Published online 27 January 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 6* ***1799–1808***
doi:10.1093/nar/gkp013

Structural insights into TDP-43 in nucleic-acid binding and domain interactions

Pan-Hsien Kuo[1,2], Lyudmila G. Doudeva[2], Yi-Ting Wang[1,2,3], Che-Kun James Shen[2] and Hanna S. Yuan[2,3,4,*]

[1]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, [2]Institute of Molecular Biology, [3]Taiwan International Graduate Program, Chemical Biology and Molecular Biophysics, Academia Sinica and [4]Graduate Institute of Biochemistry and Molecular Biology, National Taiwan University, Taipei, Taiwan, ROC

Received October 14, 2008; Revised November 27, 2008; Accepted January 7, 2009

ABSTRACT

TDP-43 is a pathogenic protein: its normal function in binding to UG-rich RNA is related to cystic fibrosis, and inclusion of its C-terminal fragments in brain cells is directly linked to frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS). Here we report the 1.65 Å crystal structure of the C-terminal RRM2 domain of TDP-43 in complex with a single-stranded DNA. We show that TDP-43 is a dimeric protein with two RRM domains, both involved in DNA and RNA binding. The crystal structure reveals the basis of TDP-43's TG/UG preference in nucleic acids binding. It also reveals that RRM2 domain has an atypical RRM-fold with an additional β-strand involved in making protein–protein interactions. This self association of RRM2 domains produced thermal-stable RRM2 assemblies with a melting point greater than 85°C as monitored by circular dichroism at physiological conditions. These studies thus characterize the recognition between TDP-43 and nucleic acids and the mode of RRM2 self association, and provide molecular models for understanding the role of TDP-43 in cystic fibrosis and the neurodegenerative diseases related to TDP-43 proteinopathy.

INTRODUCTION

Proteins constitute and carry out all kinds of intra- and extra-cellular events and therefore mutations, deletions, misfolding and aggregation of protein molecules, leading to gain or loss of protein functions, are related to numerous genetic and sporadic diseases. TAR DNA-binding protein 43 (TDP-43) is a ubiquitously expressed protein whose normal function and abnormal aggregation are directly linked to the common lethal genetic disease, cystic fibrosis (1) and to two neurodegenerative disorders: frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS) (2,3).

TDP-43 was originally identified as a transcriptional factor, repressing the transcription of *HIV-1* gene (4), mouse *SP-10* gene (5), and the expression of human cyclin-dependent kinase 6 (Cdk6) (6). TDP-43 is also a splicing factor binding to the intron 8/exon 9 junction of the cystic fibrosis transmembrane conductance regulator (CFTR) gene (7–9), and the intron 2/exon 3 junction of apoA-II gene (10) to inhibit exon splicing. The binding of TDP-43 to the UG-repeats located at the 3′-splice site of CFTR intron 8 leads to exon 9 skipping and the transcription of a shorter transcript, resulting in the expression of an inactive CFTR protein in cystic fibrosis patients (11). Moreover, TDP-43 has also been shown to be a human low molecular weight neurofilament (*h*NFL) mRNA-binding protein in spinal motor neurons (12), and a neuronal activity-responsive factor in the dendrites of hippocampal neurons (13), suggesting its involvement in regulating mRNA stability, transport and local translation in neurons. Therefore, TDP-43 is both a DNA-binding and a RNA-binding protein, bearing multiple functions in transcriptional repression, pre-mRNA splicing and translational regulation.

Recently, breakthrough studies showed that TDP-43 is the major disease protein in the pathogenesis of both FTLD with ubiquitin inclusions and ALS (14,15). FTLD, referring to a heterogeneous group of neurodegenerative disorders, is the second most common form of presenile dementia after Alzheimer's disease (3). FTLD with ubiquitin-positive, tau-negative inclusions (FTLD-U) accounts for 60% of the phenotypes associated with FTLD syndromes. ALS is the most common adult-onset motor neuron disease, and in common with FTLD, has similar ubiquitinated inclusions in the surviving motor cells (16). TDP-43 was identified as the major component protein in the ubiquitinated inclusions in both FTLD-U and ALS disorders. Pathological TDP-43 in

*To whom correspondence should be addressed. Tel: +886 2 27884151; Fax: +886 2 27826085; Email: hanna@sinica.edu.tw

the cytoplasmic and intranuclear inclusions is hyperphosphorylated, ubiquitinated and cleaved to ~25 kDa C-terminal fragments in affected brain regions. These abnormal aggregates of phosphorylated and ubiquitinated TDP-43 thus define a new group of neurodegenerative diseases, the TDP-43 proteinopathies (3,17,18). Missense mutations have now been identified in gene encoding TDP-43, *TARDBP*, in familial and sporadic ALS, providing further evidence of a direct link between TDP-43 function and neurodegeneration (16,19,20). Among the unique features of TDP-43 inclusions are that they are not amyloid deposits, and are negative for tau, α-synuclein, β-amyloid and expanded polyglutamines, indicating that they might have a distinct aggregated structure.

The *TARDBP* gene is highly conserved in human, mouse, *Drosophila melanogaster* and *Caenorhabditis elegans* (21). Sequence analysis identified two RNA-recognition motifs, RRM1 and RRM2, and a C-terminal glycine-rich domain in TDP-43 (Figure 1), similar to the domain organization of heterogeneous nuclear ribonucleoproteins (hnRNP) family proteins such as hnRNP A1 and A2/B1 (22). RRMs are common RNA-binding motifs, with two highly conserved hexameric and octameric segments denoted respectively as ribonucleoprotein 2 (RNP2) and ribonucleoprotein 1 (RNP1) (23). The conserved RNP segments in TDP-43 are involved in binding to TAR DNA sequences (4) and RNA sequences with UG-repeats (8,24). The glycine-rich C-terminal tail of TDP-43 can interact with the hnRNP family proteins to form the hnRNP-rich complex involved in splicing inhibition (25). Mutation analyses further suggest that the TDP-43 glycine-rich domain is essential for the CFTR exon 9-skipping activity (21,24,25).

Although TDP-43 is involved in diverse transcriptional and splicing events, and plays a key role in the pathologies of neurodegenerative diseases, its biochemical properties, molecular structure, domain assembly and DNA/RNA recognition mechanism are largely unknown. Here we report the crystal structure of a TDP-43 RRM2 domain in complex with a single-stranded DNA and demonstrate the basis of its TG/UG preference. The RRM2 domain has an atypical RRM-fold with an additional β-strand

