5062–5074 *Nucleic Acids Research, 2013, Vol. 41, No. 9* *Published online 21 March 2013*
doi:10.1093/nar/gkt189

Characterizing TDP-43 interaction with its RNA targets

Amit Bhardwaj, Michael P. Myers, Emanuele Buratti and Francisco E. Baralle*

International Centre for Genetic Engineering and Biotechnology (ICGEB), 34012 Trieste, Italy

Received October 11, 2012; Revised February 26, 2013; Accepted February 27, 2013

ABSTRACT

One of the most important functional features of nuclear factor TDP-43 is its ability to bind UG-repeats with high efficiency. Several cross-linking and immunoprecipitation (CLIP) and RNA immunoprecipitation-sequencing (RIP-seq) analyses have indicated that TDP-43 *in vivo* can also specifically bind loosely conserved UG/GU-rich repeats interspersed by other nucleotides. These sequences are predominantly localized within long introns and in the 3′UTR of various genes. Most importantly, some of these sequences have been found to exist in the 3′UTR region of TDP-43 itself. In the TDP-43 3′UTR context, the presence of these UG-like sequences is essential for TDP-43 to autoregulate its own levels through a negative feedback loop. In this work, we have compared the binding of TDP-43 with these types of sequences as opposed to perfect UG-stretches. We show that the binding affinity to the UG-like sequences has a dissociation constant (K_d) of ~110 nM compared with a K_d of 8 nM for straight UGs, and have mapped the region of contact between protein and RNA. In addition, our results indicate that the local concentration of UG dinucleotides in the CLIP sequences is one of the major factors influencing the interaction of these RNA sequences with TDP-43.

INTRODUCTION

43 kDa trans-active response (TAR) DNA-binding protein (TDP-43) belongs to the family of heterogeneous nuclear ribonucleoproteins and has been reported to be involved in different aspects of RNA processing such as alternative splicing, RNA stability and transcriptional regulation (1). In a range of neurodegenerative diseases such as frontotemporal lobar degeneration (FTLD) and Amyotrophic lateral sclerosis (ALS), TDP-43 is relocated from nucleus to the cytoplasm and sequestered in cytoplasmic or occasionally nuclear inclusions in which the protein is ubiquitinated, hyperphosphorylated and partially processed to yield C-terminal fragments (2,3). Mutations in the highly conserved region of the gene that encodes TDP-43 (TARDBP) have been reported in sporadic (1.5%), familial (4%) ALS and in rare cases of FTLD as well (4,5). However, over 90% of ALS and FTLD patients do not possess genetic alterations in the TARDBP locus. This is unlike the coding mutations reported in the SOD1 protein, which occurs in 20% of all familial ALS patients (6) or in the prion protein (PrP) gene in case of the inherited form (15%) of human prion diseases (7).

TDP-43 has two RNA Recognition Motifs (RRM) domains also known as RNA binding domains (RBDs). The RRM domain is the most abundant RBD and is present in ~0.5–1% of human genes (8) and is the most extensively studied RBD both in terms of structure and biochemistry (9). Usually, RRM domains contain two octameric and hexameric consensus sequences known as Ribonucleoprotein 1 (RNP1) and Ribonucleoprotein 2 (RNP2), respectively. Most commonly, three aromatic side chain residues located in the conserved RNP1 ([RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY], in β_3-strand) and RNP2 ([ILV]-[FY]-[LV]-X-N-L, in β_1-strand) sequence stretches accommodate 2 nt (10). However, deviations from this basic mode of binding are found in many RRM–RNA complexes owing to various structural features. As far as TDP-43 is concerned, it has been reported that TDP-43 binds to UG-repeats with high affinity and RRM1 is necessary and sufficient for UG binding (11). More recently, it has been shown that RRM2 also binds to nucleic acids but with reduced magnitude (12). On a global scale, attempts have recently been made to identify the RNA targets bound by TDP-43 *in vivo* using various neuronal cell lines (13–15) as well in healthy and diseased human brain tissues (16) and in mouse brains (17). These approaches have mostly used RNA immunoprecipitation-sequencing (RIP-seq) or cross-linking and immunoprecipitation (CLIP) approaches. Although these results still need to be analyzed more critically, as there is surprisingly little overlap (>10%) between some of these studies (15,18), the overall picture indicates that TDP-43 is associated

*To whom correspondence should be addressed. Tel: +39 040 3757337; Fax: +39 0403757361; Email: baralle@icgeb.org

with thousands of RNA sequences, which are mostly localized in long intronic regions of genes, their 3′ untranslated regions (3′-UTRs), as well as in several non-coding RNAs. Most of these binder sequences contain either canonical or slightly distorted UG-repeats of various lengths. However, there are also many sequences that do not contain any UG-repeats, but were still found to bind specifically to TDP-43.

The functional significance of these interactions between TDP-43 and non-UG-repeat containing RNAs as compared with UG-repeat containing RNAs is currently poorly understood. The only exception to this is represented by the presence of several CLIP-derived sequences in the 3′UTR region of TDP-43 (listed in Figure 1) called TDPBR for TDP-43 Binding Region (19). Binding of TDP-43 to the TDPBR has been found to be important for the autoregulation process of this protein (19). A more in-depth analysis of this autoregulation process established that excess TDP-43 binds to this region resulting in Polymerase II stalling, which leads either to premature transcript termination or increased removal of a normally silent intron (intron 7). Most importantly, intron 7 excision has the effect of removing the most efficient polyadenylation site of the TDP-43 gene (pA_1), forcing the polyadenylation machinery to use several alternative polyA sites that are localized downstream of pA_1. The use of these polyA sites gives rise to transcripts that are mostly nuclear retained and are thus unable to sustain efficient translation of the protein (20).

In this work, we investigated the TDP-43–RNA binding process and, in particular, how the different CLIP sequences present in the TDPBR recognize and interact with TDP-43. In fact, in most cases, protein and RNA recognize each other by an 'induced fit' causing conformational changes in the protein (21), the RNA (22) and sometimes both (23,24). Loops connecting β-strands and α-helices can also be crucial for nucleic acid recognition by providing specificity and can also undergo conformational changes on RNA binding (25). Our analysis provides direct evidence suggesting that the local concentration of UG dinucleotides in the CLIP sequences and changes in the RNA conformation (but not protein) represent the major factors influencing their binding with TDP-43 in solution. In addition, our data demonstrated that the binding affinities are 10-fold lower for TDPBR sequences than for perfect UG-repeat sequences. This difference may have a physiological significance in view of the need of multiple low affinity sequences for the proper functioning of the self-regulation process (20).

MATERIALS AND METHODS

Protein expression and purification

Expression and purification of pGEX-3X GST-TDP (101–261) was performed as previously described (26). A His-tagged TDP (101–261) was produced by a fragment encoding RRM1, and RRM2 of human TDP-43 (residues 101–261) was amplified from pGEX-3X TDP-43 using HisTDPFwd (5′-cgcggatcccagaaaacatccgatttaata-3′) and HisTDPRev (5′-cccaagcttttattcggcattggatatatgaac-3′) primers, which incorporate BamHI and HindIII restriction sites, respectively. The PCR-amplified DNA fragment was cloned into complimentary sites of expression vector pQE30 (Qiagen) to generate the N-terminal His-tagged construct pQE30HisTDP. This construct was transformed into *Escherichia coli* M15 strain and protein expression induced with 1 mM IPTG at 37°C for 4 h. Protein concentration was determined using the Bradford method. The induced cell pellet was resuspended in 1× phosphate buffered saline buffer containing protease inhibitor (Roche) and was sonicated. The supernatant fraction was loaded onto a manually packed Ni-NTA column and purified according to manufacturer's instructions. The purified protein fractions were pooled and passed through vivaspin ultracentrifugation unit (10 kDa MWC, Sartorius, Germany).

End labelling of synthetic RNA oligonucleotides:

RNA oligonucleotides were purchased from Sigma Aldrich, and 1 μM of each oligonucleotide was used for labelling using T4 polynucleotide kinase (New England Biolabs) and γ^{32}P ATP (PerkinElmer Life Science) using standard protocols.

