nature
structural &
molecular biology

Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43

Peter J Lukavsky[1,2], Dalia Daujotyte[3,5], James R Tollervey[3,5], Jernej Ule[3,5], Cristiana Stuani[4], Emanuele Buratti[4], Francisco E Baralle[4], Fred F Damberger[2] & Frédéric H-T Allain[2]

TDP-43 encodes an alternative-splicing regulator with tandem RNA-recognition motifs (RRMs). The protein regulates cystic fibrosis transmembrane regulator (*CFTR*) exon 9 splicing through binding to long UG-rich RNA sequences and is found in cytoplasmic inclusions of several neurodegenerative diseases. We solved the solution structure of the TDP-43 RRMs in complex with UG-rich RNA. Ten nucleotides are bound by both RRMs, and six are recognized sequence specifically. Among these, a central G interacts with both RRMs and stabilizes a new tandem RRM arrangement. Mutations that eliminate recognition of this key nucleotide or crucial inter-RRM interactions disrupt RNA binding and TDP-43–dependent splicing regulation. In contrast, point mutations that affect base-specific recognition in either RRM have weaker effects. Our findings reveal not only how TDP-43 recognizes UG repeats but also how RNA binding–dependent inter-RRM interactions are crucial for TDP-43 function.

TAR DNA-binding protein 43 (TDP-43) regulates alternative splicing of human disease–related genes, such as *CFTR* exon 9 (refs. 1,2), *SMN2* exon 7 (ref. 3) and *APOA2* exon 3 (ref. 4). The protein also functions in microRNA (miRNA) processing[5] and mRNA turnover[6–8], regulating its own expression through a negative feedback loop[9–11]. Moreover, the protein localizes to RNA granules, thus suggesting a role in mRNA stabilization and transport[12–15]. In 2006, TDP-43 was identified as a major protein component of cytosolic inclusions characteristic of several neurodegenerative diseases, such as frontotemporal lobar degeneration and amyotrophic lateral sclerosis[16,17]. Currently, it is unclear whether progression of neurodegeneration is linked to gain-of-function or loss-of-function mechanisms[18]. Thus, more detailed knowledge of the basic biological functions and biochemical properties of TDP-43 is essential to improve understanding of the cellular processes leading to the pathology.

TDP-43 is a member of the heterogeneous nuclear ribonucleoprotein (hnRNP) family[19] consisting of an N-terminal domain with a nuclear-localization signal, an RNA-binding domain (RBD) composed of tandem RRMs separated by a 14-aa linker, a nuclear-export signal and a C-terminal glycine-rich protein-binding domain (**Fig. 1a**). The C-terminal domain serves as a platform for interaction with other hnRNPs as well as with Drosha and Dicer complexes and is thus essential for splicing regulation and for promoting miRNA processing[5,20,21].

TDP-43 homologs are conserved throughout evolution, displaying high sequence similarity in the RBD between humans, mice, flies and worms (**Supplementary Fig. 1a**), and all have been shown to bind sequence specifically to UG repeats[2,20]. Recently, two cross-linking–immunoprecipitation (CLIP) sequencing studies confirmed preferential TDP-43 binding to long UG repeats, indicating that it binds as a homodimer[22,23]. UG-rich clusters bound by TDP-43 can be over 100 nt long, thus distinguishing TDP-43 from CUG-binding proteins, such as CELF2, which bind much shorter UG-rich clusters[22]. In addition, TDP-43 also recognizes UG-enriched sequences without long continuous UG repeats such as those found in hairpin loops of pre-miRNAs and the 34-nt TDP-43 binding region within its own 3′ untranslated region[5,10].

Both RRMs are required for high-affinity binding of single-stranded RNA containing a minimum of six UG repeats (K_d = 14.2 nM). The isolated RRMs have distinct binding characteristics: RRM2 binds shorter $(UG)_3$ repeats with higher affinity (K_d = 379 nM), and RRM1 retains most of the affinity of both RRMs for long $(UG)_6$ repeats (K_d = 65.2 nM)[1,24]. A crystal structure of RRM2 in complex with single-stranded DNA has revealed specific recognition of one TG repeat through its β-sheet surface[24] but does not explain the interplay of both RRMs. This interplay is required for binding to longer UG repeats, and it underlies the various biological functions of TDP-43. To assess how TDP-43 recognizes long single-stranded RNA targets, we determined the solution structure of the entire TDP-43 RBD bound to UG-rich RNA and tested the importance of the identified TDP-43–RNA and RRM-RRM interactions for splicing regulation of *CFTR* exon 9.

RESULTS

Overview of TDP-43 RBD bound to UG-rich RNA

We determined the NMR structure of TDP-43 RBD (residues 102–269) bound to the UG-rich RNA oligonucleotide 5′-GUGUGAAUGA AU-3′ (termed AUG12), which we derived from recent individual-nucleotide-resolution CLIP data (**Supplementary Fig. 1b** and

---

[1]Central European Institute of Technology, Masaryk University, Brno, Czech Republic. [2]Institute of Molecular Biology and Biophysics, Eidgenössische Technische Hochschule Hönggerberg, Zürich, Switzerland. [3]Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. [4]International Centre for Genetic Engineering and Biotechnology, Trieste, Italy. [5]Present addresses: Lexogen GmbH, Vienna, Austria (D.D.), Buck Institute, Novato, California, USA (J.R.T.), and Institute of Neurology, University College London, London, UK (J.U.). Correspondence should be addressed to P.J.L. (peter.lukavsky@ceitec.muni.cz) or F.H.-T.A. (allain@mol.biol.ethz.ch).

Received 30 May; accepted 17 September; published online 17 November 2013; doi:10.1038/nsmb.2698