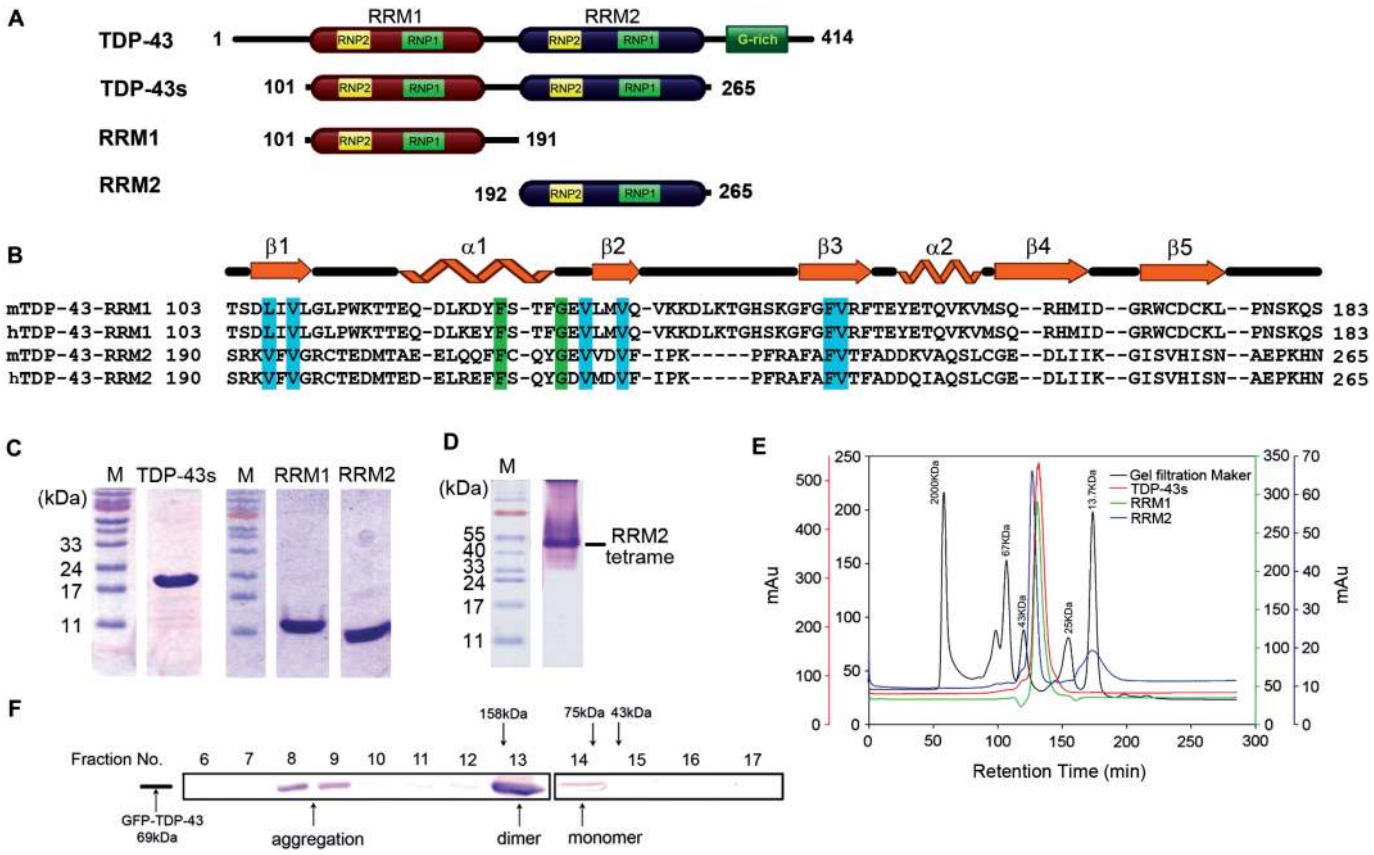


Figure 1. Domain structures and assembly of TDP-43 proteins. (**A**) TDP-43 has two RRM domains, RRM1 and RRM2, and a C-terminal glycine-rich domain. In this study, three truncated TDP-43 proteins were constructed: TDP-43s, RRM1 and RRM2. (**B**) Amino-acid sequences of mouse and human TDP-43 in the RRM1 and RRM2 domains. The secondary structures listed above the sequences are derived from the crystal structure of RRM2–DNA complex. (**C**) The purity of TDP-43 truncated proteins was assayed by 12.5% SDS–PAGE. (**D**) RRM2 appeared as a tetramer in the native 20% PAGE gel. (**E**) Gel filtration (Superdex 200) profiles of TDP-43 truncated proteins show that TDP-43s, RRM1 and RRM2 all had a molecular weight of ~40 kDa, suggesting that the recombinant TDP-43s was a homodimer, and RRM1 and RRM2 were homotetramers. (**F**) The GFP-fused TDP-43, with a molecular weight of 69 kDa, was expressed in human 293T cells for size exclusion chromatography analysis. The fractionated cell extract eluted from a Superdex 200 column were blotted by TDP-43 antibodies. GFP-TDP-43 was mainly eluted with a size of a dimer. The molecular weight markers are: aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (43 kDa).

involved in making domain–domain interactions. The role of this β-strand in forming a highly thermal-stable higher order complex is discussed.

MATERIALS AND METHODS

Protein expression and purification

The gene fragments encoding mouse TDP-43s (residues 101–265), RRM1 (residues 101–191) and RRM2 (residues 192–265) were amplified by PCR. All of the PCR products were digested with two restriction enzymes, *Bam*HI and *Hind*III, and then inserted into pQE30 expression vector (Qiagen, USA) to generate the N-terminal His-tagged constructs. All of the truncated TDP-43 proteins were over-expressed in *Escherichia coli* M15 strain. Cells were grown in LB medium to a density of ~0.5 OD_{600} and induced using 0.8 mM IPTG for 22 h at 20°C.

Cell extracts were loaded onto a Ni-NTA affinity column (Qiagen, USA) and washed with a step gradient of buffer containing 0.5 M imidazole, 100 mM NaCl and 20 mM Tris pH 7.6. Peak fractions were dialyzed against 10 mM β-mercaptoethanol and 20 mM Tris pH 7.6 and then applied to a HiTrap heparin column (GE, USA). Protein samples were eluted with a step gradient of buffer containing 1 M NaCl, 10 mM β-mercaptoethanol and 20 mM Tris pH 7.6. The eluted proteins were dialyzed against 5 mM NaCl, 10 mM β-mercaptoethanol and 20 mM Tris–HCl at pH 7.6 and then purified on a Superdex 200 gel column. The purified protein samples were concentrated to ~20 mg/ml with Vivaspin ultrafiltration unit (Sartorius, Germany).

The GFP-fused TDP-43 was expressed in human 293T cells which were maintained as monolayers in DMEM supplemented with 100 units/ml penicillin-streptomycin and 10% FCS (GIBCO) at 37°C under 5% CO_2. The plasmid encoded GFP-TDP-43 was constructed by insertion of the full-length TDP-43 cDNA into the EGFP-N3 (Promega, USA). The plasmid was transfected into 293T cells using the calcium phosphate precipitation method. After 12 h of transfection, the medium was changed, and the cells were further incubated for 24 h. The cells extracts were loaded into a Superdex 200 column (GE, USA) in the buffer containing 50 mM Tris–HCl, 400 mM NaCl at pH 7.5. The collected fractions were concentrated for western blot analysis.

Filter binding assay

All of the RNA (and DNA) substrates for filter-binding assays were 5′-end labeled with [γ-^{32}P]ATP by T4 PNK. The labeled RNA (10 pmol) was then incubated with truncated proteins for 10 min at room temperature in binding buffer containing 20 mM Tris–HCl at pH 8.0. The mixture was filtered through a BA 85 nitrocellulose membrane (Schleicher and Schuell, USA) overlaid on a nylon membrane (Roche, Germany) in a 60-well slot blot apparatus (Bio-Rad, USA). After extensive washing, the protein–RNA complex-bound nitrocellulose membrane and free-RNA-bound nylon membrane were air dried and exposed to a phosphor imaging plate. Film signals were counted by Luminescent image analyzer LAS-1000plus (Fujifilm, Japan) and the affinity was calculated using the Hill equation with three parameters. The K_d values were deduced from the protein concentrations at which half of the RNA substrates were protein-bound.

Circular dichroism (CD)

The thermal denaturing melting points of TDP-43 truncated proteins were measured three times by a CD spectrometer AVIV CD400. The CD spectra were scanned from 25°C to 85°C at a wavelength of 218 nm and the melting point was estimated by AVIV program. The protein concentration was 0.1 mg/ml in a buffer containing 300 mM NaCl in 50 mM phosphate buffer (pH 7.0).