TDP–RNA binding assay

The interaction of different RNA oligonucleotides with GST-TDP (101–261) was assessed by Electromobility shift assay (EMSA). For qualitative binding analysis, 125 nM of GST-TDP (101–261) was mixed with 50 nM of various γ^{32}P ATP labelled RNA oligonucleotides. Binding reactions (20 μl) were incubated in 1X binding buffer [10 mM NaCl, 10 mM Tris (pH 8.0), 2 mM MgCl2, 5% (v/v) glycerol, 1 mM DTT] at 25°C for 20 min before loading onto a 1.5 mm thick non-denaturing polyacrylamide gel (6%) in 0.5× TBE. Similar experiments were performed where the binding reactions were cross-linked for 2 min using UV (Stratalinker 1800; Stratagene) before loading onto the gel. Gels were run at 100–120 V for 2 h at 4°C. Gels were dried, exposed to autoradiographic film and a cyclone phosphorimager screen was used to allow for accurate quantification.

Determination of dissociation constant

The binding affinities of TDP-43 for CLIP34nt, CLIP6 and CLIP34_UG6 RNA oligonucleotides were determined using quantitative EMSA analysis as described previously (27). Briefly, RNA oligonucleotide concentration was varied while keeping the GST-TDP (101–261) protein concentration constant. The dissociation constant (K_d) was determined using a protein concentration of 25 nM for $(UG)_6$ and Clip34nt_UG6, whereas 125 nM of protein was used for the rest of the Clip sequences. The K_d was also determined using GST-TDP_{mut2} (101–261), which contains Phe229Leu and Phe231Leu mutation in the RRM2 domain. Quantification of protein–RNA complex was performed with OptiQuant image analysis software (Packard Instrument Co.). The binding curves were plotted and analyzed using Prism (GraphPad software).

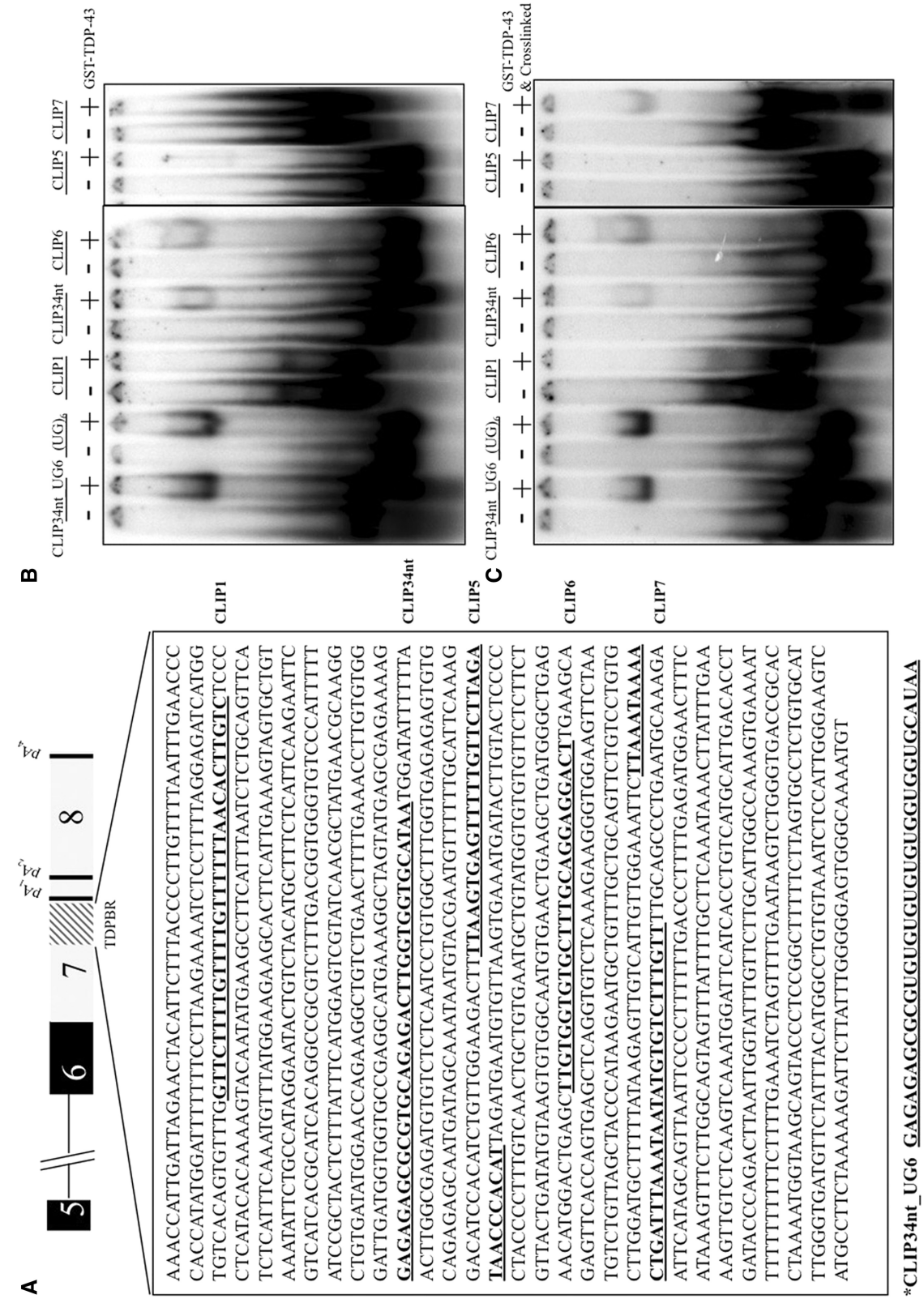

Figure 1. (**A**) Illustrates a schematic representation of TDP-43 gene showing the coding exons (black boxes), untranslated region (grey box), TDPBR (shaded box) and polyadenylation sites (black lines). The CLIP sequences analyzed in this study are shown in bold and underlined. Sequence of mutated CLIP34nt (CLIP34nt_UG6) is also shown below. (**B**) Shows an EMSA binding analysis of CLIP sequences: γ-P^{32} ATP labelled RNA oligonucleotides of CLIP sequences were mixed with GST tagged TDP-43 (101–261) and subjected to EMSA analysis. (**C**) Shows the binding profile of (**B**) following UV-cross-linking.

Lysine acetylation of His-TDP and His-TDP–RNA complex

Acetylation of lysine residues of His-TDP (101–261) was carried out under native conditions in the presence or absence of synthetic RNA oligos Which includes $(UG)_6$, CLIP34nt and CLIP6. His-TDP and various TDP–RNA complexes were prepared in 0.5X binding buffer [10 mM NaCl, 10 mM Tris(pH 8.0), 2 mM MgCl2, 5% (v/v) glycerol, 1 mM DTT] + 100 mM TEAB, pH8.5. In the case of the His-TDP–RNA complex, protein to RNA ratio was adjusted to 1:5. His-TDP (101–261) was labelled with 5 mM acetic anhydride and His-TDP–RNA complex was labelled with 5 mM hexa-deuteroacetic anhydride at 25°C for 1 h, and the reaction was stopped using 100 mM diammonium phosphate, pH8.0. Labelled His-TDP (101–261) and His-TDP–RNA complexes were mixed and subjected to SDS-PAGE. Gels were stained using Coommasie brilliant blue.

In-gel digestion and mass spectrometry analysis

Acetic anhydride labelled protein samples were subjected to in-gel digestion with chymotrypsin (Roche) at 25°C for 12 h. Extracted peptides were analyzed with electrospray mass spectrometry (Amazon ETD, Bruker) coupled with Easy NLC II (Bruker).

Dried samples were resuspended in 0.2% Acetic Acid in water and analyzed by LC-MS/MS. The chromatography was developed with a discontinous gradient of MeOH from 0 to 65% in 60 or 120 min using a 10×0.075 mm column packed with Halo RPamide resin. The column terminated in a pulled electrospray tip, which was mounted with a custom built source (Phoenix S & T) onto a Bruker Amazon ETD mass spectrometer. Electrospray was performed with an applied voltage of 2800–3000 V and nebulizer gas set at 5–6 psi. Alternating CID and ETD MS/MS spectra were collected using standard methods.

The raw MS/MS data was converted to mgf, mzxml and xml formats using Data Analysis (v4.0 SP5, Bruker). During this conversion, the spectra are also deconvoluted, and spectra that did not contain peaks separated by amino acid masses were excluded from further analysis. The resulting files were analyzed using the GPM (version 2.2.1) and MASCOT search engine (version 2.4).