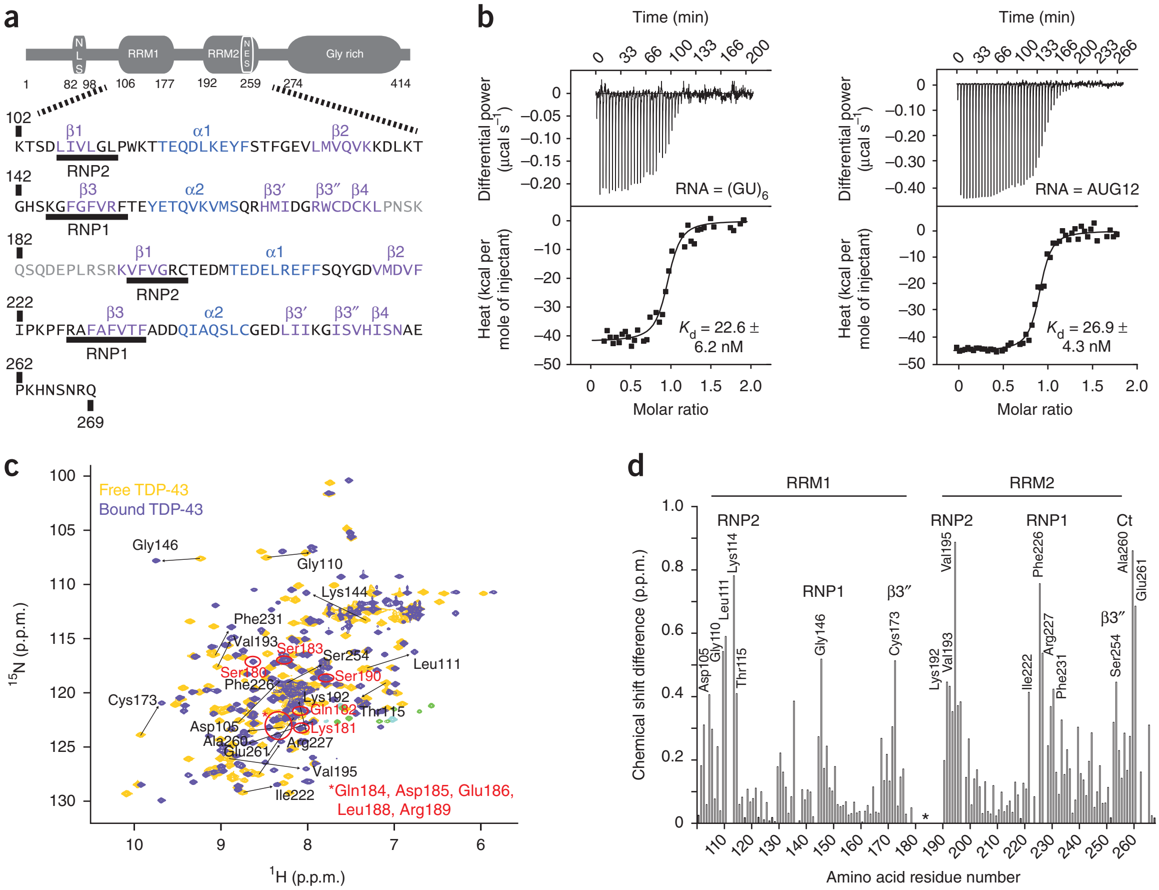

Figure 1 Interaction of TDP-43 RBD with UG-rich RNA. (**a**) Schematic representation of the TDP-43 domains. Domain boundaries are numbered according to the full-length protein sequence (NP_031401). The sequence of recombinant TDP-43 RBD used in this study (aa 102–269), containing the conserved RNP2 and RNP1 motifs, is shown. β-strands and β-hairpins are shown in purple, α-helices in blue and linker amino acids in gray. (**b**) Affinity of TDP-43 RBD for $(GU)_6$ and AUG12 RNA, determined by ITC. The fitted K_d values with fitting errors are indicated. (**c**) NMR spectra of free and bound TDP-43. Superposition of ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled TDP-43 RBD in the free (yellow) and AUG12 RNA-bound (purple) 1:1 complex form. Folded NH$^{\varepsilon}$ cross-peaks from arginine guanidinium groups are shown in light blue (free) and green (bound), respectively. The largest chemical-shift perturbations (>0.4 p.p.m.) observed upon RNA binding are indicated by black arrows. The linker residues (aa 180–190) missing in the free protein but observed in the bound form are circled in red. (**d**) Chemical-shift perturbations of TDP-43 RBD upon TNA binding. The combined chemical-shift perturbations ($\Delta\delta = [(\delta H^N)2 + (\delta N/6.51)2]^{1/2}$) of TDP-43 RBD amide resonances upon AUG12 RNA binding; residues with the largest chemical-shift perturbations are indicated (RNPs, β-hairpins (β3″) and the C terminus (Ct)). The asterisk indicates amino acids located within the linker that are missing in the ^{1}H-^{15}N HSQC of the free protein.

Supplementary Table 1 and **Supplementary Note**)[22]. Isothermal titration calorimetry (ITC) shows that binding of AUG12 RNA occurs at a 1:1 ratio with a similar K_d value of 26.9 nM as compared to 22.6 nM for $(GU)_6$ RNA (**Fig. 1b**). The affinity of $(GU)_6$ RNA determined here is consistent with the previously reported K_d value of 14.2 nM determined by filter binding assay[24]. NMR titration of TDP-43 RBD with AUG12 RNA indicates an intermediate exchange regime and saturation of the binding site at a 1:1 ratio consistent with the results from ITC measurements (**Supplementary Fig. 2a**,**b**). Numerous large chemical-shift changes throughout the ^{1}H-^{15}N HSQC spectrum indicate major structural rearrangements of the protein upon RNA binding (**Fig. 1c**,**d**). Use of ^{13}C-^{15}N–labeled AUG12 RNA prepared by *in vitro* transcription (**Supplementary Note** and **Supplementary Fig. 1c–h**) greatly aided resonance assignments (**Supplementary Fig. 2c**,**d**) and thus yielded a large number of unambiguous intra-RNA (279) and intermolecular (321) NOE-derived distance restraints in addition to 4,748 intraprotein NOEs; this allowed for a high-resolution structure determination of this 25-kDa RNA–protein complex. The final ensemble displays high precision and contains 20 conformers with an r.m.s.d. of 0.90 Å for all heavy atoms (**Fig. 2a** and **Table 1**).

Both RRMs of TDP-43 RBD display the canonical RRM fold (β1α1β2β3α2β4) with an additional β-hairpin (β3′β3″), inserted between α2 and β4, that expands the β-sheet surface accessible for RNA binding (**Fig. 2b**). The AUG12 RNA adopts a single-stranded conformation and binds an extended, positively charged binding groove on the protein surface (**Fig. 2c** and **Supplementary Note**). This groove comprises the canonical binding site on the β-sheet surface of the RRMs, including both β-hairpins as well as the linker between the RRMs and C-terminal residues following RRM2 (**Figs. 1c**,**d** and **2a**,**c**). The AUG12 RNA binds to this groove in a 5′-to-3′ direction from RRM1 to RRM2 (**Fig. 2a–c**). This is rather surprising because tandem RRMs usually bind RNA in a 5′-to-3′ direction from RRM2 to RRM1, as seen in the Sex-lethal[25], PABP[26], HuD[27], Nucleolin[28], Hrp1 (ref. 29) and U2AF65 (ref. 30) tandem RRM–RNA complexes (**Supplementary Fig. 3a**). The commonly seen arrangement of tandem RRMs places the β2 strand of the C-terminal RRM2 spatially close to the β4 strand of the N-terminal RRM1 (β4β2 type) so that the partially helical linker of 11–14 aa needs to bridge only two β strands (**Fig. 2d**). In the TDP-43 tandem RRMs, both the directionality of RNA binding and the RRM orientation are reversed (β2β4 type), and thus the 14-aa linker needs to bridge four

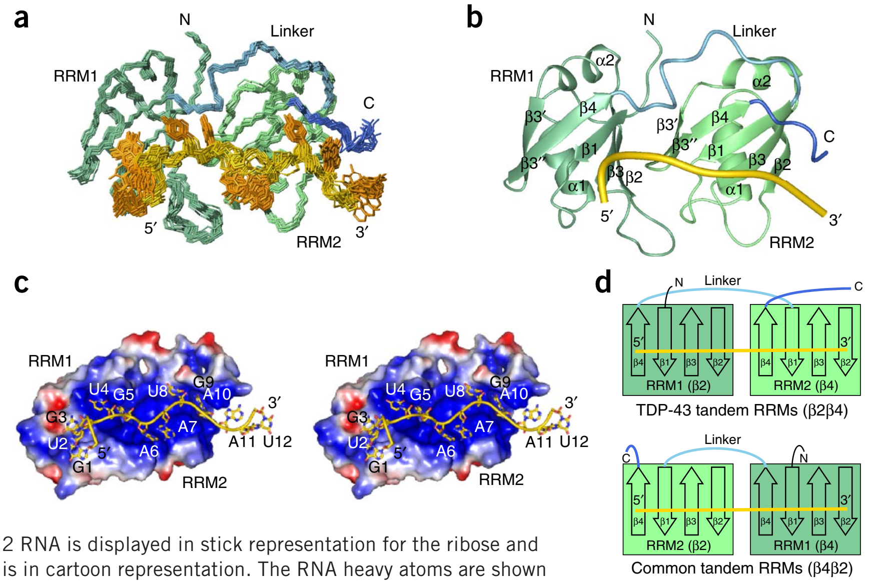

Figure 2 Structural overview of the TDP-43 RBD–AUG12 RNA complex. (**a**) Structural ensemble of the TDP-43 RBD–AUG12 RNA complex. Heavy-atom superposition of the ensemble of the 20 lowest-energy structures (**Table 1**). For the protein backbone, the N terminus and RRM1 are shown in darker green, RRM2 in lighter green, the linker in light blue and the C terminus in dark blue. RNA heavy atoms of the ribose-phosphodiester backbone are shown in gold (omitting phosphate and 2′-OH oxygens) and those of the bases in orange. (**b**) Ribbon representation of the lowest-energy TDP-43 RBD–AUG12 RNA complex structure. The RNA is shown as a golden tube passing through ribose C4′-C3′ atoms. Color scheme as in **a**. (**c**) Wall-eyed stereo image of the lowest-energy TDP-43 RBD–AUG12 RNA complex. The protein surface is colored according to its surface potential, with red and blue indicating negative and positive potential values, respectively. The AUG12 RNA is displayed in stick representation for the ribose and base heavy atoms, and the phosphodiester backbone is in cartoon representation. The RNA heavy atoms are shown in gold (carbon), red (oxygen) and blue (nitrogen). (**d**) Schematic representation of a new tandem RRM arrangement in TDP-43 RBD. The linker path in the new RRM1(β2)-RRM2(β4) arrangement observed in the TDP-43 complex (top) is compared with the common tandem RRM2(β4)-RRM1(β2) arrangement (bottom). Color scheme as in **b**. All figures were generated with PyMOL (http://www.pymol.org/) or MOLMOL[34].

