG box, which could mediate additional protein interactions. Tra2-β1 could also contribute to exon 7 splicing activation by promoting the recruitment of spliceosome components through one or both of its RS domains. Finally, the SR protein SRSF9 (SRp30c), which is also part of the supraspliceosome (61), was proposed to be recruited to SMN2 exon 7 by Tra2-β1 (62). SRSF9 could therefore be involved as well in the recruitment of the spliceosome using its C-terminal RS domain. Further investigation is now

needed to identify the exact contribution of these three proteins to SMN2 exon 7 splicing.

ACCESSION NUMBERS

We deposited the chemical shifts of hnRNP G RRM bound to 5'-AUCAAA-3' RNA to the BioMagResBank under accession number 19382. We have deposited the coordinates of the hnRNP G RRM-AUCAAA structures in the Protein Data Bank (PDB) under the PDB ID: 2mb0.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGMENTS

We thank Prof. James Stevenin for providing the hnRNP G clone, Dr Markus Blatter for his help with structure calculations, Dr Julien Boudet for his help with the ITC data analysis, Dr Mario Schubert, Dr Christophe Maris and Dr Fred Damberger for their help with NMR spectroscopy settings, Dr Juan Gerez, Georg Dorn and Michal Walczak for their help with protein expression and purification.

FUNDING

Suisse National Science Foundation (SNF), National Centres of Competence in Research (NCCR) Structural Biology and the Alternative Splicing Network of Excellence EURASNET [to F.H.T.A.]; ETH [ETH-03 11-1 to A.M.]; Fondation Suisse de Recherche sur les Maladies Musculaires (FSRMM) and Spinal Muscular Atrophy (SMA)-Europe [to A.C.]. Funding for openaccess charge: ETH, Zurich.

Conflict of interest statement. None declared.

REFERENCES

- Lorson, C.L., Rindt, H. and Shababi, M. (2010) Spinal muscular atrophy: mechanisms and therapeutic strategies. *Hum. Mol. Genet.*, **19**, R111–R118.
- Pearn, J. (1978) Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. *J. Med. Genet.*, **15**, 409–413.
- Feldkotter, M., Schwarzer, V., Wirth, R., Wienker, T.F. and Wirth, B. (2002) Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am. J. Hum. Genet.*, **70**, 358–368.
- Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Violette, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M. *et al.* (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80**, 155–165.
- Rochette, C.F., Gilbert, N. and Simard, L.R. (2001) SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to Homo sapiens. *Hum. Genet.*, **108**, 255–266.
- Kashima, T. and Manley, J.L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.*, **34**, 460–463.
- Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.*, **30**, 377–384.
- Novelli, G., Calza, L., Amicucci, P., Giardino, L., Pozza, M., Silani, V., Pizzuti, A., Gennarelli, M., Piombo, G., Capon, F. *et al.* (1997) Expression study of survival motor neuron gene in human fetal tissues. *Biochem. Mol. Med.*, **61**, 102–106.
- Mattis, V.B., Bowerman, M., Kothary, R. and Lorson, C.L. (2008) A SMNDelta7 read-through product confers functionality to the SMNDelta7 protein. *Neurosci. Lett.*, **442**, 54–58.
- Bebee, T.W., Gladman, J.T. and Chandler, D.S. (2010) Splicing regulation of the survival motor neuron genes and implications for treatment of spinal muscular atrophy. *Front. Biosci.*, **15**, 1191–1204.