Circular dichroism (CD)

The conformation of GST-TDP (101–261), RNA and GST-TDP–RNA complex was evaluated using a JASCO J-810 spectropolarimeter. All the samples were prepared in $0.5 \times$ binding buffer [10 mM NaCl, 10 mM Tris (pH 8.0), 2 mM MgCl2, 5% (v/v) glycerol, 1 mM DTT]. The CD spectrum was recorded from 300 to 198 nm using a cylindrical quartz cell with a path length of 0.5 mm. Firstly, CD spectra of protein, RNA and protein–RNA complex were recorded at a fixed concentration of protein and RNA oligonucleotides. CD spectrum of protein–RNA complex was then subtracted with the respective protein and RNA alone spectra to determine the binding induced conformational changes in RNA and protein, respectively. On the other hand, in the titration experiments, the RNA concentration was kept constant in each reaction and incubated with increasing concentrations of GST-TDP (101–261) protein. The Clip34nt_UG6 RNA was also titrated against various concentrations of GST-TDP_{mut1} (101–261), which contains the Phe147Leu and Phe149Leu mutation in RRM1 domain, and GST-TDP_{mut2} (101–261) protein, which contains the Phe229Leu and Phe231Leu mutation in the RRM2 domain. Individual RNA and protein spectra were also recorded for each concentration and used as a control. CD spectrum for each protein–RNA complex was subtracted from the control protein spectra of the same concentration and then compared with that of control RNA spectra to determine the conformational changes in RNA induced by protein binding. In all the experiments, the spectrum was collected with a scan rate of 20 nm min^{-1} and a width of 1 nm. Scans were repeated three times and then averaged.

RESULTS

Binding affinity of TDP-43 to various CLIP RNAs

One of the key events that control TDP-43 autoregulation is its interaction *in vivo* with the TDPBR, an ~600-nt long region, which lies within the 3′UTR (19,20). In our previous studies, we identified a 34 nt sequence, named as CLIP34nt, within the TDPBR region as a specific binding site for TDP-43 (19). However, band shift analyses indcated that other sequences within this region were also binding to TDP-43. This observation has functional significance, as autoregulation in a suitable 3′UTR reporter construct was abolished only when the entire TDPBR was removed (19). In keeping with this observation, the recently published iCLIP analysis performed in human brain and cell lines predicted the existence of additional CLIP sequences originated from the TDPBR (16). Based on preliminary sequence analysis according to their length and nucleotide composition, see below, we selected five of these CLIP sequences for further study (see Figure 1A for a schematic diagram of their position and sequence within the TDPBR).

During qualitative binding analysis, we found that only CLIP34nt (19) and CLIP6 RNA showed strong binding to GST-TDP (101–261) with and without UV cross-linking (Figure 1B and C). On the other hand, CLIP7 RNA showed the binding only after UV cross-linking (Figure 1C), while CLIP1 and CLIP5 did not show any visible EMSA shift in either condition (Figure 1B and C). This could be owing to the fact that *in vivo* binding of TDP-43 to these other sequences (CLIP1 and 5) may require the presence of additional factors that are not present in the *in vitro* reaction. To explore the interplay between UG and CLIP motifs, we also designed a mutant called CLIP34nt_UG6 (Figure 1A, bottom). In this mutant, the non-canonical UG-repeats of CLIP34nt were converted into a perfect $(UG)_6$-repeat by replacing the least number of nucleotides (Figure 1A). The CLIP34nt_UG6 was found to interact with the same intensity of $(UG)_6$ RNA (Figure 1B and C), indicating

towards the central role played by UG-repeats in TDP–RNA interaction. On the other hand, the GST-TDP_{mut1} (101–261) did not show any visible EMSA shift with any of the RNA oligonucleotide used in this study (Supplementary Figure S1).

Subsequently, the K_d values for CLIP34nt, CLIP6 and CLIP34nt_UG6 were calculated using GST-TDP (101–261) using quantitative EMSA (27) (Figure 2, Table 1). For this experiment, we also used a GST-TDP_{mut2} (101–261), which was previously shown to bind to UG-repeated sequences in EMSA analysis (11). This mutant was chosen to gain additional quantitative insight in the role played by RRM2 Phenylalanines 229 and 231 in the binding of TDP-43 to CLIPs and UG_6. Using this mutant, we observed that all K_d values for all the CLIP sequences were comparable with those observed for GST-TDP (101–261). This result suggested that, unlike Phenylalanines 147 and 149 in RRM1 (11), the corresponding residues of RRM2 do not play a major role in the recognition and binding of TDP-43 with any of these sequences (Table 1).

Finally, it was interesting to note that with GST-TDP (101–261) the mutated CLIP34nt (CLIP34nt_UG6) displayed a K_d of 2–5 nM, which is comparable with that of $(UG)_6$ RNA (Table 2) but 10 times stronger than the CLIP34nt (Table 1), thus confirming the central role played by UG-repeats in the TDP–RNA interaction and in keeping with the data obtained in the EMSA and cross-linking experiments (Figure 1B, C and 2A–D).

Overall, these results are consistent with previous data that showed TDP-43 is capable of interacting with high affinity to single-stranded UG-repeats, with binding strength strongly increasing with the length of these repeats (11,12,27). Moreover, it was interesting to note that the K_d values of CLIP34nt and CLIP6 with distorted $(UG/GU)_5$-repeats are comparable with that of $(UG)_4$-repeats (12,27). Similarly, the K_d values for CLIP1, CLIP5 and CLIP7 containing $(UG)_1$, $(UG)_1$ and $(UG)_2$ as a longest UG-repeats, respectively, could not be determined and followed the same trend like $(UG)_1$- and $(UG)_2$-repeats (Table 2).

Analysis of the various TDP–RNA complexes using lysine acetylation and mass spectrometry

It has previously been shown that Phe147 and Phe149 in RRM1 of TDP-43 plays a major role in interacting with nucleic acids (11). Because Lys145 lies in close proximity to these two active Phe residues, we were able to take advantage of a lysine labelling strategy to determine whether any of the target RNA molecules (UG-repeats and CLIP RNAs) are binding this site or use different faces of TDP-43 to attain stable binding. Importantly, lysine labelling has several advantages, as it has been shown that the reactivity of the ε-amino group of lysine residues is directly proportional to their surface accessibility (28,29) and their reactivity can be influenced by factors that change their surface accessibility or alter their pK_a (30). Therefore, lysine labelling can provide increased resolution and the ability to pinpoint subtle changes in accessibility as a function of RNA binding.

In our experiments, we found that TDP-43 lysine labelling with acetic anhydride to be robust and gentle and did not interfere with RNA gel shifting at concentration up to 5 mM (data not shown). In addition, binary comparisons can be performed via mass spectrometry using different stable isotope versions of acetic anhydride. The schematic representation of the work flow is shown in Supplementary Figure S2. Briefly, free His-TDP (101–261) was treated with 5 mM acetic anhydride and His-TDP (101–261) in the presence of RNA was treated with equimolar concentration of hexadeuteroacetic anhydride (heavy). The two reactions were mixed, separated by SDS-PAGE and subjected to in-gel digestion with chymotrypsin, which resulted in a mixture of acetylated and heavy acetyl derivative of the same peptides. These peptides were then separated and analyzed by LC-MS/MS. Acetylated and heavy acetylated peptides were identified based on their mass shifts of 42 and 45 Da per lysine residue, respectively. In addition, unacetylated lysine residues were also found indicating that these lysines are buried in the structure and have poor solvent accessibility in the test conditions. To maintain the homogeneity in the text, we used the amino acid numbering as per the full-length wild-type TDP-43. We were particularly interested in the Chym132-147 peptide (^{132}MVQVKKDLKTGHSKGF^{147}), which contains part of the RRM1 and includes the Phe147 residue required for RNA binding. For our purposes, Chym132-147 was of considerable importance as it originates by the chymotryptic digestion at Phe147, dissecting the active site residues of RRM1, and hence can provide information about RNA recognition and binding. In the case of free His-TDP (101–261), we observed that all the four lysine residues were labelled uniformly with acetic anhydride, indicating that in the absence of nucleic acid, these lysines are solvent accessible. However, in case of His-TDP–$(UG)_6$, His-TDP–CLIP34nt and His-TDP–CLIP6 complex, we observed partial labelling of K^{145} (underlined in Figure 3) with d_3-acetic anhydride (Figure 3). Comparative analysis of the light and heavy form of this peptide clearly showed that partial labelling of K^{145} was greatly enhanced by the presence of RNA as compared with free His-TDP (Figure 3). These results demonstrate that the binding to these RNAs protects K^{145} residues and makes it less accessible for acetic anhydride labelling. All the results obtained using this technique concern RRM1. The role of RRM2 in binding to RNA is still unknown although it has been previously claimed that low affinity RNA or DNA binding may occur through the RRM2 (12). This strategy does not provide any appropriate answer about the role of RRM2 domain in RNA binding because the nearest lysine residue is four residues away and lies on a different chymotryptic peptide.