β-strands instead of two and adopts an extended conformation. The need for a longer linker path might explain why this new arrangement (β2β4 type) is found more rarely than is the other type.

^{15}N relaxation studies indicate that in the absence of RNA the two RRMs do not tumble completely independently. The linker shows conformational exchange indicating dynamics at the RRM-RRM interface, which is stabilized upon RNA binding (**Supplementary Table 2** and **Supplementary Note**).

Specific readout of six U or G nucleotides by the TDP-43 RBD

The RNA interaction surface of TDP-43 RBD accommodates 10 out of 12 nt of AUG12 RNA (**Fig. 2a**,**c**). Nucleotides G1, U2, G3, U4 and G5 lie on the RRM1 β-sheet surface, A6 and A7 stack and act as a spacer between the two RRMs, and U8, G9 and A10 interact with RRM2 and C-terminal residues, whereas A11 and U12 do not make any contacts with the protein (**Fig. 2c**). Of the 10 nt in contact with the protein surface, 6 nt are recognized sequence specifically (G1, G3, U4, G5, U8 and G9). All are Us or Gs, and this explains the binding preference of TDP-43 for UG-rich sequences.

On RRM1, the 5′ residue G1 stacks with Trp113 from loop 1 of RRM1 and shows an intramolecular interaction with G3 N7 through its amino group (**Fig. 3a**). Through this intra-RNA hydrogen bond, G1 is indirectly sequence specifically recognized by the RRM, because only a G base can form such a hydrogen bond (**Supplementary Fig. 2d** and **Supplementary Note**). This structural organization is reminiscent of but not identical to the intramolecular contacts between G2 and A4 or G2 and G4 found in the RRM of Fox-1 (ref. 31) or the qRRM1 of hnRNP F[32], respectively (**Supplementary Figs. 3b** and **4a**). G3 adopts a characteristic *syn* conformation and is placed on the β-sheet surface for specific recognition through hydrogen bonds to the side chains of Arg171 (β3′β3″) and Asp174 (β4) (**Fig. 3a**). The ribose moiety of G3 is held in place through hydrogen bonds of the 2′-OH group with both the Gly110 amide proton and the Lys145 backbone carbonyl oxygen as well as through hydrophobic interactions with Leu109 from RNP2 (β1). The following residue, U4, is also specifically recognized. Hydrophobic interactions with Ile107 from RNP2 (β1) and Phe147 from RNP1 (β3) position the base and ribose moieties on the β-sheet surface (**Fig. 3a**). The entire Watson–Crick face of U4 is read out by hydrogen-bonding with Lys176 (β4), the backbone carbonyl oxygen of Leu177 and the main chain and side chain of Asn179, both from the inter-RRM linker (**Fig. 3a**).

To test the importance of the interactions of G1, G3 and U4 with RRM1, we replaced the key interacting residues of TDP-43 RBD individually with alanine and tested the effect of these substitutions on binding affinity by ITC. The Trp113 mutation results in a modest three-fold loss in binding affinity compared to that of wild-type TDP-43 RBD, whereas double amino acid substitutions disrupting the specific readout of the bases of G3 (R171A D174A) or U4 (K176A N179A) lead to a substantial 20-fold and 6-fold reduction in affinity, respectively (**Table 2**).

RRM2 shows only two nucleotide-specific interactions, namely with U8 and G9. Stacking interactions of Phe194 from RNP2 (β1) with U8 and of Phe231 from RNP1 (β3) with G9, together with hydrophobic interactions of Phe229 from RNP1 (β3) and of Phe221 (β2) with the U8 and G9 ribose moieties, position these two nucleotides on the β-sheet surface (**Fig. 3b**). Similarly to U4 in RRM1, the entire Watson–Crick face of U8 is involved in specific interactions to residues from RNP1 and RNP2 in RRM2. U8 is recognized by hydrogen bonds with Ser258 (β4), the backbone carbonyl oxygen of Asn259 and the backbone amide of Glu261 from the C terminus (**Fig. 3b**). The G base of G9 also forms multiple hydrogen bonds with the protein. Base-specific hydrogen bonds are present with Lys192 (β3) and with the backbone carbonyl oxygen of Glu261 of the C terminus (**Fig. 3b**). Moreover, a sequential RNA hydrogen bond is formed between the amino group of G9 and the U8 phosphate oxygen, and G9 contacts the Lys263 side chain from the C terminus: the aliphatic chain stacks on top of the base, and the ammonium group forms a hydrogen bond with the 2′-OH (**Fig. 3b**). The side chains interacting with G9 form a pronounced cavity into which G9 is deeply buried, and this requires G9 to adopt a *syn* conformation similar to that of G3 in RRM1 ((**Fig. 3b**), **Supplementary Fig. 4b** and **Supplementary Note**). Finally, the A10 base and ribose protons display several NOEs to side chain protons of Phe221 and Lys263, thus indicating interaction of A10 with the C terminus of TDP-43 RBD, but we could

Table 1 Structural statistics of the TDP-43 RBD in complex with AUG12 RNA

	Protein	RNA–protein	RNA
NMR distance and dihedral constraints			
Distance constraints			
Total NOE	4,748		279
Intra-residue	956		198
Inter-residue			
Sequential ($\lvert i-j\rvert = 1$)	1,232		74
Medium-range ($\lvert i-j\rvert < 5$)[a]	962		7
Long-range ($\lvert i-j\rvert > 5$)[a]	1,598		0
Intermolecular		321	
Hydrogen bonds[b]	48	8	0
Total dihedral angle restraints[c]	0		3
ϕ	0		–
ψ	0		–
Structure statistics[d]			
Violations (mean ± s.d.)			
Distance constraints (Å) > 0.3		0.3 ± 0.5	
Dihedral angle constraints (°) > 5		0.0	
Max. dihedral angle violation (°)		0.29 ± 0.04	
Max. distance constraint violation (Å)		0.21 ± 0.70	
Deviations from idealized geometry			
Bond lengths (Å)		0.0037 ± 0.0001	
Bond angles (°)		1.619 ± 0.012	
Average pairwise r.m.s. deviation (Å)[d]			
Heavy	0.81 ± 0.13	1.33 ± 0.20	0.91 ± 0.12
Backbone	0.57 ± 0.12	–	–

[a] Medium- and long-range NOEs include 46 NOEs between the N terminus and RRM1 (33 NOEs) and the linker (13); 36 NOEs between RRM1 and RRM2; and 175 NOEs between the linker and RRM1 (58), RRM2 (81) and the C terminus (36). [b] Hydrogen-bond constraints were identified in the course of structure refinement for slowly exchanging amide protons in D_2O and imino resonances protected from exchange with H_2O. [c] Dihedral-angle constraints for three χ angles (G1, G3 and G9) were derived from H1′-H8 cross-peak intensities in a 2D NOESY spectrum acquired with 50-ms mixing time. [d] Protein and RNA r.m.s.d. values were calculated with residues Lys 102–267 for the protein and 1–10 for the RNA for the ensemble of 20 refined structures from a total of 50 calculated structures.

not deduce any sequence-specific A10-protein interactions from the ensemble of structures.

Single- and double-alanine point mutants of TDP-43 RRM2 show only weak effects on the overall binding affinity (**Table 2**). Taken together, these data agree with previous results that indicated a weaker contribution to the overall TDP-43 RNA binding affinity from RRM2 compared to that of RRM1 and suggest a different functional role for the two domains[1,24].