Crystallization, structural determination and refinement

The RRM2 used for crystallization was purified with a procedure slightly different from the one used for biochemical analysis. All the Tris–HCl buffer was replaced by PBS buffer at pH 7.9, and 1 mM DTT was added in the last purification step using a HiTrap heparin column. The eluted proteins were dialyzed against 1% glycerol and 20 mM Tris–HCl at pH 7.9, and concentrated to ~6 mg/ml. The purified RRM2 was mixed with a single-stranded DNA with a sequence of 5′-TTGAGCGTT-3′ in a one-to-one molar ratio. Crystals of RRM2–DNA complex were grown by hanging drop vapor-diffusion method at room temperature, by mixing 1 μl of protein–DNA solution with 1 μl of reservoir solution containing 2 M $(NH_4)_2SO_4$, 0.1 M phosphate-citrate at pH 4.2. The reservoir solution contained an additional 10% glycerol.

The X-ray diffraction data were collected at beamline 13C1 of the NSRRC in Hsinchu, Taiwan at −150°C. The data were processed and scaled by HKL2000 (26) and all of the diffraction statistics are listed in Table 1. The RRM2–DNA complex crystallized in the F222 cubic space group, with one molecule per asymmetric unit. The structure of the complex was solved by molecular replacement using the NMR structure of human TDP-43–RRM2 (PDB accession code: 1WF0) as the searching model. After refinement of the protein molecule, the structure of DNA was built as guided by the (2Fo-Fc) Fourier maps. The structure model was then manually rebuilt with WinCoot and refined with CCP4. The data collection and refinement statistics are summarized in Table 1. The coordinates and structural factors of the RRM2–DNA complex have been deposited in the Protein Data Bank with a PDB ID of 3D2W.

RESULTS

TDP-43 is a dimer with four RRM domains

To study the biochemical properties of TDP-43, we constructed three truncated forms of mouse TDP-43, which shares a high sequence identity of 96.1% with human TDP-43 (Figure 1). TDP-43s covered both RRM domains from residues 101–285, RRM1 covered the first RRM from residues 101–191, and RRM2 covered the second RRM from residues 192–285. The three truncated forms

Table 1. Crystallographic statistics of RRM2–DNA complex crystal

Data collection and processing	Values
Wavelength (Å)	1.0
Space group	F222
Cell dimensions ($a/b/c$) (Å)	41.283/87.290/125.345
Resolution (Å)	1.65
Observed reflections	63 632
Unique reflections	13 674
Redundancy[a]	4.7 (4.5)
Completeness[a] (%)	98.7 (97.0)
R_{sym}[a] (%)	5.8 (32.5)
$I/\sigma(I)$ [a]	22.2 (3.5)
Refinement statistics	
Resolution range	25.46–1.65
Reflections (work/test)	12633/1010
R-factor/R-free (%)	20.7/24.6
Non-hydrogen atoms	
Protein	561
Solvent molecules	127
Model quality	
r.m.s. deviations in	
Bond length (Å)	0.009
Bond angle (°)	1.257
Average B-factor (Å^2)	23.5
Ramachandran plot (%)	
Most favored	93.8
Additionally allowed	6.2
Generously allowed	0
Disallowed	0

[a]The last shell (1.70–1.65 Å) statistics are listed in parenthesis.

of TDP-43 fused with an N-terminal His-tag were over-expressed in *E. coli* and purified by chromatographic methods using a Ni^{2+}-NTA agarose affinity column, followed by a HiTrap heparin and a Superdex 200 gel filtration column. All three recombinant proteins had homogeneity greater than 98% as analyzed by SDS–PAGE (Figure 1C).

The calculated molecular weights of the truncated proteins were 20659.59 Da for TDP-43s, 12351.09 Da for RRM1 and 10080.45 Da for RRM2. However, the gel filtration profiles gave molecular weights of approximately 40 kDa for all of these three proteins (Figure 1E). RRM2 was further applied to a native PAGE where it appeared as a homotetramer with a molecular weight between 40 kDa and 55 kDa (Figure 1D). These results show that TDP-43s is a homodimer with four RRM domains. On the other hand, RRM1 and RRM2 are homotetramers, with four copies of the RRM in each tetrameric assembly.

To further determine the oligomerization of TDP-43 in a cellular environment, a GFP-fused TDP-43 was expressed in human 293T cells for size exclusion chromatography analysis. The cell extract was fractionated by a Superdex 200 gel filtration column and the fractionated eluents were blotted by TDP-43 antibodies. As shown in Figure 1F, the GFP-TDP-43 with a molecular weight of 69 kDa was eluted mainly with a size of a dimer (~150 kDa). Small portions of the GFP-TDP-43 were eluted as monomers (~70 kDa) and large-size aggregates. This result thus suggests that TDP-43 forms a homodimer in a cellular environment.

TDP-43s binds ssDNA and dsDNA with preference for TG-rich sequences

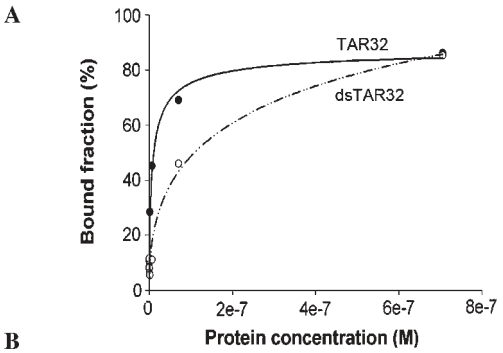
To find out the DNA-binding affinity of TDP-43 truncated proteins, the three truncated proteins were incubated with single-stranded and with double-stranded 32-mer DNA of HIV-1 TAR sequences (TAR32). The binding affinities between TDP-43 fragments and DNA were analyzed by nitrocellulose filter-binding assays. TDP-43s bound the single-stranded 32-mer DNA of TAR sequence (TAR32) slightly better than the double-stranded DNA, with a K_d of 10.6 ± 0.2 nM for ssTAR32 and 83.8 ± 9.5 nM for dsTAR32 (Figure 3). RRM1 also bound single- and double-stranded TAR32 DNA with comparable affinities ($K_d = 6.2 \pm 0.6$ nM for ssTAR32 and 16.7 ± 1.0 nM for dsTAR32), whereas RRM2 had one-order lower binding affinities ($K_d = 137.5 \pm 14.8$ nM for ssTAR32 and 600.5 ± 96.8 nM for dsTAR32). These results suggest that TDP-43s binds both double-stranded and single-stranded DNA of TAR sequences, and that both RRM1 and RRM2 domains can bind DNA.

To find out whether TDP-43 binds site-specifically to TG-rich DNA sequences, we synthesized a number of 12-mer DNAs bearing either TAR sequences, TG repeats, CA repeats or a P12 random sequence without any TG repeats (Figure 2C). The TDP-43s K_d values for single-stranded and double-stranded TAR and TG repeats (ranging between 89.3 nM and 147.1 nM) were one to two orders of magnitude less than those for non-TG sequences (382.5 nM for a single-stranded CA repeat and 1315 nM for a double-stranded P12 random sequence). These data suggest that TDP-43s prefers to bind to TG-rich 12-mer DNAs with at least an order higher affinity than non-TG sequences.

TDP-43s prefers to bind UG-rich RNA

A previous study showed that the GST-fused TDP-43 recombinant protein can bind a single-stranded RNA with a minimum number of six UG repeats and the binding affinity increases with the number of repeats (8). To further quantify the binding affinity of TDP-43 for RNA, we synthesized a series of RNA nucleotides with or without UG repeats of different lengths. TDP-43s bound single-stranded UG-repeat RNAs with high affinity [$K_d = 14.2 \pm 1.4$ nM for single-stranded $(UG)_6$, see Figure 3]. Consistent with previous results, TDP-43s bound RNA with more UG-repeats with higher affinity, as evidenced by the K_d values of 3060 nM for $(UG)_3$, 115.3 nM for $(UG)_4$, 14.2 nM for $(UG)_6$ and 2.79 nM for $(UG)_8$. These results suggest that TDP-43s can bind single-stranded RNA with at least three UG repeats and the binding affinity increases by about one order with each addition of two UG repeats.