- Horne, C. and Young, P.J. (2009) Is RNA manipulation a viable therapy for spinal muscular atrophy? *J. Neurol. Sci.*, **287**, 27–31.
- Hofmann, Y. and Wirth, B. (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum. Mol. Genet.*, **11**, 2037–2049.
- Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J. and Eperon, I.C. (2000) RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Hum. Mol. Genet.*, **9**, 685–694.
- Clery, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S. and Allain, F.H. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. *Nat. Struct. Mol. Biol.*, **18**, 443–450.
- Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T. *et al.* (2011) Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res.*, **39**, 1538–1553.
- Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J. and Wirth, B. (2000) 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. U.S.A.*, **97**, 9618–9623.
- Soulard, M., Della Valle, V., Siomi, M.C., Pinol-Roma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J.C., Monod, G., Dreyfuss, G. *et al.* (1993) hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res.*, **21**, 4210–4217.
- Delbridge, M.L., Lingenfelter, P.A., Distech, C.M. and Graves, J.A. (1999) The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nat. Genet.*, **22**, 223–224.
- Lingenfelter, P.A., Delbridge, M.L., Thomas, S., Hoekstra, H.E., Mitchell, M.J., Graves, J.A. and Distech, C.M. (2001) Expression and conservation of processed copies of the RBMX gene. *Mamm. Genome*, **12**, 538–545.
- Nasim, M.T., Chernova, T.K., Chowdhury, H.M., Yue, B.G. and Eperon, I.C. (2003) HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. *Hum. Mol. Genet.*, **12**, 1337–1348.
- Wang, J., Gao, Q.S., Wang, Y., Lafyatis, R., Stamm, S. and Andreadis, A. (2004) Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J. Neurochem.*, **88**, 1078–1090.
- Wang, Y., Wang, J., Gao, L., Stamm, S. and Andreadis, A. (2011) An SRP75/hnRNPG complex interacting with hnRNPE2 regulates the 5' splice site of tau exon 10, whose misregulation causes frontotemporal dementia. *Gene*, **485**, 130–138.
- Takemoto, T., Nishio, Y., Sekine, O., Ikeuchi, C., Nagai, Y., Maeno, Y., Maegawa, H., Kimura, H. and Kashiwagi, A. (2007) RBMX is a novel hepatic transcriptional regulator of SREBP-1c gene response to high-fructose diet. *FEBS Lett.*, **581**, 218–222.
- Omura, Y., Nishio, Y., Takemoto, T., Ikeuchi, C., Sekine, O., Morino, K., Maeno, Y., Obata, T., Ugi, S., Maegawa, H. *et al.* (2009) SAFB1, an RBMX-binding protein, is a newly identified regulator of hepatic SREBP-1c gene. *BMB Rep.*, **42**, 232–237.
- Zhao, S., Korzan, W.J., Chen, C.C. and Fernald, R.D. (2008) Heterogeneous nuclear ribonucleoprotein A/B and G inhibits the transcription of gonadotropin-releasing-hormone 1. *Mol. Cell. Neurosci.*, **37**, 69–84.
- Shin, K.H., Kang, M.K., Kim, R.H., Christensen, R. and Park, N.H. (2006) Heterogeneous nuclear ribonucleoprotein G shows tumor suppressive effect against oral squamous cell carcinoma cells. *Clin. Cancer Res.*, **12**, 3222–3228.
- Shin, K.H., Kim, R.H., Kim, R.H., Kang, M.K. and Park, N.H. (2008) hnRNP G elicits tumor-suppressive activity in part by upregulating the expression of Txnip. *Biochem. Biophys. Res. Commun.*, **372**, 880–885.