G5 stabilizes RRM interactions crucial for RNA binding

Three nucleotides of AUG12 RNA, namely G5, A6 and A7, are located at the interface between the two TDP-43 RRMs (**Figs. 2c** and **3c**). Whereas the bases of A6 and A7 point toward the outside, G5 is buried in a deep central pocket at the junction between the two β-sheets (**Figs. 2c** and **3c**). All three nucleotides seem to act as a 'glue' between the two RRMs, with G5 being contacted by side chains from both domains while A6 and A7 stack and contact basic side chains (A6 to Lys145 of RRM1 and A7 to Arg227 of RRM2), thereby bridging the two β2-β3 loops. Arg197 in RRM2 further contacts the A7 phosphate between the two nucleotides. RNA binding induces many inter-RRM interactions including a small hydrophobic core formed by Leu131, Met132 (in RRM1) and Ile249 (RRM2); a salt bridge involving Arg151 (RRM1) and Asp247 (RRM2); and a hydrogen bond between the side chain of Lys136 (RRM1) and the carbonyl oxygen of Arg197 (**Fig. 3d** and **Supplementary Fig. 4c**).

G5 adopts an *anti* conformation and stacks with Phe149 from RNP1 (β3) of RRM1 (**Fig. 3c**). The Watson–Crick face is hydrogen-bonded to the amide proton and the carboxyl group of Asp105 (β1 on RRM1) and the side chain hydroxyl of Ser254 (β-hairpin of RRM2) (**Fig. 3c**). In addition, the G5 sugar-phosphate backbone interacts with the side chain of Lys145 from RRM1. A6 and A7 are not recognized in a sequence-specific manner (**Fig. 3c** and **Supplementary Note**). In contrast, sequence-specific contacts to G5 are absolutely crucial for binding. Disruption of the G5 base-stacking (F149A) or the recognition of the Watson–Crick face of G5 (D105A S254A) leads to strong decreases in affinity, by 15- or 18-fold, respectively (**Table 2**).

Likewise, we observed a strong decrease in affinity when the inter-RRM interactions were disrupted in the vicinity of G5. Altering the hydrophobic contacts between Met132 and Ile249 (M132A I249A) leads to a seven-fold decrease in affinity, and eliminating a single salt bridge between Arg151 and Asp247 (R151A D247A) is very deleterious to RNA binding (>37-fold decrease in affinity) (**Table 2**). For a control, we also mutated two interface residues that are not in direct contact, namely Leu139 and Glu200 (**Supplementary Fig. 4c**), and this leads to only a modest, three-fold decrease in affinity (**Table 2**). We can conclude, on the basis of these data, that recognition of the central G nucleoside G5 is crucial for complex formation, because it requires contacts to both RRMs. Also, G5 binding apparently induces inter-RRM interactions that are not established in the absence of RNA (described above).

Taken together, our data show that only nucleotides G1, G3, U4, G5, U8 and G9 are sequence-specifically recognized by the TDP-43 RBD surface, whereas U2, A6, A7 and A10 are not. Furthermore, G5, A6 and A7 function in stabilizing the complex by mediating inter-RRM interactions. Our data define a minimal consensus sequence, 5′-GNGUGNNUGN-3′, recognized by TDP-43 for high sequence specificity and binding affinity (**Supplementary Note**).

Regulation of *CFTR* exon 9 splicing by TDP-43

To assess the functional importance of the RNA-protein contacts revealed by our solution NMR structure of the TDP-43 RBD in complex with UG-rich RNA, we performed *in vivo* splicing assays, using a *CFTR* exon 9–splicing minigene reporter system. The latter contains the *CFTR* exon 9 sequence, the splicing junctions and part of the flanking introns, with the TG11-T5 repeat at the 3′ splice site constituting the functional TDP-43–binding site (**Fig. 4a**)[2]. As previously described[21], the splicing assay was coupled with RNA interference–mediated knockdown of endogenous TDP-43 and add-back of short interfering RNA (siRNA)-resistant wild-type or mutant full-length TDP-43 protein to characterize structural determinants of TDP-43 splicing function. Knockdown of endogenous TDP-43 leads to 54% exon 9 inclusion (**Supplementary Fig. 5a** and **Supplementary Table 3**). Overexpression of siRNA-resistant wild-type (WT) or F4L-mutant TDP-43 (with F147L, F149L, F229L and F231L mutations), which completely abolish RNA binding affinity[1], leads to 24% (for functional TDP-43 WT) or 62% (for functionally impaired TDP-43 F4L) exon 9 inclusion, corresponding to a 2.5-fold increase

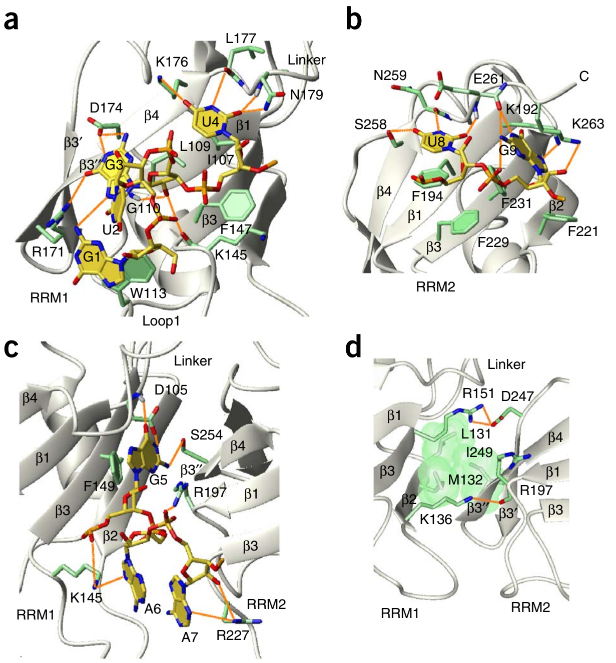

Figure 3 Recognition of UG-rich RNA by TDP-43 RBD. (**a**) Recognition of AUG12 RNA nucleotides G1–U4 by TDP-43 RRM1. The structure of the complex is displayed in gray ribbon (protein backbone) and stick (RNA) representation. In all panels, the heavy atoms of interacting protein side chain and backbone atoms are displayed in green (carbon), red (oxygen), light gray (sulfur), blue (nitrogen) and gray (amide protons). The RNA heavy atoms are shown in gold (carbon), red (oxygen), orange (phosphorus) and blue (nitrogen). Hydrogen bonds are represented by orange dashed lines. (**b**) Recognition of AUG12 RNA nucleotides U8 and G9 by TDP-43 RRM2. (**c**) Recognition of AUG12 RNA nucleotides G5–A7 at the TDP-43 RRM1-RRM2 interface. (**d**) Contacts between RRMs of TDP-43 RBD. For hydrophobic side chains of L136, M137 and I249, the Van der Waals surface is displayed in light green. The main interdomain contacts are displayed. In all figures, only the intermolecular interactions that are most commonly observed in the ensemble of structures are highlighted.

as compared to that of WT (**Fig. 4b**, **Supplementary Fig. 5a** and **Supplementary Table 3**).

We evaluated the role of specific amino acids involved in RNA recognition for splicing activity by testing the ability of alanine mutants of TDP-43 to rescue splicing activity. With the exception of Lys263, all tested residues important for recognition showed a weaker repressing effect with 1.1- to 1.9-fold increase in exon 9 inclusion as compared to that of WT (**Fig. 4b**). In addition, we combined mutations of amino acids mediating specific recognition of G3, U4 and G9 and tested their effect on exon 9 inclusion (**Fig. 4b**,**c**). The double mutant R171A D174A, for example, which impairs G3 recognition and results in a 20-fold decrease in RNA affinity, shows a 1.8-fold increase in exon 9 inclusion (**Table 2** and **Fig. 4b**).