The truncated mutant RRM1 bound RNA in a way similar to that of TDP-43s. RRM1 also preferred to bind to RNA with more UG-repeats, with a K_d of 4690 nM for $(UG)_3$ and 65.2 nM for $(UG)_6$. However, the binding affinity trend of RRM2 to UG-repeat RNA was different: it bound to ss$(UG)_3$ with higher affinity (379 nM) than to ss$(UG)_6$ (4240 nM). This result implies



B

Apparent K_d between TDP-43 truncated proteins and DNA substrates

	K_d (nM)	
	TAR32	dsTAR32
TDP-43s (a.a. 101-285)	10.6 ± 0.2	83.8 ± 9.5
RRM1 (a.a. 101-191)	6.2 ± 0.6	16.7 ± 1.0
RRM2 (a.a. 192-285)	137.5 ± 14.8	600.5 ± 96.8

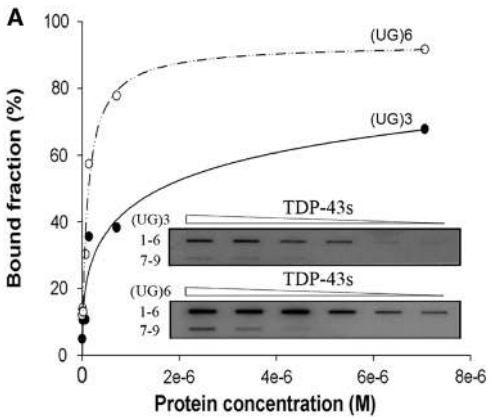
C

Apparent K_d between TDP-43s and 12-bp DNA

Oligo	K_d (nM)
Single-stranded DNA	
TAR12a sense	107.5 ± 3.5
TAR12b sense	219 ± 19.7
(TG)6	89.3 ± 19.3
(CA)6	382.5 ± 174.6
Double-stranded DNA	
dsTAR12a	96.4 ± 0.8
dsTAR12b	147.1 ± 97.4
ds(TG)6	127 ± 25.4
P12	1315 ± 205

TAR12a: 5′-CTTTTTGCCTGT-3′
TAR12b: 5′-TGGGTCTCTCTG-3′
P12: 5′-CGGGATATCCCG-3′

Figure 2. Binding affinities of TDP-43 truncated proteins and DNA, analyzed by nitrocellulose filter-binding assays. (**A**) TDP-43s bound to both single-stranded and double-stranded TAR DNA sequence (32-mer). The 5′-end ^{32}P-labeled DNA (10 pmol) was incubated with 0.01–40 μM TDP-43s and the resultant protein–DNA complexes trapped in the nitrocellulose filters were quantified. (**B**) Summary of the apparent dissociation constants of TDP-43 truncated proteins and TAR 32-mer ssDNA and dsDNA. Both RRM1 and RRM2 domains were capable of DNA binding. (**C**) Summary of the apparent dissociation constants of TDP-43s and various 12-mer ssDNA and dsDNA. TDP-43s prefers to bind to TG-rich 12-mer DNAs with affinity at least one order higher than non-TG sequences.



B

Apparent K_d between TDP-43 truncated proteins and RNA substrates

	K_d (nM)		
	TDP-43s (a.a. 101-285)	RRM1 (a.a. 101-191)	RRM2 (a.a. 192-285)
Single-stranded RNA			
(UG)3	3060 ± 2100	4690 ± 629	379 ± 123
(UG)4	115.3 ± 16.69		
(UG)6	14.2 ± 1.4	65.2 ± 1.8	4240 ± 381
(UG)8	2.79 ± 0.14		
U12	300.5 ± 79.9		
A12	904.5 ± 205.7		

Figure 3. Binding affinities of TDP-43 truncated proteins and RNA, analyzed by nitrocellulose filter-binding assays. (**A**) TDP-43s bound to single-stranded RNA containing three and six UG repeats. (**B**) Summary of the apparent dissociation constants of TDP-43 truncated proteins and RNAs. TDP-43s prefers to bind UG-rich 12-mer ssRNA with affinity at least one order higher than RNA without any UG-repeats.

that RRM2 might have a different conformation or a different tertiary structure of its domain assembly than TDP-43s and RRM1. Moreover, the trend of binding affinity to $(UG)_6$ RNA was: TDP-43s > RRM1 > RRM2, suggesting that both RRM domains are necessary for achieving the best binding affinity of TDP-43.

To further confirm that TDP-43 prefers UG-rich sequences, we prepared 12-mer RNAs without any UG-repeat: a poly(U) sequence (U_{12}) and a poly(A) sequence (A_{12}). Compared to 12-mer UG-repeat RNAs, the dissociation constants between TDP-43s and U_{12} and A_{12} were 20–60-fold higher, in the range of 300–900 nM. This result verifies that TDP-43s prefers to bind UG-rich 12-mer RNA with at least one order higher affinity than RNA without any UG-repeats.

Overall crystal structure of RRM2–DNA complex

To reveal the molecular basis underlying the interactions between TDP-43 and DNA/RNA, all of the truncated forms of TDP-43 were screened for co-crystallization conditions with RNA and DNA of various sequences and lengths. Only RRM2 co-crystallized with a single-stranded 10-mer DNA (5′-GTTGAGCGTT-3′) with a preferred TG binding site at the third and fourth positions (underlined). The RRM2–DNA complex crystallized in cubic F222 space group with one molecule per asymmetric unit. The structure of the complex was solved by molecular replacement using the NMR solution structure of human TDP-43 RRM2 domain (PDB entry code: 1WF0, unpublished results) as the search model.

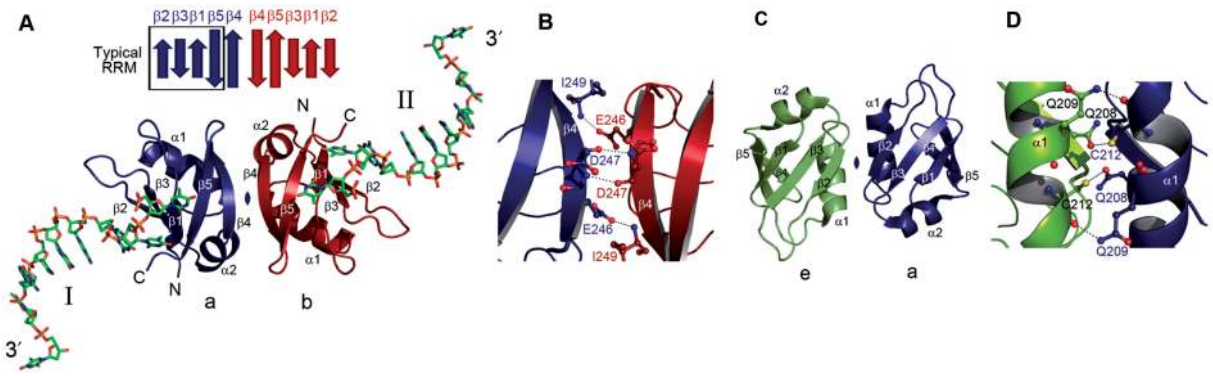


Figure 4. Crystal structure of RRM2–DNA complex. (**A**) A ribbon model of RRM2 dimer bound to single-stranded DNAs. Molecule **a** (in blue) is related to molecule **b** (in red) by a 2-fold crystallographic symmetry axis. DNA molecules are displayed as stick models. A classical RRM domain contains four β-strands (β2-β3-β1-β5) as marked by the rectangle on the top of the structure; however, TDP-43 RRM2 has an extra β4 strand next to β5. (**B**) A pair of hydrogen bonds are formed between the main-chain atoms of Asp247 (Asp247-O to Asp247-N), and a pair of hydrogen bonds are formed between Glu245 (Oε2) and Ile249 (N) between the two antiparallel β4 strands to stabilize the RRM2 dimeric structure. (**C**) The RRM2 domain (monomer **a**) also interacts with the neighboring 2-fold symmetry-related molecule (monomer **e**). This view is rotated ~90 degrees vertically to that of panel A. (**D**) Four hydrogen bonds were formed between monomer **a** and **e**: Gln209 (Nε2) to Cys212 (O), and Glu208 (Oε1) to Cys (Sγ).