Presence of conformational changes in target RNA on complex formation

The importance of UG-repeats with respect to the interaction with TDP-43 is well documented but the role played by the tertiary structure of various RNA targets has never been studied before. To investigate structural

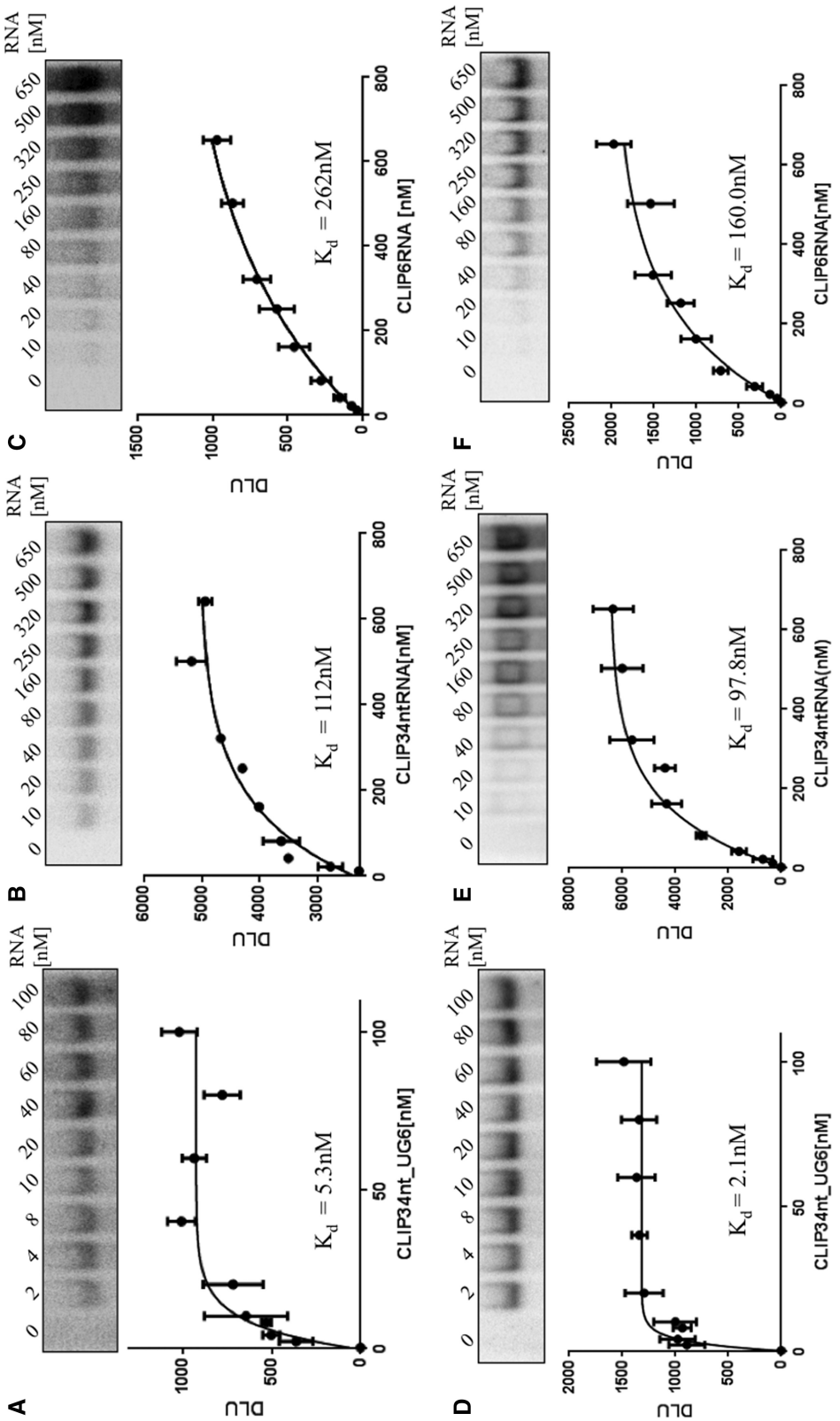

Figure 2. Quantitative EMSA analysis of various TDP–RNA complexes. Panel (**A–C**) shows the gel profiles and binding curves plotted using quantitative EMSA analysis for CLIP34nt_UG6, CLIP34nt and CLIP6 with GST-TDP (101–261). Panel (**D–F**) show the gel profile and binding curves plotted using quantitative EMSA analysis for CLIP34nt_UG6, CLIP34nt and CLIP6 with GST-TDP_{mut2} (101–261 F229L and F231L). The concentration (nM) of each probe used to determine the K_d is mentioned on the top of each gel. Each experiment was repeated at least three times to plot the binding curves.

Table 1. Experimentally determined Kd values of various CLIP sequences in this study

RNA Samples	Dissociation Constant (K_d) [nM]	
	GST-TDP (101–261)	GST-TDP$_{mut2}$ (101–261)
CLIP34nt	112	97.8
CLIP6	262	160
CLIP34nt_UG6	5.3	2.1

Table 2. Experimentally determined Kd values of various UG-repeats

Sequence	Kd value (nM)	Reference
$(UG)_1$	Could not be determined	12
$(UG)_2$	Could not be determined	12
$(UG)_3$	3060 ± 2100	12
$(UG)_4$	115.3 ± 16.69	12
$(UG)_5$	30 ± 3	27
$(UG)_6$	8.0 ± 0.7	27
$(UG)_8$	2.79 ± 0.14	12

changes on TDP–RNA interaction, CD spectra were collected for GST-TDP (101–261), RNAs and GST-TDP–RNA complexes. A typical CD spectrum with a negative peak at 208 nm and a positive peak at 265 nm corresponds to A-form double-helix conformation in RNA (31,32), and a decrease of peak intensities at 208 and 265 nm is an indication of RNA denaturation (33). Far-UV CD spectra of the various RNA oligonucleotides were first recorded at a fixed protein/RNA concentration (Supplementary Figure S3). All the RNA oligonucleotides displayed A-type helix conformation under experimental conditions except $(UG)_6$ RNA. Nonetheless, this experiment clearly showed conformational changes in the case of the $(UG)_6$ and CLIP34nt_UG6 sequences, whereas CLIP34nt, CLIP6 and CLIP1 (used as non-binding control) displayed only subtle changes after their interaction with the protein.

To further explore the conformational changes in RNA due to protein binding, titration experiments were then performed in which a fixed concentration of all the four RNA samples [CLIP34nt, CLIP6, CLIP34nt_UG6 and $(UG)_6$] were titrated against various concentration of GST-TDP (101–261) (Figure 4). In case of $(UG)_6$ RNA, the peaks (at 275 and 200 nm) were found to deviate from the typical A-type helix spectrum and were of less intensity, presumably owing to its weak tendency towards proper base stacking. A clear reduction in the peak intensities (at 200 and 275 nm) was observed for $(UG)_6$ RNA with increasing concentrations of GST-TDP (101–261), and this could be interpreted as a change in RNA conformation due to TDP-43 binding (Figure 4A). Compared with this situation, the subtracted CD spectra of CLIP34nt displayed subtle changes at 265 nm, whereas the subtracted spectra of CLIP6 RNA showed that the negative peak at 210 nm (in unbound state) shifted to 200 nm on binding with protein, but the peak at 268 nm remained unchanged. These subtle changes may be because of the presence of comparatively weak interaction of CLIP34nt and CLIP6 with GST-TDP (101–261) with respect to $(UG)_6$ RNA (Figure 4B and C). Finally, in the case of the CLIP34nt_UG6 RNA sample, it was observed that the intensity at 255 nm decreased with increasing GST-TDP (101–261) concentrations like it was observed with $(UG)_6$ RNA (Figure 4D). The CD titration experiment was also performed with GST-TDP$_{mut2}$ using CLIP34nt_UG6 RNA, and this mutant displayed the same binding affinity as GST-TDP (101–261) to all the target RNA oligos. The substracted spectra displayed the same pattern like the one observed for GST-TDP (101–261), with just some subtle difference at 15 μM protein concentration (Figure 4E).