Comparison of the results from affinity measurement (**Table 2**) and splicing repression (**Supplementary Table 3**) suggests that RNA binding affinity is not the only property controlling the strength of splicing repression by TDP-43. Sequence-specific interactions of U8 and G9 on RRM2, which contribute much less to the overall RNA binding affinity of TDP-43 compared to those of RRM1 (**Table 2**), are of equal functional importance for splicing as are interactions of G1, G3 and U4 on RRM1, which provide high-affinity binding. In fact, despite the modest loss in binding affinity observed for F194A (U8) or K192A E261A (G9), we observed 1.8- and 1.5-fold increases in exon 9 inclusion, respectively (**Table 2** and **Fig. 4b**,**c**), showing that specific RNA recognition by both RRMs is crucial for the splicing activity of TDP-43.

Specific recognition of the central G nucleoside G5 is essential both for high RNA binding affinity and splicing activity of TDP-43 (**Fig. 4b**). The single mutant F149A and the double mutant D105A S254A, which display 15- and 18-fold loss in binding affinity, respectively, also impair splicing activity and lead to 2.7- and 2.4-fold increases, respectively, in exon 9 inclusion (**Fig. 4b**,**c**). We observed essentially no splicing repression for these mutants. In addition, mutation of Arg197, which interacts with the phosphate group between A6 and A7, leads to only a four-fold loss of RNA binding affinity but a large 2.1-fold increase in exon 9 inclusion (**Table 2** and **Fig. 4b**), results suggesting that neutralization of the phosphate charge at the RRM-RRM interface is also required for TDP-43 splicing activity.

Our structure also revealed RRM-RRM interactions that are induced upon interaction of TDP-43 RBD with UG-rich RNA and are essential for RNA binding (**Fig. 3d**). Single and double mutants of TDP-43 targeting the hydrophobic contact between Met132 and Ile249 and the salt bridge between Arg151 and Asp247 show large (1.7- to 2.5-fold, respectively) levels of exon 9 inclusion, thus indicating, again, almost no repression (**Fig. 4b**,**c** and **Supplementary Table 3**). These results also reflect the mutants' greatly reduced RNA binding affinity (**Table 2**) and underline the importance of these RRM-RRM contacts induced by RNA binding for TDP-43 function

Table 2 Binding affinities of wild-type and mutant TDP-43 RBD and UG-rich RNA

Protein[a]	RNA[b]	K_d (nM)[c]	Affinity factor[d]
Wild type	GUGUGUGUGUGU	22.6 ± 6.2	0.8
Wild type	GUGUGAAUGAAU (AUG12)	26.9 ± 4.3	1.0
Wild type	GUGUGAUGAAU (AUG11)	33.2 ± 4.4	1.2
Wild type	GUGUGUGUGAAU (AUG12-UG)	32.4 ± 5.9	1.2
Wild type	GUGUGUGAAUGAAU (AUG14)	70.0 ± 9.6	2.6
W113A (G1)	AUG12	85.5 ± 14.8	3.2
R171A D174A (G3)	AUG12	529.1 ± 26.1	19.7
K176A ASN179A (U4)	AUG12	155.8 ± 9.6	5.8
F149A (G5)	AUG12	393.7 ± 15.8	14.6
D105A S254A (G5)	AUG12	471.7 ± 36.5	17.5
R197A (A7)	AUG12	109.4 ± 13.4	4.1
S258A (U8)	AUG12	27.5 ± 4.3	1.0
F194A (U8)	AUG12	122.1 ± 8.3	4.5
K263A (G9)	AUG12	35.6 ± 7.9	1.3
K192A E261A (G9)	AUG12	64.9 ± 9.2	2.4
K181A (rrm12)	AUG12	28.8 ± 3.4	1.1
R151A D247A (rrm12)	AUG12	>1,000	>37
M132A I249A (rrm12)	AUG12	191.9 ± 21.7	7.1
L139A E200A (control)	AUG12	66.7 ± 7.9	2.5

[a]AUG12 RNA residue recognition or RRM interactions affected by TDP-43 mutation based on observed interactions in the structural ensemble of the complex are indicated in parentheses. [b]Mutated residues in comparison to AUG12 RNA are underlined. [c]All complexes formed with 1:1 stoichiometry ($n = 1$), except for K176A, N179A, R197A, M132A and I249A, which showed best fit at $n = 0.5$. Experiments were performed in technical duplicates, and the K_d values are from one representative experiment and are reported with fitting error. [d]Affinity factors for variant RNA or protein relative to wild-type TDP-43–AUG12 RNA binding.

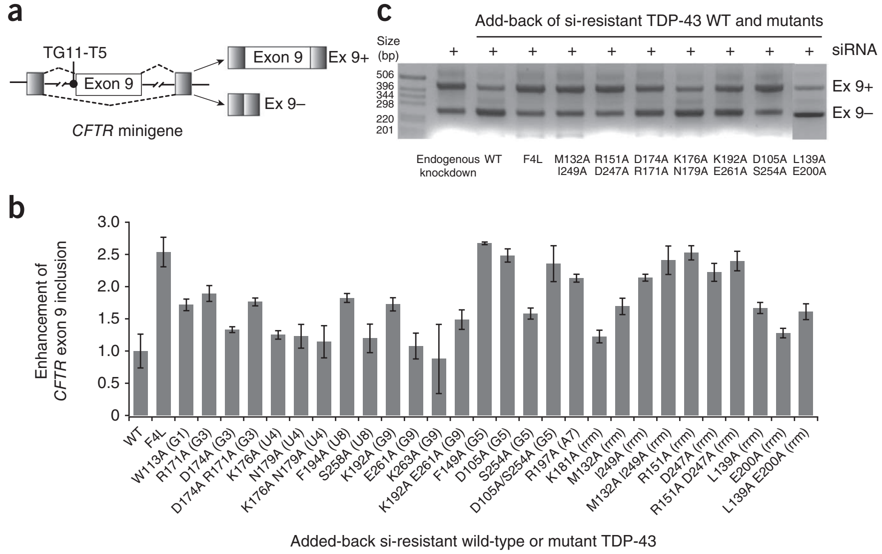

Figure 4 Impairment of *CFTR* exon 9 splicing by alanine mutagenesis of residues at the RRM-RRM or TDP-43 RBD–AUG12 RNA interfaces. (**a**) Schematic representation of the *CFTR* exon 9 (ex 9) minigene construct used in transfection experiments. Lines represent introns; open and shaded boxes, exons. The intronic TG11-T5–repeat TDP-43–binding site is indicated. The two major splice forms are shown on the right. (**b**) *CFTR* exon 9 inclusion upon knockdown of the endogenous TDP-43 protein and add-back of siRNA (si)-resistant wild-type TDP-43 (WT), an RNA binding–impaired mutant (F4L) and TDP-43 variant proteins carrying single- or double-alanine substitutions. The splicing assays were performed in biological triplicates (three individual transfection experiments). The average enhancement by *CFTR* exon 9 inclusion normalized to WT levels is shown. Error bars, s.d. from three independent transfection experiments. Mutations, as well as the interaction affected by the corresponding mutation, are indicated along the *x* axis. (**c**) Agarose gel analysis of *CFTR* exon 9 splicing efficiency of double-alanine TDP-43 mutants. Levels of *CFTR* exon 9 inclusion (Ex 9+) in each sample measured by densitometric scanning and analyzed with ImageJ are shown. The last lane with the double mutant L139A E200A originates from a different agarose gel, which was run and analyzed under the same conditions. Further details and the uncropped original gel are in **Supplementary Figure 5**.

in splicing. In contrast, alanine mutation of Leu139 and Glu200, which are also located in the binding cleft between the RRMs but do not mediate any RRM-RRM interactions, show weaker effects on both RNA binding and splicing activity (**Table 2** and **Fig. 4b**).

In summary, this implies that splicing repression by TDP43 is defined by three features of the interaction with UG-rich RNA: (i) high RNA binding affinity mainly provided by RRM1, (ii) sequence-specific recognition of UG-repeats encoded by both RRMs and (iii) precise positioning of the two RRMs relative to each other, mediated by the intermolecular contacts to G5, A6 and A7 and by the induced inter-RRM interactions.