The final model contained one RRM2 molecule (residues 190–261) and nine out of ten nucleotides (T2–T10) with an R-factor of 20.7% and an R-free of 24.6% for 63 632 reflections up to a resolution of 1.65 Å.

RRM2 was a tetramer in low salt conditions, however, RRM2 appeared as a dimer in the crystals grown from acidic high-salt conditions. A gel filtration analysis further confirmed that RRM2 indeed dissociated into dimers under crystallization conditions of 0.1 M phosphate-citrate and 2.0 M $(NH_4)_2SO_4$ at pH 4.2 (data not shown). The ribbon model of RRM2 dimer is shown in Figure 4A, where RRM2 molecules **a** and **b** are related to each other by a crystallographic 2-fold symmetry. A typical RRM contains a four-stranded β-sheet (23), however, RRM2 in TDP-43 has an additional β4 next to β5 (β2-β3-β1-β5-**β4**). Thus, the RRM2 has a αβ sandwich structure containing a five-stranded β-sheet packed with two α-helices. The dimeric interface is formed mainly through the interactions of the two antiparallel β4 strands (residues 245–250), with a buried solvent accessible surface area of 685 $Å^2$. Two hydrogen bonds were formed between the main-chain atoms of Asp247 (Asp247-O to Asp247-N, **a** to **b** and **b** to **a**), and two more hydrogen bonds were formed between Glu245 (Oε2) and Ile249 (N) in the two antiparallel β4 strands to stabilize the dimeric structure (Figure 4B). The RRM2 dimer thus has an extended β-sheet consisting of 10 antiparallel β-strands. The RRM2 molecule (monomer **a**) also interacted with another 2-fold symmetry-related RRM2 molecule (monomer **e**) via β2 and α1 (Figure 4C and D). Four hydrogen bonds (**a** to **e** and **e** to **a**) were formed between the side-chain atoms in α1 helix: Gln209 (Nε2) to Cys212 (O), and Gln208 (Oε1) to Cys212 (Sγ). The buried surface between monomer **a** and **e** was 901 $Å^2$, slightly higher than the interface between monomer **a** and **b**.

The single-stranded DNA was bound to the accessible surface of the β-sheet of RRM2 with the three 5′-end nucleotides, T2, T3 and G4, interacting extensively with the β-sheet residues, whereas the 3′-end nucleotides stretched away and made no interaction with RRM2. The single-stranded DNA (strand I), bound to RRM2 monomer **a**, further interacted with the DNA (strand III) that was bound to the neighboring RRM2 molecule (monomer **c**), forming a double-stranded-like conformation (Figure 5). Watson–Crick base pairs were formed between G6 and C7 of strand I and III. A5 of strand I also interacted with G8 of strand III, whereas T9 of strand I was stacked with G4 of strand III. These interactions likely stabilized the crystal packing between RRM2–DNA complexes and therefore led to the successful crystallization of a high-resolution crystal.

Specific interactions between TDP-43 and TG sequence

Why does TDP-43 prefer to bind TG-rich and UG-rich sequences? The interactions between RRM2 and DNA are schematically displayed in Figure 5B. The three 5′-end nucleotides, particularly T3 and G4, played the key role in the interactions. The thymine of T3 made three hydrogen bonds to RRM2: O4 forming hydrogen bonds with Ser258 (Oγ); N3 with Asn259 (backbone O); O2 with Glu261 (backbone N). Besides hydrogen bonding, Phe194 and Phe231, the two aromatic residues conserved in RNP2 and RNP1 segments, also stacked or formed van der Waals' interactions with thymine of T3 (Figure 5C). The guanine of G4 made four hydrogen bonds with RRM2: N7 and O6 with Lys192 (Nζ), and N1 and N2 with Glu261 (backbone O). The aromatic side chain of Phe221 and Phe231 stacked or formed van der Waals' interactions with the guanine of G4.

By replacing the thymine T3 with a uracil and a cytosine, we modeled a uracil and a cytosine bound at the T3-binding pocket. We found that a cytosine can make optimally two hydrogen bonds with RRM2, instead of three hydrogen bonds originally identified with thymine. On the contrary, a uridine at the same location may form optimally three hydrogen bonds with RRM2, providing a structural basis for the specific uridine preference in this pocket (Figure 5D). Similarly, we replaced the guanine