On the other hand, the subtracted CD spectra of CLIP34nt_UG6 remained almost unchanged when titrated against GST-TDP$_{mut1}$ (101–261) (Figure 4F). This observation clearly shows that the observed change in CLIP34nt_UG6 spectra is specific with respect to its interaction with the protein. These evident conformational changes in CLIP34_UG6 RNA can be attributed to the presence of continuous UG-repeats (≥6). Finally, the far-UV CD spectra of CLIP34nt and CLIP34_UG6 RNA revealed that the position of positive peak shifted from 264 nm (in case of CLIP34nt) to 255 nm in case of the mutant CLIP (Figure 4F).

Lack of conformational changes in TDP-43 protein on complex formation

Regarding the effects of RNA binding on protein conformation, it has been shown that during the formation of most protein–RNA complexes, the protein, the RNA and sometimes both can undergo conformational changes such as with the TAR–TAT complex (34) and U1A–RNA complex (24,34). However, no such study has been carried out with respect to the TDP-43–$(UG)_n$ complex or TDP-43–CLIP complex. Therefore, in parallel to these analyses, we also analyzed the conformational changes in the GST-TDP (101–261) protein itself following RNA interaction using circular dichroism. The far-UV CD spectra of GST-TDP (101–261) alone showed two negative peaks at 208 and 222 nm, characteristic of α-helical structure (Figure 5B). It is evident from the CD analysis performed at fixed protein and RNA concentrations that the resultant subtracted CD spectra of GST-TDP (101–261) did not show any significant conformational changes in the presence of the various RNA oligonucleotides (Figure 5B–D). In fact, considering that both protein and RNA contribute at 208 nm, even the small changes observed around 208 nm in the subtracted CD spectra of GST-TDP cannot be attributed solely to a change in protein conformation alone. Nevertheless, subtle localized changes in GST-TDP (101–261) conformation cannot be ruled out and more advanced studies (i.e. NMR or radiographic crystallography) might be required to investigate this possibility. To confirm that a significant amount of GST-TDP (101–261) was bound to RNA while recording the CD spectrum, we performed the native EMSA gels using the same samples prepared during CD

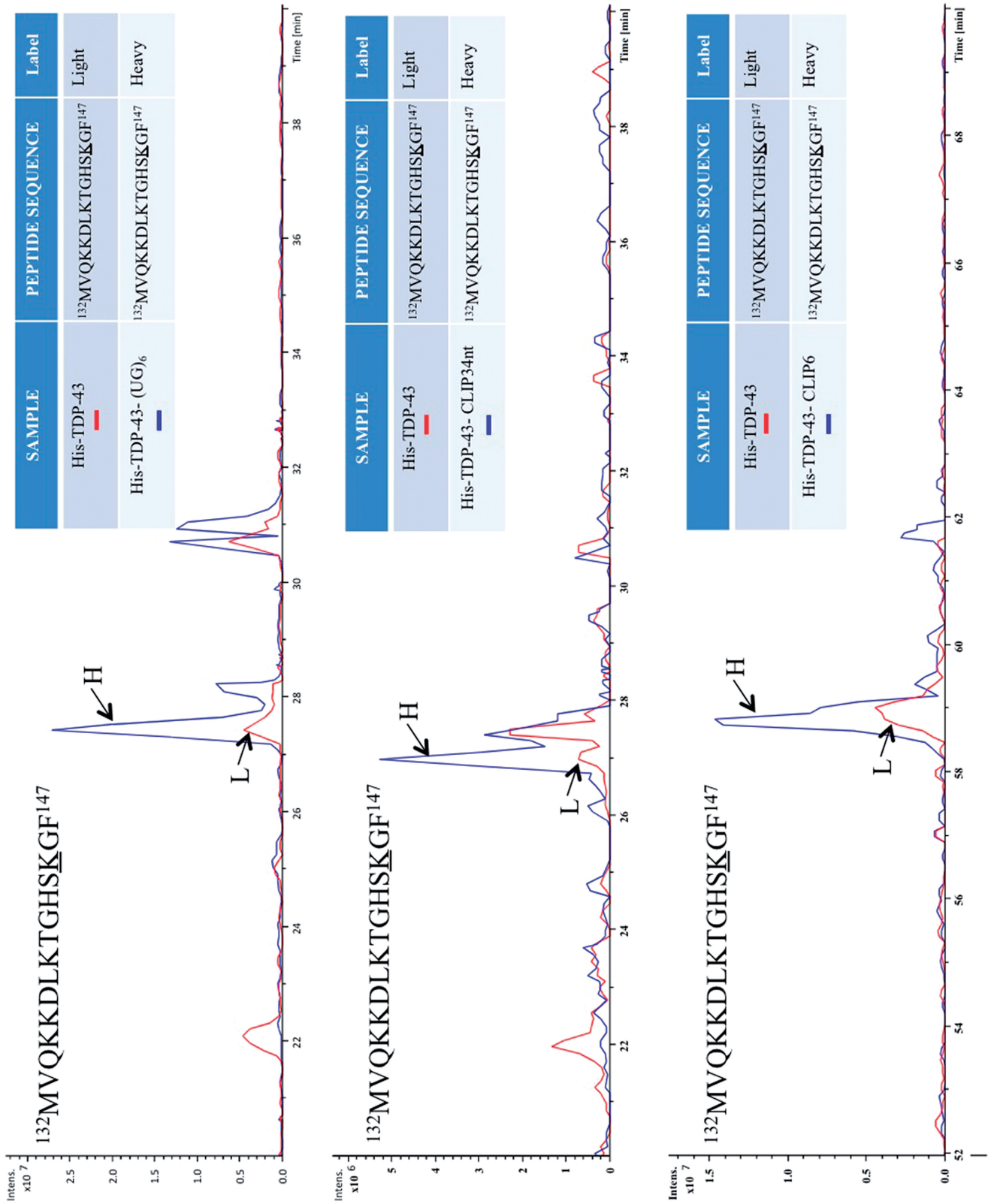

Figure 3. Extracted ion Chromatograms of acetylated His-TDP (101–261). The extracted ion chromatograms for the acetylated chymotryptic peptide ^{132}MVQKKDLKTGHSKGF^{147} (Chym132–147) were generated for samples in the absence and presence of RNA. Importantly, the light acetylation was used to label TDP-43 in the absence of RNA (red traces, Peak 'L'), while the heavy reagent was used to label TDP-43 in the presence of RNA (blue traces, Peak 'H'). For each binary comparison, the samples were mixed before chymotryptic digestion and analyzed in the same LC-MS/MS run. The production of the acetylated Chym132–147 peptides from no RNA controls are shown compared with TDP-43 + $(UG)_6$ (**A**), TDP-43 + CLIP34nt (**B**) and TDP-43 + CLIP6 (**C**). The *x*-axis shows the intensity of the eluting peptides and, in all cases, the presence of RNA inhibited the acetylation of Lysine145 (underlined) as indicated by the greater intensity of the blue traces relative to the red traces. *y*-axis shows the elution time of these peptides.