DISCUSSION

We determined the solution NMR structure of the TDP-43 tandem RRMs in complex with UG-rich RNA and show the molecular details of target RNA recognition by this human alternative-splicing factor. Most RRM proteins contain multiple RRM units separated by short linkers to create a continuous surface for recognition of longer RNA sequences of 8–10 nt with high affinity and specificity[33]. Our structure shows that the way in which this surface is created and stabilized is very different in TDP-43 compared to other tandem RRMs. Normally, RNA complexes of common tandem RRMs[25–27,29] use the shorter path of the linker from the β4 strand of RRM1 to the β1 strand of RRM2 (β4β2 type), spanning just two β-strands to position the two RRMs so that RRM2 binds the 5′ side of the RNA and RRM1 binds the 3′ side. However, in the complex of TDP-43 tandem RRMs, the linker takes the longer path (β2β4 type) with the opposite positioning of the RRMs relative to the RNA directionality (**Fig. 2d** and **Supplementary Fig. 3a**). This longer linker path allows the linker to extensively participate in interactions with the RNA, RRM1 and RRM2 as well as with the N and C termini of the protein (**Fig. 3**, **Supplementary Fig. 4c,d** and **Table 1**). This alternative arrangement could be important to position the protein termini upon target-RNA binding for interaction with protein partners, such as hnRNP A2, which are required for splicing regulation by TDP-43 (refs. 20,21).

In addition, the way in which RNA binding mediates RRM-RRM interaction in TDP-43 is unique among the proteins containing tandem RRMs. So far, a fixed positioning between RRMs could be mediated by the RNA itself, as in Sex-lethal[25], HuD[27] or Nucleolin[28], for example. In the case of these tandem RRMs, the RRMs move independently in the free state, and upon binding RNA they adopt a fixed position relative to one another (**Supplementary Fig. 3a**). The RNA acts as glue between the two RRMs, but very few inter-RRM interactions are found in these complexes. The other extreme case is represented by PABP[26], in which the RRMs already interact with each other in the free state through a large interdomain interface, and the RNA lies on this extended, preformed RNA-binding platform (**Supplementary Fig. 3a**). In the TDP-43–RNA complex, however, both extensive, direct inter-RRM interactions and interactions mediated by the RNA (G5, A6 and A7) are present (**Supplementary Fig. 3a**). Also unique is the requirement for interactions with both RRMs in G5 sequence-specific recognition. Thus, in TDP-43 the arrangement of the RRMs with respect to one another is dictated by sequence-specific recognition of G5. This fixing of the relative orientation of the RRMs is clearly important for splicing activity *in vivo*, because mutations affecting the inter-RRM interface or the contact to G5 or A7 have a disproportionately strong effect on exon 9 inclusion. These observations suggest that splicing activity may be regulated in TDP-43 because the RRM-RRM interactions induced by binding of the central residues are necessary for the correct positioning of RRMs and RNA and potentially for binding of additional protein partners in the splicing reaction.

The structure of the complex reveals that a consensus sequence (5′-GNGUGNNUGN-3′) is recognized by TDP-43 for high sequence specificity and binding affinity (**Supplementary Note**). This consensus motif can now explain the identification of TDP-43 RNA-binding sites that do not match the motif of long continuous UG repeats. This degenerate RNA-recognition consensus sequence of TDP-43 clearly contrasts with other tandem RRMs that create high affinity and specificity through the continuous recognition of 6–9 nt (refs. 25–27,29).

Our combined structural, biophysical and functional investigations of TDP-43 binding to UG-rich RNA also suggest a strong correlation between binding affinity and inhibitory splicing function for interaction mediated by RRM1 but not RRM2. In isolation, the second RRM, as compared to RRM1, was previously shown to bind UG-rich RNA weakly, and we show that alanine mutation in the recognition sites on RRM2 has only little impact on the overall RNA binding affinity of TDP-43. Nevertheless, both RRM1 and RRM2 interactions are nucleotide specific for UG repeats and thus contribute to TDP-43 splicing function. The possible role of RRM2 could be to help direct the path of bound UG-rich RNA for productive TDP-43 dimerization on pre-mRNA target sites. This in turn, together with reversed directionality of RNA binding, could be essential for properly directing the recruitment of other less sequence-specific splicing factors, such as hnRNP A2, to their target sites on the pre-mRNA to coordinate alternative splicing. Finally, our structural insights also lay the foundation for future investigations of the other multiple biological functions of TDP-43 in RNA metabolism and how their loss can lead to neurodegeneration.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The chemical shifts of the TDP-43 RBD–AUG12 RNA complex have been deposited in the Biological Magnetic Resonance Data Bank under accession number 19290. The coordinates of the structural ensemble were deposited in the Protein Data Bank under accession number 4BS2.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank C. Maris for help with setup of filtered NMR experiments and denaturing RNA purification; M. Blatter and D. Theler for help with structure calculation and validation; and A. Clery and J. Boudet for help with ITC measurement and data analysis. We also thank S. Chesnov from the Functional Genomics Center Zürich for MS analyses. This work was supported by the project 'Central European Institute of Technology' (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund, by grants from the Human Frontier Science Program (RGP0024/2008) and by the Internationalization of the Structural Biology Research Program (CZ.1.07/2.3.00/20.0042) to P.J.L. and from Agenzia di Ricerca per la Sclerosi Laterale Amiotrofica (TARMA project) and Thierry Latran Foundation (REHNPALS) to C.S., E.B. and F.E.B. and by a Sinergia grant from the Swiss National Science Foundation (CRSII3 136222) to F.H.-T.A.

AUTHOR CONTRIBUTIONS

P.J.L., D.D., E.B. and F.H.-T.A. designed the project. J.R.T. and J.U. performed and analyzed the individual-nucleotide-resolution CLIP data. D.D. and J.R.T. expressed the initial TDP-43 protein construct and performed band-shift assays to optimize the UG-rich RNA sequence. P.J.L. optimized the TDP-43 protein construct, prepared all samples for structural studies and ITC measurements and performed all NMR experiments, NMR data analysis, structure calculations and ITC measurements and analysis. F.F.D. assisted with setup of specialized NMR experiments. F.F.D. and F.H.-T.A. helped with NMR data analysis and F.F.D. with structure calculation and validation. C.S. and E.B. designed and performed the minigene add-back experiments; E.B. and F.E.B. participated in their interpretation. P.J.L., F.F.D., E.B. and F.H.-T.A. wrote the manuscript; all authors discussed the results and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Buratti, E. & Baralle, F.E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276**, 36337–36343 (2001).
2. Buratti, E. *et al.* Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping. *EMBO J.* **20**, 1774–1784 (2001).
3. Bose, J.K., Wang, I.F., Hung, L., Tarn, W.Y. & Shen, C.K. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J. Biol. Chem.* **283**, 28852–28859 (2008).
4. Mercado, P.A., Ayala, Y.M., Romano, M., Buratti, E. & Baralle, F.E. Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res.* **33**, 6000–6010 (2005).
5. Kawahara, Y. & Mieda-Sato, A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl. Acad. Sci. USA* **109**, 3347–3352 (2012).
6. Ayala, Y.M., Misteli, T. & Baralle, F.E. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *Proc. Natl. Acad. Sci. USA* **105**, 3785–3789 (2008).
7. Fiesel, F.C. *et al.* Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. *EMBO J.* **29**, 209–221 (2010).
8. Godena, V.K. *et al.* TDP-43 regulates *Drosophila* neuromuscular junctions growth by modulating Futsch/MAP1B levels and synaptic microtubules organization. *PLoS ONE* **6**, e17808 (2011).
9. Avendaño-Vazquez, S.E. *et al.* Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* **26**, 1679–1684 (2012).
10. Ayala, Y.M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277–288 (2011).
11. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **14**, 459–468 (2011).
12. Wang, I.F., Wu, L.S., Chang, H.Y. & Shen, C.K. TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *J. Neurochem.* **105**, 797–806 (2008).
13. Colombrita, C. *et al.* TDP-43 is recruited to stress granules in conditions of oxidative insult. *J. Neurochem.* **111**, 1051–1061 (2009).
14. Dewey, C.M. *et al.* TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Mol. Cell Biol.* **31**, 1098–1108 (2011).
15. McDonald, K.K. *et al.* TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum. Mol. Genet.* **20**, 1400–1410 (2011).
16. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
17. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
18. Buratti, E. & Baralle, F.E. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biol.* **7**, 420–429 (2010).
19. Krecic, A.M. & Swanson, M.S. hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363–371 (1999).
20. Ayala, Y.M. *et al.* Human, *Drosophila*, and *C. elegans* TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **348**, 575–588 (2005).
21. D'Ambrogio, A. *et al.* Functional mapping of the interaction between TDP-43 and hnRNP A2 *in vivo*. *Nucleic Acids Res.* **37**, 4116–4126 (2009).
22. Tollervey, J.R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).
23. Xiao, S. *et al.* RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell Neurosci.* **47**, 167–180 (2011).
24. Kuo, P.H., Doudeva, L.G., Wang, Y.T., Shen, C.K. & Yuan, H.S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808 (2009).
25. Handa, N. *et al.* Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
26. 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).
27. Wang, X. & Tanaka Hall, T.M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141–145 (2001).
28. Allain, F.H., Bouvet, P., Dieckmann, T. & Feigon, J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870–6881 (2000).
29. Pérez-Cañadillas, J.M. Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. *EMBO J.* **25**, 3167–3178 (2006).
30. Mackereth, C.D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408–411 (2011).
31. Auweter, S.D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
32. Dominguez, C., Fisette, J.F., Chabot, B. & Allain, F.H. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* **17**, 853–861 (2010).
33. Cléry, A., Blatter, M. & Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
34. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, 29–32 (1996).