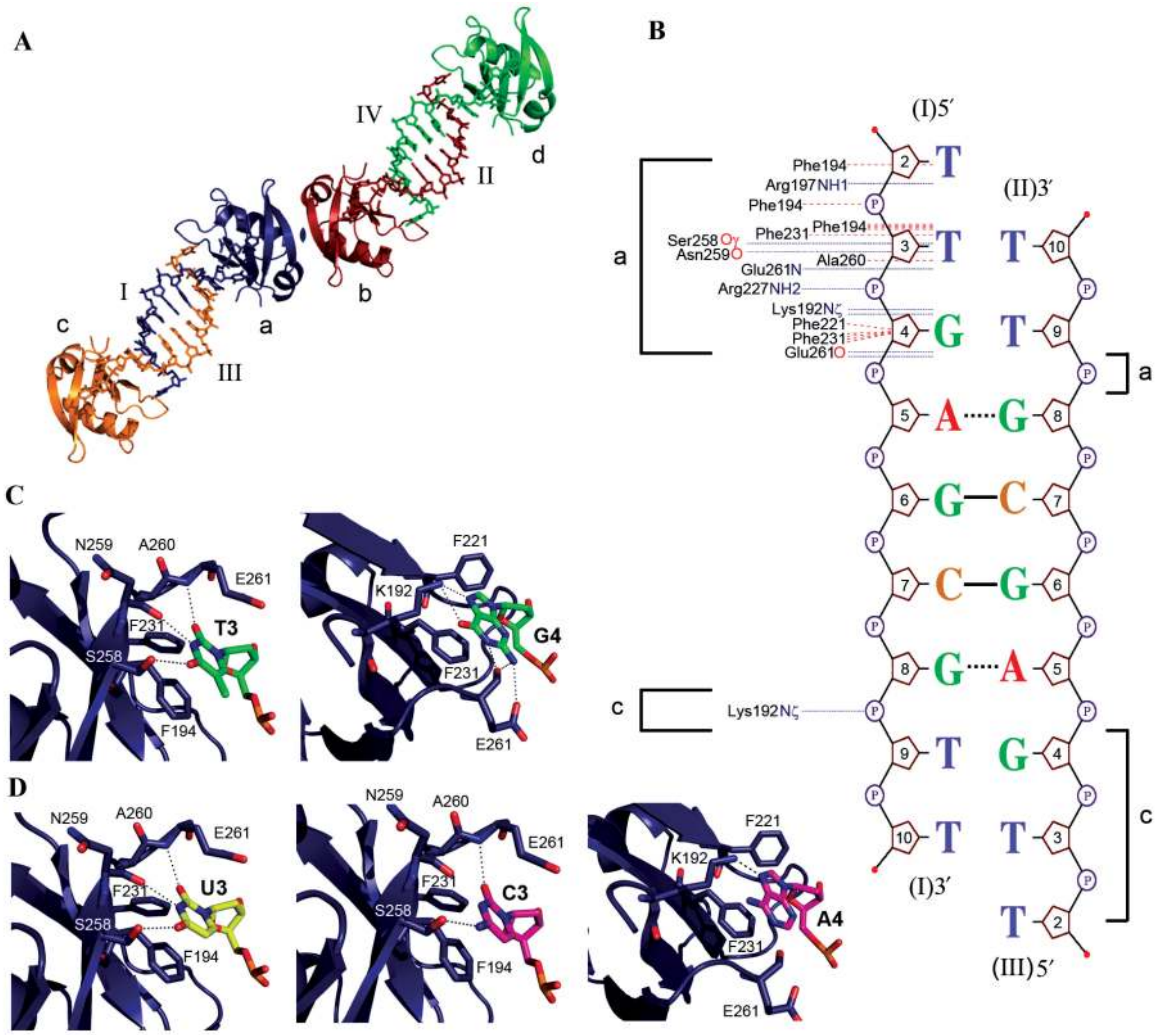


Figure 5. Interactions between TDP-43 RRM2 and DNA. (**A**) Single-stranded DNA bound to RRM2 formed a double-stranded-like conformation, with the DNA bound to the neighboring RRM2 molecule. Four RRM2–DNA complexes are shown here to demonstrate the interactions between DNA strand I (bound to molecule a) and strand III (bound to molecule c). (**B**) Schematic diagrams of the detailed contacts between RRM2 and DNA. Hydrogen bonds are shown by blue dotted line, and non-bonded contacts are shown by red dotted line. Watson–Crick base pairs were formed between G6 and C7 (indicated by solid line) of strand I and III. (**C**) Extensive hydrogen-bond networks and non-bonded interactions are identified in the T3-binding pocket (left) and G4-binding pocket (right) in TDP-43 RRM2–DNA crystal structure The bases of T3 and G4 stack with Phe194 and Phe231, respectively. (**D**) A uracil (U3, left) and a cytosine (C3, middle) are modeled at the T3-binding pocket, whereas an adenine is modeled at the G4 pocket (A4, right).

base in G4 with an adenine, and found that the possible hydrogen bond number was reduced from 4 to 1. This study thus elucidates the structural basis of TG- and UG-preference of TDP-43 in DNA/RNA binding.

RRM2 domain forms a highly thermal-stable assembly

The RRM2 dimer (monomers **a** and **b**) observed in the crystals should be highly stable since a 10-stranded anti-parallel β-sheet was formed in the dimer. To examine the thermal stability of the TDP-43 truncated proteins, thermal denaturation experiments were carried out and the melting of β-strand structure was monitored by CD at 218 nm (Figure 6). TDP-43s had a melting point of 49.7 ± 0.9°C, whereas RRM1 had a comparable melting temperature of 49.5 ± 0.7°C in 200 mM NaCl at pH 7.0. However, the RRM2 was not melted with retained β-strand structure up to 85°C, suggesting that its melting point was increased by more than 35°C as compared to those of TDP-43s and RRM1. This result shows that RRM2 forms a highly stable structure and that it has distinct biophysical properties as compared to TDP-43s and RRM1.

A close look at the crystal packing of RRM2–DNA complexes further shows that RRM2 dimers were assembled into a fibril-like solenoid structure (Figure 6C). The dimeric TDP-43 RRM2 was packed against the neighboring 2-fold-symmetry related dimers through the interactions between the β2 strands and α1 helices. Each RRM2 domain used the β4 strand on one edge of the β-sheet to interact with the RRM2 domain within the dimer, whereas it used the α1/β2 on the other edge of

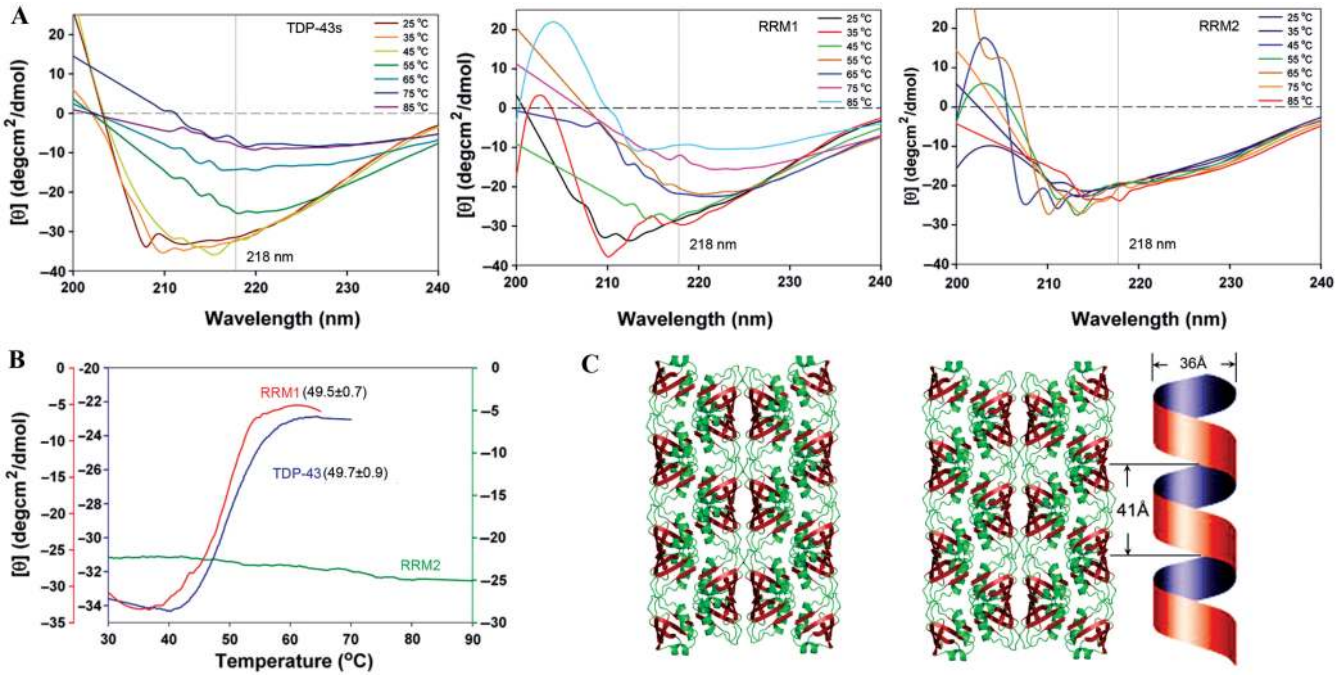


Figure 6. The RRM2 domain of TDP-43 forms a highly thermal-stable assembly as monitored by CD. (**A**) Thermal denaturation of TDP-43s, RRM1 and RRM2, was assayed by CD from 25 to 85°C in 200 mM NaCl at pH 7.0. (**B**) The melting point, monitored at a wavelength of 218 nm, was 49.7 ± 0.9°C for TDP-43s, and 49.5 ± 0.7°C for RRM1. RRM2 was not melted up to 85°C, suggesting that it was highly stable and had a melting point greater than 85°C. (**C**) The stereo view of the crystal packing of RRM2 shows that the RRM2 dimers interact with the neighboring dimers to generate a left-handed super-helix structure. For clarity, DNA molecules have been removed and two super helices packed side by side are shown here.

the β-sheet to interact with the neighboring RRM2 dimers. The 10-stranded β-sheets in the RRM2 dimer thus were wrapped into a left-handed super helix with α-helices buried inside forming the hydrophobic core. This 'super-helix' had diameters of 36 Å and 28 Å with a helical pitch of 41 Å (Figure 6C). Although it remains unclear why RRM2 domain was more resistant to thermal denaturation, the structural study suggests that the high melting point of RRM2 was likely due to protein dimerization and formation of higher order thermal-stable assemblies.