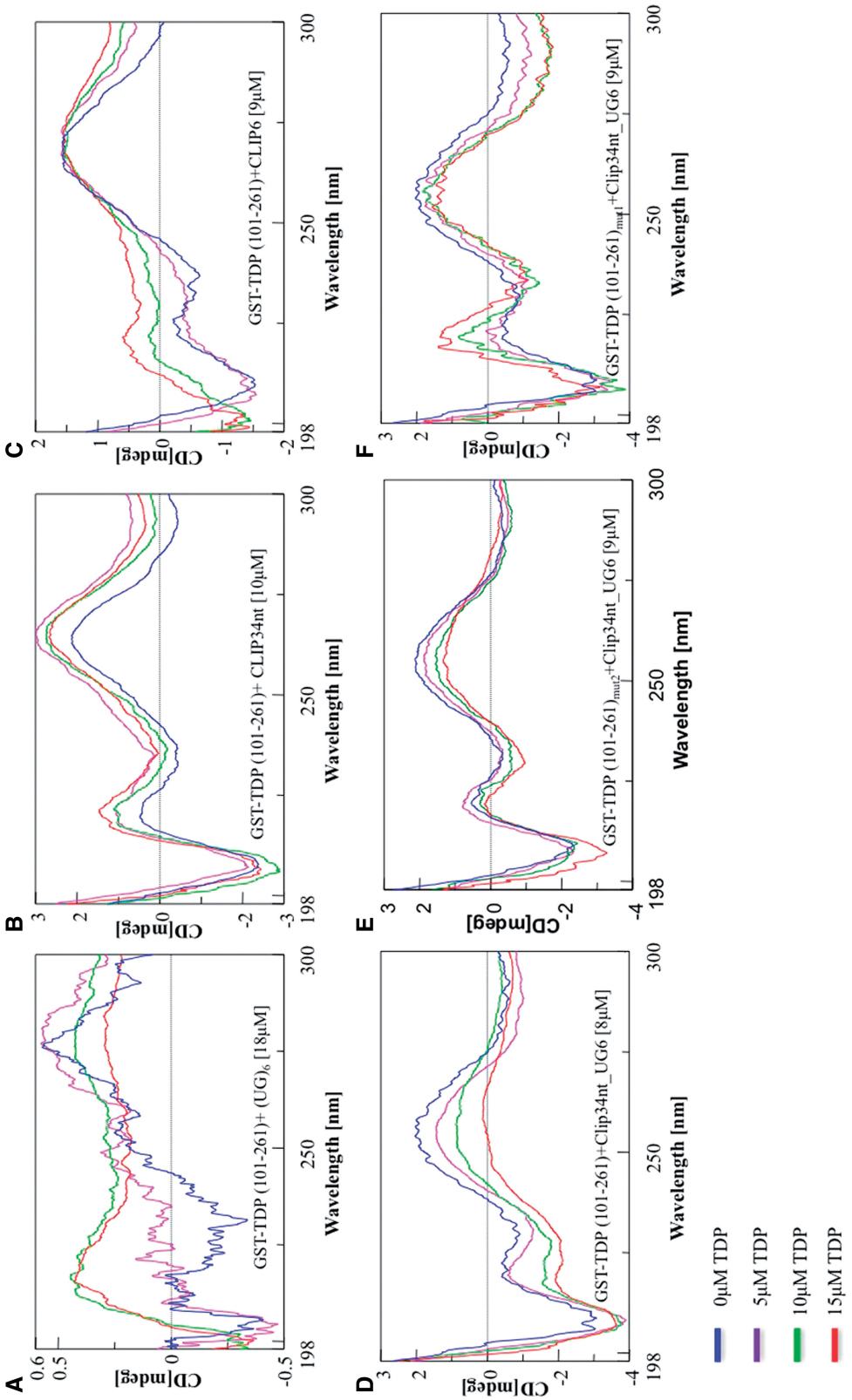

Figure 4. Titration experiments for various TDP–RNA complexes using circular dichroism. Starting from 0 μM (blue), 5 μM (violet), 10 μM (green) and 15 μM (red) of GST-TDP (101–261) were used to perform titration experiments against a fixed concentration of various RNAs. Panels **A**, **B**, **C** and **D** shows the subtracted CD spectra of $(UG)_6$ (18 μM), CLIP34nt (10 μM), CLIP6 (9 μM) and CLIP34nt_UG6 (8 μM) in the presence of various concentration of GST-TDP (101–261), respectively. Panel **E** and **F** show the subtracted CD spectra of CLIP34nt_UG6 (8 μM) in the presence of various concentrations of GST-TDP_{mut2} (101–261) and GST-TDP_{mut1} (101–261), respectively.

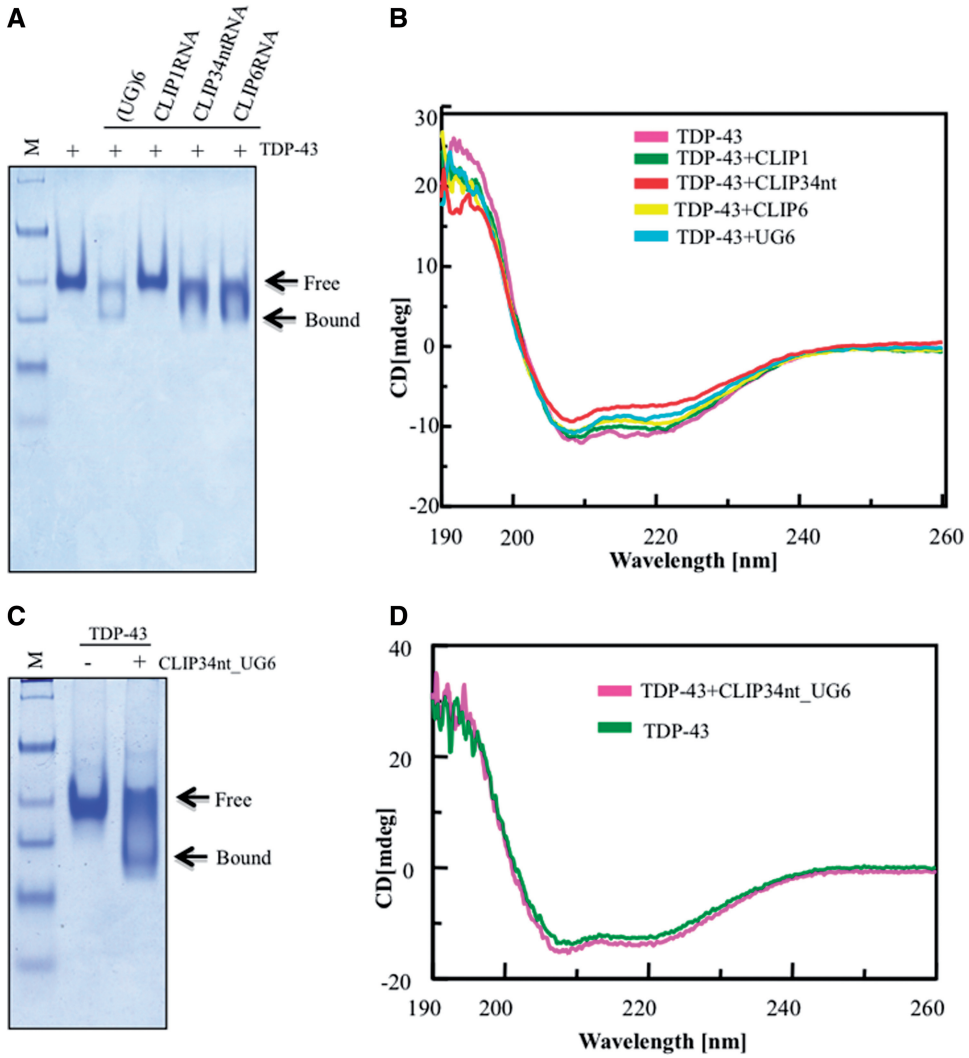

Figure 5. Circular dichroism analysis of GST-TDP (101–261) in the presence of RNA: (**A**) lane 1 of coomassie-stained EMSA gel shows the mobility of free GST-TDP (101–261) (arrow 1). In lane 2, a faster-moving band appeared in the presence of $(UG)_6$ RNA (arrow 2), indicating the formation of a stable protein–RNA complex. Mobility of GST-TDP (101–261) did not change in the presence of CLIP1 RNA (negative control, lane3). Lane (4–5) shows that the addition of CLIP34nt and CLIP6 resulted in the appearance of a smear indicating the formation of weak complex. (**B**) All the samples from A were subjected to circular dichroism, and no significant changes were observed. (**C**) Appearance of a faster-moving band showing the stable complex between GST-TDP (101–261) and CLIP34nt_UG. (**D**) Far-UV CD spectra of GST-TDP (101–261) in the presence of CLIP34nt_UG.

analysis. Figure 5A shows the mobility of free GST-TDP (101–261) under coomassie-stained native EMSA gel (lane 2, marked as 'free'). In Figure 5A, lane 3, a faster-moving band appeared in the presence of $(UG)_6$ RNA (marked as 'bound'), indicating the formation of a stable protein–RNA complex. As expected, mobility of TDP-43 did not change in the presence of CLIP1 RNA (non-binding control, lane 4), and the addition of CLIP34nt and CLIP6 resulted in the appearance of a smear, which is often indicative of the formation of a weak complex (Figure 5A, lane 5–6). The appearance of a well-defined faster-moving band was also observed using this technique for GST-TDP-CLIP34nt_UG (Figure 5C, lane 2–3).