ONLINE METHODS

cDNA cloning of TDP-43 RBD and PCR site-directed mutagenesis. The cDNA fragment encoding human TDP-43 RBD (102–269) was prepared by PCR from GST-TDP-43 (ref. 2) with appropriate primers and subcloned into pET28a vector (Novagen) with a TEV instead of the original thrombin-cleavage site with the NheI and XhoI restriction sites. Protein mutants were obtained by PCR-based site-directed mutagenesis with the pET28a TDP-43 RBD (102–269) plasmid as a DNA template according to the QuikChange protocol (Stratagene).

Expression and purification of the recombinant TDP-43 RBD protein. Human TDP-43 RBD (102–269) recombinant protein and mutant proteins were expressed in *E. coli* BL21(DE3) codon plus cells (Novagen) in LB rich or M9 minimal media supplemented with $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-glucose. Protein expression was induced at an OD_{600} of 0.8 by addition of 1 mM isopropyl β-D-thiogalactoside (IPTG) and further incubation for 14 h at 16° C. Cells were harvested by centrifugation at 4°C, 15 min at 2,600*g*, and the cell pellet was resuspended in lysis buffer (20 mM HEPES, pH 7.5, 1 M NaCl, 10% glycerol (w/v), 30 mM imidazole and 5 mM 2-mercaptoethanol). After lysis with Emulsiflex (Avestin) and centrifugation at 4 °C for 20 min at 43,000*g*, the supernatant was loaded onto two 5-mL HiTrap chelating columns (HiTrap, GE Healthcare) in series, charged with nickel sulfate. After column loading, the column was washed with 50 ml of lysis buffer, and protein was eluted over a 90-ml gradient with elution buffer (lysis buffer with 300 mM imidazole). For cleavage of the His_6 tag, the pooled fractions were dialyzed against 2 L of lysis buffer (500 mM NaCl and no imidazole) in the presence of TEV protease (1 mg/100 ml) at RT overnight. We reloaded TEV cleavage reaction mixtures onto HiTrap chelating columns to remove the His_6-TEV protease and the His_6-tag fusion as well as minor contaminating proteins. After purification, the proteins were dialyzed against 3× 1 L storage buffer (50 mM potassium phosphate buffer, pH 6.8, 300 mM NaCl and 5 mM 2-mercaptoethanol), frozen and stored at −20 °C.

RNA transcription and purification. Unlabeled RNA oligonucleotides were purchased from Dharmacon, deprotected according to the manufacturer's instructions, lyophilized and resuspended in water or NMR buffer (50 mM potassium phosphate buffer, pH 6.8, 2.5 mM 2-mercaptoethanol).

For isotope-labeled RNA oligonucleotides (5′-GUGUGAAUGAAU-3′) a *cis*-hammerhead construct was prepared by PCR with two overlapping primers (5′-GAGCAAGCTTAATACGACTCACTATAGGGAGAAATTCATTCACACCTGATGAGTCCGTGAGGACGAAAG-3′ and 5′-gctcgaattcgaagacaaattcattcacactagaccgaagtctttcgtcctcacggactcatcagg-3′) and subcloned into pUC18 as previously described[35]. Labeled RNA oligonucleotides (^{15}N or ^{13}C-^{15}N) were transcribed *in vitro* from linearized plasmid DNA with T7 RNA polymerase as previously described[36]. Optimal magnesium chloride concentrations for efficient transcription (14 mM) and efficient hammerhead cleavage (18 mM) were determined by 25-μL trial reactions. After 4 h of incubation at 37°C, the reaction mixture was boiled for 5 min and slowly cooled to RT, and the magnesium chloride concentration was increased to 18 mM to complete hammerhead cleavage overnight at RT. All RNAs were purified by denaturing anion-exchange chromatography on a preparative Dionex DNAPac PA-100 column (22 × 250 mm) at 85 °C (refs. 37,38). After removal of salts and urea by *n*-butanol extraction, the salt-free RNA was lyophilized and resuspended in water or NMR buffer.

Preparation of RNA–protein complexes. Purified TDP-43 RBD and 5′-GUGUGAAUGAAU-3′ RNA were assembled into complexes at a 1:1.1 molar ratio in the presence of RNase inhibitor (Ambion), concentrated with Vivaspin 20-mL centrifugal devices with 5,000 MWCO (Sartorius) and passed over a Sephadex S-200 gel-filtration column (GE Healthcare) with 20 mM HEPES, pH 7.5, 150 mM NaCl and 5 mM 2-mercaptoethanol as gel-filtration buffer at 4 °C. Fractions corresponding to the 1:1 complex were pooled, concentrated to 400–700 μM and dialyzed two times against 2 L NMR buffer with Slide-A-Lyzer 0.5-mL dialysis cassettes with 5,000 MWCO (Thermo Scientific).

NMR measurement. All NMR measurements for the RNA–protein complex were performed in NMR buffer at 298 K unless otherwise noted with Bruker AVIII-500 MHz, AVIII-600 MHz, AVIII-700 MHz and Avance-900 MHz spectrometers all equipped with cryoprobes. Data were processed with Topspin 3.0 (Bruker) and analyzed with CARA (http://cara.nmr.ch/) or Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Protein sequence–specific backbone and side chain assignments were achieved with unlabeled RNA in complex with ^{13}C-^{15}N–labeled protein to collect 2D ^1H-^{15}N HSQC, 2D ^1H-^{13}C HSQC, 3D HNCA, 3D HNCO, 3D CBCACONH, 3D HNCACB, 3D HBHACONH, 3D HC(C)H TOCSY, 3D (H)CCH TOCSY, and 3D NOESY ^1H-^{13}C HSQC aliphatic experiments[39]. Aromatic proton assignments were performed with 2D ^1H-^1H TOCSY, 2D long-range ^1H-^{15}N HSQC (for histidines) and 3D NOESY ^1H-^{13}C HSQC aromatic experiments. A 3D NOESY ^1H-^{15}N HSQC was collected on a complex of unlabeled RNA with ^{15}N-labeled protein.

Preliminary RNA resonance assignments of unlabeled RNA in complex with ^{15}N-labeled or ^{13}C-^{15}N–labeled protein were performed with 2D ^1H-^1H TOCSY, 2D ^1H-^1H NOESY and 2D ^{13}C 1F-filtered 2F-filtered NOESY[40] experiments in 100% D_2O. Imino protons were assigned on the basis of 2D ^1H-^{15}N HSQC and 2D HNCO experiments[41]. Ribose and aromatic proton assignments for RNA were achieved with ^{13}C-^{15}N-labeled RNA in complex with ^{15}N-labeled protein to collect 2D ^1H-^{13}C HSQC, 3D HC(C)H TOCSY, 3D HC(C)H COSY, 3D NOESY ^1H-^{13}C HSQC ribose, and 3D NOESY ^1H-^{13}C HSQC aromatic experiments optimized for RNA[42].