DISCUSSION

Both RRM1 and RRM2 of TDP-43 are involved in site-specific DNA/RNA interactions

About 2% of the gene products in humans contain at least one RNA recognition motif (RRM), making RRM one of the most abundant protein domains (23). Most RRM proteins have two to six copies of RRMs, which are folded into an αβ sandwich structure with a four β-stranded pleated sheet packed against two α-helices. Crystal structures of a number of RRM proteins containing two tandem RRM motifs, in a way similar to TDP-43, have been reported, including hnRNP A1 (27,28), Hud (29), Sxl (30), PABP (31) and FIR (32). hnRNP A1 forms a dimer when it is bound with single-stranded nucleic acids, whereas Hud, Sxl and PABP are monomers with similar relative domain arrangement between RRM1 and RRM2. The superposition of TDP-43 RRM2 domain onto the RRM2 domain of two representative RRM proteins, hnRNP A1 dimer (PDB entry code: 1U1O) and Hud monomer (PDB entry code: 1FXL), showed that the RRM2 domains in these proteins share a similar RRM fold, except that TDP-43 RRM2 has an extra β4 strand that is absent in hnRNP A1 and Hud (Figure 7).

The superposition of TDP-43 onto hnRNP A1 further shows that the single-stranded nucleic acid molecules are all bound to the β-sheet of RRMs at similar locations (displayed as a tube in Figure 7). The thymine in T3 is bound to a binding pocket in TDP-43 and this pocket is also identified in hnRNP A1, HuD and Sxl where it binds to adenine, uracil and guanine, respectively (Figure 7D). Different hydrogen-bonding networks and non-bonded interactions differentiate the binding specificity of these RRM proteins. The two aromatic residues, Phe231 (conserved in RNP1) and Phe194 (conserved in RNP2), are involved in stacking with T3. Previous mutational studies showed that the conserved Phe residues in RRM1 of TDP-43 play a key role in nucleic-acid recognition (8). Therefore the RRM1 domain of TDP-43 likely interacts with nucleic acids, particularly with TG-rich sequence, in a way similar to that of RRM2 domain.

However, the G4 guanine-binding pocket identified in TDP-43 is only found in hnRNP1 A1 but not in HuD, Sxl and PABP, indicating that TDP-43 is structurally more closely related to hnRNP1 A1 (Figure 7E). Our biochemical data first suggest that TDP-43 is a homodimer containing four copies of RRM. Secondly, TDP-43 shares not only an identical domain organization with two RRM followed by a glycine-rich domain and also the highest

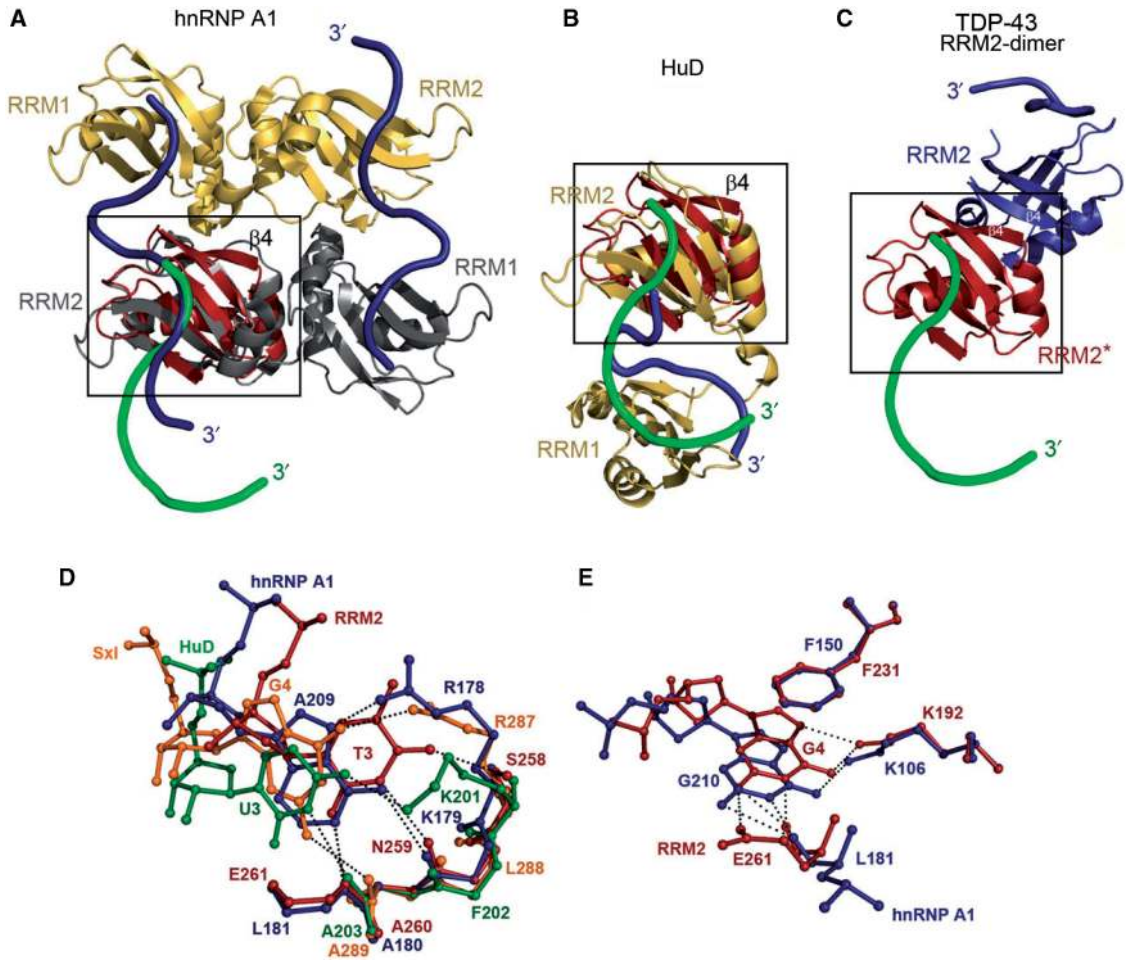


Figure 7. Comparison of RRM domain assembly between TDP-43, hnRNP A1 and HuD. (**A**) The RRM2 domain of TDP-43 (in red) was superimposed onto the RRM2 of hnRNP A1. hnRNP A1 is a homodimer (in yellow and gray) bound to two strands of RNA (schematically displayed as a navy blue tube). The DNA bound to TDP-43 is displayed in green. (**B**) The RRM2 of TDP-43 (red) was superimposed onto the RRM2 of HuD (yellow). All the RRM2 domains in hnRNP A1, HuD and TDP-43 were fixed in the same orientation as marked by a black frame. (**C**) The dimeric interface in the TDP-43 RRM2 dimer is atypical, compared to those found in hnRNP A1 and HuD. (**D**) A similar T3-binding pocket was identified in several RRM proteins: TDP-43 in red (bound to T3); hnRNP A1 in navy blue (bound to A209); HuD in green (bound to U3); and Sxl in orange (bound to G4). (**E**) A similar G4 binding pocket in TDP-43 (in red) was only identified in hnRNP A1 (in navy blue, bound to G210). The PDB entry codes of the structures used in this figure are: 1U1O for hnRNP A1, 1FXL for HuD and 1B7F for Sxl.

sequence identity (21%) with hnRNP A1. Thirdly, TDP-43 shares the closest structural similarity in nucleic-acid binding with hnRNP1. Therefore, based on these three lines of evidence, we suggest that TDP-43 forms a homodimer with a domain arrangement similar to that of hnRNP A1. Also similar to hnRNPA1, both RRM1 and RRM2 are involved in nucleic-acid interactions.