This data indicates that a significant amount of protein was bound to the RNA while performing the CD spectra. It follows that binding to RNA has negligible effect on TDP-43 conformation, which is the complete opposite situation compared with multi-domain splicing factor U2AF65 that also contains two RRM domains in two distinct 3D arrangements in the absence and presence of a high-affinity RNA ligand (35). Furthermore, the lack of labelling of K^{145} when the protein is bound to the RNA indicate that the interaction is blocking access to this lysine residue rather than the RNA inducing a conformational change in TDP-43, which buries K^{145} in the protein.

DISCUSSION

Protein–RNA complexes are involved in regulating many cellular and viral processes (36,37). In this study, we have used EMSA, circular dichroism and mass spectrometry to

understand the structural aspects of the interaction of TDP-43 with UG-repeated sequences and non-UG putative binding sites from the TDPBR region of its own RNA. In particular, we focused on five CLIP sequences from the 3′UTR region of TDP-43 pre-RNA owing to the key role they play in TDP-43 autoregulation (19,20).

Regarding the two positive hits analyzed in this study, CLIP34nt and CLIP6 were found to contain partial UG-repeats, UGGUGGUG and UGUGGUGUG as the longest (UG/GU)-repeats, respectively. Interestingly, these non-canonical (UG/GU)-repeats in CLIP34nt and CLIP6 can be arranged in several possible ways to form UG or GU sequences and at least four UG-repeats can be found. In fact, CLIP34nt and CLIP6 produce a K_d value equivalent to a canonical $(UG)_4$-repeat. To further clarify this, we mutated the original CLIP34nt by introducing a perfect $(UG)_6$-repeat in the middle of the sequence. This mutated CLIP34nt sequence (CLIP34_UG6) bound to GST-TDP (101–261) with similar affinity to that of $(UG)_6$ both qualitatively and quantitatively. The K_d values for all the three RNA were also determined using GST-TDP_{mut2} and were found to be comparable with those obtained using wild-type GST-TDP (101–261). These results suggest that the phenylalanines in the RRM2 domain do not play an active role in recognizing and binding with these sequences. Furthermore, K_d values could not be determined for CLIP1, CLIP5 and CLIP7 RNA, suggesting that the UG dinucleotides play a central role in TDP-43 recognition and point towards a major role played by RRM1 domain in binding with these sequences. During most of the protein–RNA complex formation, the protein, the RNA and sometimes both undergo conformational changes (24,34,38). However, there is no experimental evidence supporting any kind conformational rearrangement in case of the TDP–RNA complex. The Far-UV CD spectra of $(UG)_6$ and CLIP34_UG6 shows that these RNA molecules undergo a significant conformational change in the presence of GST-TDP (101–261) but remained unchanged in the presence of GST-TDP_{mut1} and provide valuable insights about the binding of UG-repeats and UG-repeat-containing sequences with TDP-43. On the other hand, CD analysis showed that TDP-43 does not undergo any major conformational changes in the presence of RNA, which is in complete contrast to the situation in the PrP where host-encoded stimulatory RNA molecules have been shown to have a role in stoichiometric transformation of PrP^c to PrPres (39).

On the basis of the present results and on previous data, which suggested that Phe147 and Phe 149 make aromatic stacking interaction with uracil (U) and guanine (G), respectively (27), we are proposing a binding model for the TDP–RNA complex (Figure 6). According to this model, as TDP-43 gets near to an RNA, its RRM1 domain binds and falls off from UG nucleotides present in the sequence (Figure 6A). In keeping with this, it should be noted the differential K_d values for various $(UG)_n$-repeats suggests that off-rate (K_d) from one or two UGs is fast and hence could not be determined (Tables 1 and 2). Longer repeats would be important in maintaining higher local concentration of UGs, thus providing ample opportunity for TDP-43 to bind another UG once dissociated from the previous one. As a result, the dissociation rate (K_d) decreases with the increasing length of $(UG)_n$-repeats (Figure 6B). Finally, it should also be noted that besides the key interaction of TDP-43 RRM1 with a UG sequence, several other factors may influence the binding of this protein *in vivo*. For example, RNA secondary and tertiary structures may also provide another controlling factor during the formation of TDP-43–RNA complex and is well supported by our CD analysis. Furthermore, it has been shown, for example, that TDP-43 does not interact with UG-repeats if they are extensively involved

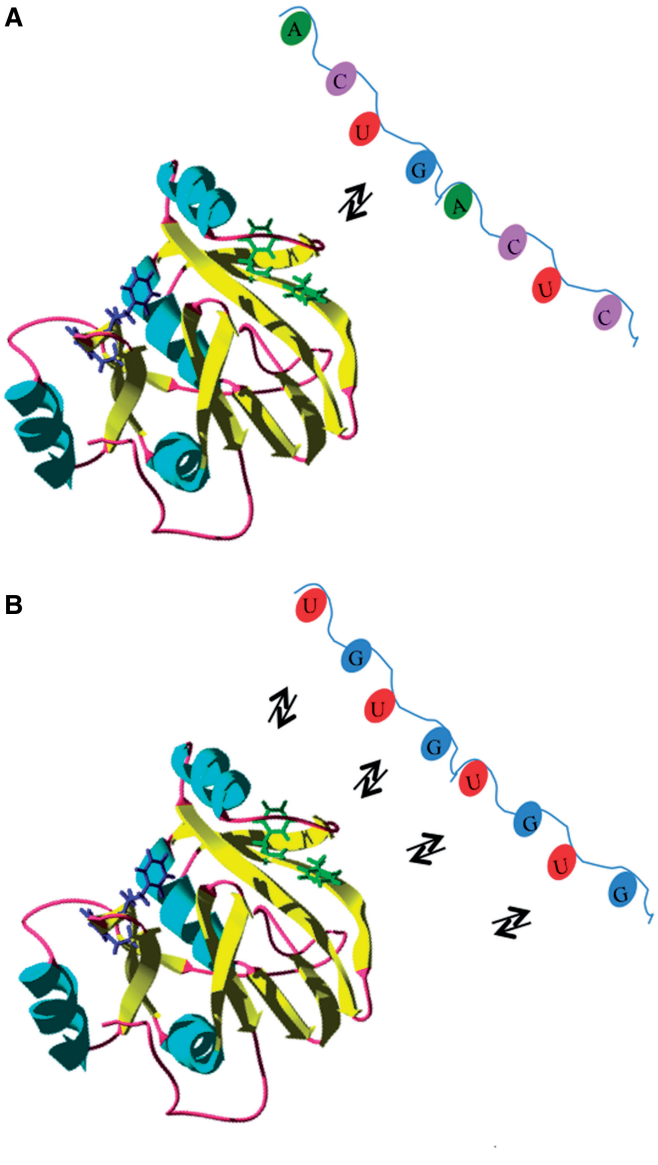

Figure 6. Proposed binding mechanism of TDP-43 (via RRM1) and multi-site RNA targets containing UG-repeats. In this model, long UG-repeats maintain a high local concentration of UGs, which in turn increases the probability of binding other shorter or single UGs stretches that prevent the protein to diffuse away before rebinding can occur.

in base pairing (40) and may explain the ambiguity behind the unpredictable nature of TDP-43 binding to apparently long UG-repeats.

Therefore, our results have provided additional information with regard to the way TDP-43 binds to RNA. In particular, an important observation is represented by the ability of TDP-43 to 'unfold' the RNAs it binds in a manner that does not involve ATP hydrolysis. From a practical point of view, our results will be useful in identifying what are the most likely targets for the binding of this factor and provide a mechanistic framework for TDP-43 affects on RNA structure. Specifically, the frequency of UG-repeats near to each other and the low propensity of the sequence to fold into a highly structured RNA secondary structure would be the most important factors when evaluating whether any given sequence represents a strong binding site.