28. Martinez-Arribas,F., Agudo,D., Pollan,M., Gomez-Esquer,F., Diaz-Gil,G., Lucas,R. and Schneider,J. (2006) Positive correlation between the expression of X-chromosome RBM genes (RBMX, RBM3, RBM10) and the proapoptotic Bax gene in human breast cancer. *J. Cell. Biochem.*, **97**, 1275–1282.
29. Dichmann,D.S., Fletcher,R.B. and Harland,R.M. (2008) Expression cloning in *Xenopus* identifies RNA-binding proteins as regulators of embryogenesis and RbmX as necessary for neural and muscle development. *Dev. Dyn.*, **237**, 1755–1766.
30. Tsend-Ayush,E., O'Sullivan,L.A., Grutzner,F.S., Onnebo,S.M., Lewis,R.S., Delbridge,M.L., Marshall Graves,J.A. and Ward,A.C. (2005) RBMX gene is essential for brain development in zebrafish. *Dev. Dyn.*, **234**, 682–688.
31. Adamik,B., Islam,A., Rouhani,F.N., Hawari,F.I., Zhang,J. and Levine,S.J. (2008) An association between RBMX, a heterogeneous nuclear ribonucleoprotein, and ARTS-1 regulates extracellular TNFR1 release. *Biochem. Biophys. Res. Commun.*, **371**, 505–509.
32. Matsunaga,S., Takata,H., Morimoto,A., Hayashihara,K., Higashi,T., Akatsuchi,K., Mizusawa,E., Yamakawa,M., Ashida,M., Matsunaga,T.M. *et al.* (2012) RBMX: a regulator for maintenance and centromeric protection of sister chromatid cohesion. *Cell Rep.*, **1**, 299–308.
33. Shin,K.H., Kim,R.H., Kang,M.K., Kim,R.H., Kim,S.G., Lim,P.K., Yochim,J.M., Baluda,M.A. and Park,N.H. (2007) p53 promotes the fidelity of DNA end-joining activity by, in part, enhancing the expression of heterogeneous nuclear ribonucleoprotein G. *DNA Repair*, **6**, 830–840.
34. Adamson,B., Smogorzewska,A., Sigoillot,F.D., King,R.W. and Elledge,S.J. (2012) A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.*, **14**, 318–328.
35. Dreumont,N., Bourgeois,C.F., Lejeune,F., Liu,Y., Ehrmann,I.E., Elliott,D.J. and Stevenin,J. (2010) Human RBMY regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9G8 and Tra2- β . *J. Cell Sci.*, **123**, 40–50.
36. Kanhoush,R., Beenders,B., Perrin,C., Moreau,J., Bellini,M. and Penrad-Mobayed,M. (2010) Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus*, **1**, 109–122.
37. Heinrich,B., Zhang,Z., Raitkin,O., Hiller,M., Benderska,N., Hartmann,A.M., Bracco,L., Elliott,D., Ben-Ari,S., Soreq,H. *et al.* (2009) 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.
38. Skrisovska,L., Bourgeois,C.F., Steff,R., Grellscheid,S.N., Kister,L., Wenter,P., Elliott,D.J., Stevenin,J. and Allain,F.H. (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.*, **8**, 372–379.
39. Geertsma,E.R. and Dutzler,R. (2011) A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry*, **50**, 3272–3278.
40. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93–158.
41. Peterson,R.D., Theimer,C.A., Wu,H. and Feigon,J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR*, **28**, 59–67.
42. Lee,W., Revington,M.J., Arrowsmith,C. and Kay,L.E. (1994) A pulsed field gradient isotope-filtered 3D ¹³C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.*, **350**, 87–90.
43. Herrmann,T., Guntert,P. and Wuthrich,K. (2002) 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.
44. Herrmann,T., Guntert,P. and Wuthrich,K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
45. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
46. Case,D.A., Cheatham,T.E. 3rd, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
47. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
48. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
49. Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55, 29–32.
50. Grellscheid,S.N., Dalglish,C., Rozanska,A., Grellscheid,D., Bourgeois,C.F., Stevenin,J. and Elliott,D.J. (2011) Molecular design of a splicing switch responsive to the RNA binding protein Tra2 β . *Nucleic Acids Res.*, **39**, 8092–8104.
51. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
52. Lorson,C.L. and Androphy,E.J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum. Mol. Genet.*, **9**, 259–265.
53. Liu,Y., Bourgeois,C.F., Pang,S., Kudla,M., Dreumont,N., Kister,L., Sun,Y.H., Stevenin,J. and Elliott,D.J. (2009) The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon. *PLoS Genet.*, **5**, e1000707.
54. Kelly,R.C., Jensen,D.E. and von Hippel,P.H. (1976) DNA “melting” proteins. IV. Fluorescence measurements of binding parameters for bacteriophage T4 gene 32-protein to mono-, oligo-, and polynucleotides. *J. Biol. Chem.*, **251**, 7240–7250.
55. Auweter,S.D., Oberstrass,F.C. and Allain,F.H. (2007) Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities. *J. Mol. Biol.*, **367**, 174–186.
56. Shin,K.H., Kim,R.H., Yu,B., Kang,M.K., Elashoff,D., Christensen,R., Pucar,A. and Park,N.H. (2011) Expression and mutation analysis of heterogeneous nuclear ribonucleoprotein G in human oral cancer. *Oral Oncol.*, **47**, 1011–1016.
57. Michlewski,G., Sanford,J.R. and Caceres,J.F. (2008) The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol. Cell*, **30**, 179–189.
58. Ngo,J.C., Giang,K., Chakrabarti,S., Ma,C.T., Huynh,N., Hagopian,J.C., Dorrestein,P.C., Fu,X.D., Adams,J.A. and Ghosh,G. (2008) A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Mol. Cell*, **29**, 563–576.
59. Ouyang,Y.Q., zur Hausen,A., Orlowska-Volk,M., Jager,M., Bettendorf,H., Hirschfeld,M., Tong,X.W. and Stickeler,E. (2011) Expression levels of hnRNP G and hTra2- β 1 correlate with opposite outcomes in endometrial cancer biology. *Int. J. Cancer*, **128**, 2010–2019.
60. Grellscheid,S., Dalglish,C., Storbeck,M., Best,A., Liu,Y., Jakubik,M., Mende,Y., Ehrmann,I., Curk,T., Rossbach,K. *et al.* (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2 β in development. *PLoS Genet.*, **7**, e1002390.
61. Hegele,A., Kamburov,A., Grossmann,A., Sourlis,C., Wowro,S., Weimann,M., Will,C.L., Pena,V., Luhrmann,R. and Stelzl,U. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell*, **45**, 567–580.
62. Young,P.J., DiDonato,C.J., Hu,D., Kothary,R., Androphy,E.J. and Lorson,C.L. (2002) SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 β 1. *Hum. Mol. Genet.*, **11**, 577–587.