TDP-43 RRM2 domain has a unique atypical dimeric interface

Comparison of the protein interfaces among RRM family proteins, we notice that the dimeric interface between the two RRM2 domains is exclusively observed only in the crystal structure of TDP-43 RRM2–DNA complex. In hnRNP A1, the monomer containing RRM1–RRM2 is packed against the other monomer in an antiparallel orientation, so that RRM1 interacts with RRM2 intermolecularly and intramolecularly (Figure 7A). In Hud, Sxl and PABP, RRM1 is packed against RRM2 of the same molecule using a different interface. However, the RRM2 domain in TDP-43 has a unique dimeric interface that has not been observed previously among other RRM proteins.

This dimeric interface between the two RRM2 domains in TDP-43 is unique since it is formed by the β4 strand that is only present in TDP-43 but not in other RRM proteins. We further show that the melting point of RRM2 domain was greater than 85°C, at least 35°C higher than that of TDP-43s (49.7 ± 0.9°C) and RRM1 (49.5 ± 0.7°C), suggesting that RRM2 was assembled into a thermal-stable structure, that must be different from the native state. It remains unclear whether the atypical dimerization of the RRM2 domains are related to the pathogenic inclusions of TDP-43 C-terminal fragments observed in FTLD and ALS (for a recent review, see (33)). This well-organized thermal-stable RRM2 dimeric structure thus offers a testable model for the

study of the pathogenic aggregates in the group of neurodegenerative diseases known as TDP-43 proteinopathies.

FUNDING

This work has been supported by grants from the Academia Sinica and the National Science Council, Taiwan, ROC. We thank the staff of the beamlines BL-13B1 and BL-13C1 at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, ROC. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine. Funding for open access charge: Academia Sinica, Taiwan.

Conflict of interest statement. None declared.

REFERENCES

1. Buratti,E. and Baralle,F.E. (2008) Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Front. Biosci.*, **13**, 867–878.
2. Talbot,K. and Ansorge,O. (2006) Recent advances in the genetics of amyotrophic lateral sclerosis and frontotemporal dementia: common pathways in neurodegenerative disease. *Human Mol. Genet.*, **15**, 182–187.
3. Kwong,L.K., Uryu,K., Trojanowski,J.Q. and Lee,V.M.-Y. (2008) TDP-43 proteinopathies: neurodegenerative protein misfolding diseases without amyloidosis. *Neurosignals*, **16**, 41–51.
4. Ou,S.-H.I., Wu,F., Harrich,D., Garcia-Martinez,L.F. and Gaynor,R.B. (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.*, **69**, 3584–3596.
5. Abhyankar,M.M., Urekar,C. and Reddi,P.P. (2007) A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues- role for TDP-43 in insulator function. *J. Biol. Chem.*, **282**, 36143–36154.
6. Ayala,Y.M., Misteli,T. and Baralle,F.E. (2008) TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *Proc. Natl Acad. Sci. USA*, **105**, 3785–3789.
7. 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.
8. 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.
9. Ayala,Y.M., Pagani,F. and Baralle,F.E. (2006) TDP43 depletion rescues aberrant *CFTR* exon 9 skipping. *FEBS Lett.*, **580**, 1339–1344.
10. Mercado,P.A., Ayala,Y.M., Romano,M., Buratti,E. and Baralle,F.E. (2005) 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.
11. Mantovani,V., Garagnani,P., Selva,P., Rossi,C., Ferrari,S., Cenci,M., Calza,N., Cerreta,V., Luiselli,D. and Romeo,G. (2007) Simple method for haplotyping the poly(TG) repeat in individuals carrying the IVS8 5T allele in the CFTR gene. *Clin. Chem.*, **53**, 531–533.
12. Strong,M.J., Volkening,K., Hammond,R., Yang,W., Strong,W., Leystra-Lantz,C. and Shoesmith,C. (2007) TDP43 is a human low molecular weight neurofilament (*h*NFL) mRNA-binding protein. *Mol. Cell Neurosci.*, **35**, 320–327.
13. Wang,I.-F., Wu,L.-S., Chang,H.-Y. and Shen,C.-K.J. (2008) TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *J. Neurochem.*, **105**, 797–806.
14. Neumann,M., Sampathu,D.M., Kwing,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.
15. Arai,T., Hasegawa,M., Akiyama,H., Ikeda,K., Nonaka,T., Mori,H., Mann,D., Tsuchiya,K., Yoshida,M., Hashizume,Y. *et al.* (2006) TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Comm.*, **351**, 602–611.
16. Sreedharan,J., Blair,I.P., Tripathi,V.B., Hu,X., Vance,C., Rogelj,B., Ackerley,S., Durnall,J.C., Williams,K.L., Buratti,E. *et al.* (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science*, **319**, 1668–1672.
17. Cairns,N.J., Neumann,M., Bigio,E.H., Holm,I.E., Troost,D., Hatanpaa,K.J., Foong,C., White,C.L. III, Schneider,J.A., Kretzschmar,H.A. *et al.* (2007) TDP-43 in familial and sporadic frontotemporal lobar degeneration with ubiquitin inclusions. *Am. J. Pathol.*, **171**, 227–240.
18. Forman,M.S., Trojanowski,J.Q. and Lee,V.M.-Y. (2007) TDP-43: a novel neurodegenerative proteinopahty. *Curr. Opin. Neurobiol.*, **17**, 548–555.
19. Deerlin,V.M.V., Leverenz,J.B., Bekris,L.M., Bird,T.D., Yuan,W., Elman,L.B., Clay,D., Wood,E.M., Chen-Plotkin,A.S., Martinez-Lage,M. *et al.* (2008) *TARDBP* mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol.*, **7**, 409–416.
20. Gitcho,M.A., Baloh,R.H., Chakraverty,S., Mayo,K., Norton,J.B., Levitch,D., Hatanpaa,K.J., White,C.L. III, Bigio,E.H., Caselli,R. *et al.* (2008) TDP-43 A315T mutation in familial motor neuron disease. *Ann. Neurol.*, **63**, 535–538.
21. Wang,H.-Y., Wang,I.-F., Bose,J. and Shen,C.-K.J. (2004) Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics*, **83**, 130–139.
22. Dreyfuss,G., Matunis,M.J., Pinol-Roma,S. and Burd,C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.*, **62**, 289–321.
23. Maris,C., Dominguez,C. and Allain,F.H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
24. 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.
25. Buratti,E., Brindisi,A., Giombi,M., Tisminetzky,S. and Ayala,Y.M. (2005) TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail, an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon9 splicing. *J. Biol. Chem.*, **280**, 37572–37584.
26. Otwinowski,Z. and Minor,W. (1993) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
27. Ding,J., Hayashi,M.K., Zhang,Y., Manche,L., Krainer,A.R. and Xu,R.-M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.*, **13**, 1102–1115.
28. Oubridge,C., Ito,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) Crystal structure at 1.92Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
29. Wang,X. and Hall,T.M.T. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.*, **8**, 141–145.
30. Crowder,S.M., Kanaar,R., Rio,D.C. and Alber,T. (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc. Natl Acad. Sci. USA*, **96**, 4892–4897.
31. Deo,R.C., Bonanno,J.B., Sonenberg,N. and Burley,S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, **98**, 835–845.
32. Crichlow,G.V., Zhou,H., Hsiao,H.-h., Frederick,K.B., Debrosse,M., Yang,Y., Folta-Stogniew,E.J., Chung,H.-J., Fan,C., De La Cruz, E.M. *et al.* (2008) Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. *EMBO J.*, **27**, 277–289.
33. Wang,I.-F., Wu,L.-S. and Shen,C.-K.J. (2008) TDP-43: an emerging new player in neurodegenerative diseases. *Trends Mol. Medicine*, **14**, 479–485.