Finally, it should also be considered that all the CLIP sequences analyzed in this study can be found in the 3′UTR of TDP-43 itself and are essential to mediate the autoregulation of this protein. Given the possible relationships between defects at the level of autoregulation and excessive production of TDP-43, which may lead to aggregation and toxic effects (41,42), understanding of the binding of TDP-43 to these specific sequences may represent the first step in devising therapeutic strategies based on titring TDP-43 level. Furthermore, the conformational changes that may lead to a prion type of disease leading to aggregation and propagation seems unlikely to happen as TDP-43 in the absence and presence of RNA does not seem to have radical structural changes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3.

ACKNOWLEDGEMENTS

We wish to thank Doriano Lamba (IC-CNR) and Giuseppe Legname (SISSA) for help with the CD analyses.

FUNDING

Associazione Ricerca Italiana sulla Sclerosi laterale Amiotrofica (AriSLA) [TARMA]. Funding for open access charge: Associazione Ricerca Italiana sulla Sclerosi laterale Amiotrofica. AriSLA grant TARMA.

Conflict of interest statement. None declared.

REFERENCES

1. Buratti,E. and Baralle,F.E. (2012) TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem. Sci.*, **37**, 237–247.
2. Neumann,M., Sampathu,D.M., Kwong,L.K., Truax,A.C., Micsenyi,M.C., Chou,T.T., Bruce,J., Schuck,T., Grossman,M., Clark,C.M. *et al.* (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, **314**, 130–133.
3. Rademakers,R., Neumann,M. and Mackenzie,I.R. (2012) Advances in understanding the molecular basis of frontotemporal dementia. *Nat. Rev. Neurol.*, **8**, 423–434.
4. Barmada,S.J. and Finkbeiner,S. (2010) Pathogenic TARDBP mutations in amyotrophic lateral sclerosis and frontotemporal dementia: disease-associated pathways. *Rev. Neurosci.*, **21**, 251–272.
5. Mackenzie,I.R., Rademakers,R. and Neumann,M. (2010) TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.*, **9**, 995–1007.
6. Ido,A., Fukuyama,H. and Urushitani,M. (2011) Protein misdirection inside and outside motor neurons in amyotrophic lateral sclerosis (ALS): a possible clue for therapeutic strategies. *Int. J. Mol. Sci.*, **12**, 6980–7003.
7. Collinge,J. (1997) Human prion diseases and bovine spongiform encephalopathy (BSE). *Hum. Mol. Genet.*, **6**, 1699–1705.
8. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
9. Maris,C., Dominguez,C. and Allain,F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
10. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
11. Buratti,E. and Baralle,F.E. (2001) 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.
12. Kuo,P.H., Doudeva,L.G., Wang,Y.T., Shen,C.K. and Yuan,H.S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.*, **37**, 1799–1808.
13. Sephton,C.F., Cenik,C., Kucukural,A., Dammer,E.B., Cenik,B., Han,Y., Dewey,C.M., Roth,F.P., Herz,J., Peng,J. *et al.* (2011) Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.*, **286**, 1204–1215.
14. Xiao,S., Sanelli,T., Dib,S., Sheps,D., Findlater,J., Bilbao,J., Keith,J., Zinman,L., Rogaeva,E. and Robertson,J. (2011) RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell Neurosci.*, **47**, 167–180.
15. Colombrita,C., Onesto,E., Megiorni,F., Pizzuti,A., Baralle,F.E., Buratti,E., Silani,V. and Ratti,A. (2012) TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *J. Biol. Chem.*, **287**, 15635–15647.
16. Tollervey,J.R., Curk,T., Rogelj,B., Briese,M., Cereda,M., Kayikci,M., Konig,J., Hortobagyi,T., Nishimura,A.L., Zupunski,V. *et al.* (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.*, **14**, 452–458.
17. Polymenidou,M., Lagier-Tourenne,C., Hutt,K.R., Huelga,S.C., Moran,J., Liang,T.Y., Ling,S.C., Sun,E., Wancewicz,E., Mazur,C. *et al.* (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.*, **14**, 459–468.
18. Buratti,E. and Baralle,F.E. (2010) The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biol.*, **7**, 420–429.
19. Ayala,Y.M., De Conti,L., Avendano-Vazquez,S.E., Dhir,A., Romano,M., D'Ambrogio,A., Tollervey,J., Ule,J., Baralle,M., Buratti,E. *et al.* (2011) TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.*, **30**, 277–288.
20. Avendano-Vazquez,S.E., Dhir,A., Bembich,S., Buratti,E., Proudfoot,N. and Baralle,F.E. (2012) Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.*, **26**, 1679–1684.
21. Delagoutte,B., Moras,D. and Cavarelli,J. (2000) tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J.*, **19**, 5599–5610.
22. Puglisi,J.D., Tan,R., Calnan,B.J., Frankel,A.D. and Williamson,J.R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, **257**, 76–80.

23. Allain,F.H., Gubser,C.C., Howe,P.W., Nagai,K., Neuhaus,D. and Varani,G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature*, **380**, 646–650.
24. Avis,J.M., Allain,F.H., Howe,P.W., Varani,G., Nagai,K. and Neuhaus,D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
25. Allain,F.H., Howe,P.W., Neuhaus,D. and Varani,G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5772.
26. Buratti,E., Dork,T., Zuccato,E., Pagani,F., Romano,M. and Baralle,F.E. (2001) Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping. *EMBO J.*, **20**, 1774–1784.
27. Ayala,Y.M., Pantano,S., D'Ambrogio,A., Buratti,E., Brindisi,A., Marchetti,C., Romano,M. and Baralle,F.E. (2005) Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.*, **348**, 575–588.
28. Glocker,M.O., Borchers,C., Fiedler,W., Suckau,D. and Przybylski,M. (1994) Molecular characterization of surface topology in protein tertiary structures by amino-acylation and mass spectrometric peptide mapping. *Bioconjug. Chem.*, **5**, 583–590.
29. Suckau,D., Mak,M. and Przybylski,M. (1992) Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc. Natl. Acad. Sci. USA*, **89**, 5630–5634.
30. Hochleitner,E.O., Borchers,C., Parker,C., Bienstock,R.J. and Tomer,K.B. (2000) Characterization of a discontinuous epitope of the human immunodeficiency virus (HIV) core protein p24 by epitope excision and differential chemical modification followed by mass spectrometric peptide mapping analysis. *Protein Sci.*, **9**, 487–496.
31. Gregoire,C.J., Gautheret,D. and Loret,E.P. (1997) No tRNA3Lys unwinding in a complex with HIV NCp7. *J. Biol. Chem.*, **272**, 25143–25148.
32. Carmona,P., Rodriguez-Casado,A. and Molina,M. (1999) Conformational structure and binding mode of glyceraldehyde-3-phosphate dehydrogenase to tRNA studied by Raman and CD spectroscopy. *Biochim. Biophys. Acta*, **1432**, 222–233.
33. Suga,K., Umakoshi,H., Tomita,H., Tanabe,T., Shimanouchi,T. and Kuboi,R. (2010) Liposomes destabilize tRNA during heat stress. *Biotechnol. J.*, **5**, 526–529.
34. Tan,R. and Frankel,A.D. (1992) Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine. *Biochemistry*, **31**, 10288–10294.
35. Mackereth,C.D., Madl,T., Bonnal,S., Simon,B., Zanier,K., Gasch,A., Rybin,V., Valcarcel,J. and Sattler,M. (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, **475**, 408–411.
36. Ramakrishnan,V. and White,S.W. (1998) Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biochem. Sci.*, **23**, 208–212.
37. Varani,G. and Nagai,K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 407–445.
38. Gubser,C.C. and Varani,G. (1996) Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry*, **35**, 2253–2267.
39. Deleault,N.R., Lucassen,R.W. and Supattapone,S. (2003) RNA molecules stimulate prion protein conversion. *Nature*, **425**, 717–720.
40. Passoni,M., De Conti,L., Baralle,M. and Buratti,E. (2012) UG Repeats/TDP-43 Interactions near 5′ splice sites exert unpredictable effects on splicing modulation. *J. Mol. Biol.*, **415**, 46–60.
41. Budini,M. and Buratti,E. (2011) TDP-43 Autoregulation: implications for disease. *J. Mol. Neurosci.*, **45**, 473–479.
42. Buratti,E. and Baralle,F.E. (2011) TDP-43: new aspects of autoregulation mechanisms in RNA binding proteins and their connection with human disease. *FEBS J.*, **278**, 3530–3538.