Intermolecular NOEs were obtained with 2D ^1H-^1H NOESY, 2D ^1H-^1H, ^{13}C 2F-filtered NOESY and 3D ^{13}C 1F-filtered 2F-edited HSQC-NOESY[43] with unlabeled RNA and ^{15}N-labeled or ^{15}N-^{13}C–labeled protein. Unambiguous assignments of intermolecular NOEs from protein side chains to RNA ribose and aromatic protons were also obtained from 3D NOESY ^1H-^{13}C HSQC ribose, and 3D NOESY ^1H-^{13}C HSQC aromatic experiments. Intermolecular NOEs between imino protons of the RNA and protein protons were assigned with a 3D ^1H-^{13}C HSQC NOESY and 2D ^1H-^1H NOESY at 278 K in H_2O.

All NOESY spectra were recorded with a mixing time of 150 ms, the 3D (H)CCH and HC(C)H TOCSY spectra with a mixing time of 23 ms and the 2D ^1H-^1H TOCSY spectra with mixing times of 50 ms.

Protein sequence–specific backbone assignments of unbound TDP-43 RBD were achieved with ^{13}C-^{15}N–labeled protein in gel-filtration buffer at 0.5 mM concentration by collection of 2D ^1H-^{15}N HSQC, 2D ^1H-^{13}C HSQC, 3D HNCO, 3D CBCACONH and 3D HNCACB experiments.

Relaxation experiments were performed on a Bruker AVIII-500MHz equipped with a cryoprobe, with a ^{13}C-^{15}N–labeled TDP-43 RBD sample (described above) at a concentration of 0.5 mM and ^{15}N-labeled TDP-43 RBD in complex with unlabeled RNA at 0.5 mM in NMR buffer. The NMR experiments were performed once. The ^{15}N T1 values were determined from eight ^1H-^{15}N spectra recorded as a pseudo-3D experiment with different delays: 80, 160, 240, 360, 480, 600, 900 and 1,200 ms and an interscan delay of 1 s. The ^{15}N T2 values were determined from nine ^1H-^{15}N correlation spectra recorded with different relaxation delays: 6.14, 18.42, 24.56, 36.83, 49.11, 61.39, 73.67, 92.08 and 104.36 ms and an interscan delay of 1 s. T1 and T2 values were extracted with a curve-fitting routine in Sparky (http://www.cgl.ucsf.edu/home/sparky/) for well-resolved amide resonances whose amide groups are rigid. Overall correlation times (τ_c) were derived from the T1/T2 ratio and calculated with the assumption of overall isotropic motion according to ref. 44. The average values and standard deviations are reported in **Supplementary Table 2**.

Structure calculation and refinement. AtnosCandid[45,46] integrated into the Unio package was used to generate preliminary structures and a list of automatically assigned NOE distance constraints for the TDP-43 RBD in complex with RNA. Peak peaking and NOE assignments were performed on 3D NOESY (^{15}N- and ^{13}C-edited) spectra. Hydrogen-bond constraints were based on hydrogen-deuterium exchange experiments of the amide protons. Constraints were introduced for backbone amide protons whose signal was visible in ^1H-^{15}N HSQC spectra after lyophilization of the complex and dissolution in D_2O. Oxygen acceptors were identified on the basis of preliminary structures calculated without hydrogen-bond constraints. Seven AtnosCandid iterations were performed, and 100 independent structures were calculated at each iteration step. Structures of the protein–RNA complexes were calculated with CYANA[45] by addition of the manually assigned intramolecular RNA and intermolecular RNA-protein distance restraints to the intramolecular protein constraints derived from AtnosCandid. For each CYANA run, 100 independent structures were calculated. The 50 lowest-energy structures were refined with the SANDER module of AMBER 7.0 with a simulated annealing protocol[47]. The 20 best structures on the basis of energy and restraint violation energies were analyzed with PROCHECK[48]. The Ramachandran plot of the TDP-43–AUG12 RNA complex indicates that 74.7% of the residues are in the most favored region, 23.4% in the additionally allowed regions, 1.6% in the generously allowed regions and 0.3% in the disallowed region.

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 with optical absorbance at 280 and 260 nm, respectively. Wild-type and mutant TDP-43 RBD proteins (5–20 μM) were titrated with 50–250 μM of RNA by 50 injections of 5 μl every 3 min at 25 °C. The ITC experiments were performed in technical duplicates. Raw data were integrated, normalized for the molar concentration and analyzed according to a 1:1 RNA/protein-ratio binding model. The fits were accomplished with the nonlinear Levenberg–Marquardt method with Origin 7.0. The values from **Table 2** are from one representative experiment and are reported with fitting error.

Coupled *in vivo* splicing and siRNA assay. The coupled *in vivo* splicing and siRNA assays were performed as previously described with add-back of siRNA-resistant wild-type or mutant full-length TDP-43 protein[2,21,49]. The procedure described in D'Ambrogio *et al.*[21] was used to compare the ability of the various siRNA-resistant TDP-43 single and double mutants to rescue the inhibition on *CFTR* exon 9 alternative splicing. Briefly, HeLa cells were plated at 30% of confluence (day 0), and two rounds of TDP-43 siRNA transfections were carried out, according to the procedure already described[49], on days 1 and 2 in order to maximize TDP-43 silencing efficiency. Transfection of 0.5 μg of the *CFTR* 155T reporter minigene[49] together with 1 μg of pFLAG-expressed proteins was performed in the afternoon on day 2. Cells were harvested on day 3, 18 h after the transfection, and total RNA was collected with EuroGold TRifast (Euroclone). Reverse transcription was performed with M-MLV Reverse Transcriptase (Invitrogen), according to the manufacturer's protocol. PCR with DNA polymerase (New England Biolabs) was carried out for 35 amplification cycles (95 °C for 45 s, 54 °C for 45 s, 72 °C for 45 s). PCR products were analyzed on 1.5% agarose gels. Expression levels of endogenous and added-back TDP-43 proteins were monitored through western blotting with a commercially available mouse monoclonal FLAG antibody (Sigma, F1804), and with in house–made mouse polyclonal antibodies against human TDP-43 and tubulin (used as a loading control). Both antibodies were used at a 1:1,000 dilution. Specificity information can be found in ref. 49. The splicing assays were done in biological triplicates (three individual transfection experiments). The band intensities were evaluated with ImageJ. The average values with standard deviation are reported (**Fig. 4** and **Supplementary Table 3**). Original gel images are shown in **Supplementary Figure 5**.

35. Lukavsky, P.J. & Puglisi, J.D. Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* **10**, 889–893 (2004).
36. Easton, L.E., Shibata, Y. & Lukavsky, P.J. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **16**, 647–653 (2010).
37. Duss, O., Maris, C., von Schroetter, C. & Allain, F.H. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res.* **38**, e188 (2010).
38. Duss, O., Lukavsky, P.J. & Allain, F.H. Isotope labeling and segmental labeling of larger RNAs for NMR structural studies. *Adv. Exp. Med. Biol.* **992**, 121–144 (2012).
39. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
40. 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).
41. Fürtig, B., Richter, C., Wohnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **4**, 936–962 (2003).
42. Lukavsky, P.J. & Puglisi, J.D. RNAPack: an integrated NMR approach to RNA structure determination. *Methods* **25**, 316–332 (2001).
43. 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).
44. Gryk, M.R., Abseher, R., Simon, B., Nilges, M. & Oschkinat, H. Heteronuclear relaxation study of the PH domain of β-spectrin: restriction of loop motions upon binding inositol trisphosphate. *J. Mol. Biol.* **280**, 879–896 (1998).
45. 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).
46. 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).
47. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
48. 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).
49. Ayala, Y.M., Pagani, F. & Baralle, F.E. TDP43 depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett.* **580**, 1339–1344 (2006).