



Article

NMR Fragment-Based Screening against Tandem RNA Recognition Motifs of TDP-43

Gilbert Nshogoza ¹, Yaqian Liu ¹, Jia Gao ¹, Mingqing Liu ¹, Sayed Ala Moududee ¹ , Rongsheng Ma ¹, Fudong Li ¹, Jiahai Zhang ¹, Jihui Wu ¹, Yunyu Shi ^{1,2} and Ke Ruan ^{1,*}

¹ Hefei National Laboratory for Physical Sciences at the Microscale, School of Life Sciences, University of Science and Technology of China, Hefei 230027, China

² CAS, Center for Excellence in Biomacromolecules, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: kruan@ustc.edu.cn; Tel.: +86-551-6360-7767

Received: 28 April 2019; Accepted: 28 June 2019; Published: 30 June 2019



Abstract: The TDP-43 is originally a nuclear protein but translocates to the cytoplasm in the pathological condition. TDP-43, as an RNA-binding protein, consists of two RNA Recognition Motifs (RRM1 and RRM2). RRM1 and RRM2 are known to involve both protein-nucleotide and protein-protein interactions and mediate the formation of stress granules. Thus, they assist the entire TDP-43 protein with participating in neurodegenerative and cancer diseases. Consequently, they are potential therapeutic targets. Protein-observed and ligand-observed nuclear magnetic resonance (NMR) spectroscopy were used to uncover the small molecule inhibitors against the tandem RRM of TDP-43. We identified three hits weakly binding the tandem RRMs using the ligand-observed NMR fragment-based screening. The binding topology of these hits is then depicted by chemical shift perturbations (CSP) of the ¹⁵N-labeled tandem RRM1 and RRM2, respectively, and modeled by the CSP-guided High Ambiguity Driven biomolecular DOCKing (HADDOCK). These hits mainly bind to the RRM2 domain, which suggests the druggability of the RRM2 domain of TDP-43. These hits also facilitate further studies regarding the hit-to-lead evolution against the TDP-43 RRM domain.

Keywords: epigenetics; protein-RNA interaction; RRM domain inhibitor; NMR fragment-based screening; TDP-43

1. Introduction

RNA recognition motifs (RRMs) play diverse roles in post-transcriptional gene expression events such as RNA transport, localization, stability, and mRNA and rRNA processing. RRM is also known as the ribonucleoproteins (RNP) domain, as it contains the short and conserved elements RNP1 and RNP2, or RNA binding domain (RBD), that are abundantly distributed in higher vertebrates [1] and ubiquitously found in all kingdoms of life, including viruses and prokaryotes. In addition, they also participate in important functions such as microRNA biogenesis, apoptosis, and cell division [2,3]. RRMs are not only known to be involved in protein–nucleotide interactions, but also in protein–protein interactions [4].

The transactive response DNA-binding Protein 43kDa (TDP-43) is a RRM-containing protein, which plays important functions in mRNA metabolism regulation, including transcription repression, exon skipping, and RNA splicing [5,6]. TDP-43 is originally a nuclear protein, but translocates to the cytoplasm upon a pathological condition. It is a ubiquitously expressed, highly conserved, and multifunctional RNA and DNA-binding protein [7]. TDP-43 stabilizes the mRNA of human low-molecular-weight neurofilament (hNFL) [8]. Depletion of TDP-43 has important consequences in essential metabolic processes in human cells, like nuclear shape deformation, apoptosis, and misregulation of the cell cycle [9]. The disruption of TDP-43 auto-regulation impacts both localization

of TDP-43 and its level, which results in TDP-43 accumulation in the cytoplasm. Based on its crucial roles in RNA processing, dysfunctional TDP-43 causes some abnormalities in alternative mRNA splicing, miRNA biogenesis, and RNA-rich granules formation [10].

The dysregulation of TDP-43 is hence associated with a variety of human diseases, especially neurodegenerative diseases, e.g., frontotemporal lobar degeneration (FTLD), amyotrophic lateral sclerosis (ALS), brain ischemia, aging, and Alzheimer's disease [11–13]. For instance, in cases of FTLD and ALS, TDP-43 is the main constituent of their ubiquitin inclusions [14]. During the stress conditions, TDP-43 is localized in the cytoplasm, with mRNA binding to its RRM and glycine-rich domain, and thus forms the isolated liquid compartment enriching the mRNA and proteins. Such stress granules (SGs) in cells and in pathological brain tissue play crucial roles in FTLD/ALS pathology [15,16]. Aggregate-prone TDP-43 variants or exposure to oxidative stress generates distinct TDP-43 inclusions devoid of SGs [17]. The toxicity of the TDP-43 overexpression requires the presence of functional RNA Recognition motifs [18–20]. Recently, the proteinopathy of both important mutations (D169G and K263E located at RRM1 and RRM2, respectively) was computationally explored and the mutants are more prone to aggregation, causing neurological disorders [21].

Apart from the TDP-43 involvement in neurodegenerative diseases, an accumulating amount of evidence suggests that TDP-43 is a cancer responsive factor. TDP-43 positively contributes to the anticancer activity for curcumin in MCF-7 cells [22] and as a tumor suppressor by partnering with the TRIM16 in inhibiting the viability and proliferation of neuroblastoma and breast cancer cells [23]. In addition, normal levels of TDP-43 might be a crucial protective factor for cells under apoptotic insult [24]. On the contrary, the TDP-43 inhibition suppressed cervical cancer cell growth and induced cell cycle arrest while its overexpression promoted cancer cell progression and drove the cell cycle [25]. TDP-43 may regulate melanoma cell proliferation and metastasis by modulating glucose metabolism [26]. TDP-43 also plays an oncogenic role in malignant glioma cell progression by stabilizing small nucleolar RNA host gene 12 (SNHG12) [27]. The findings demonstrated that TDP-43 regulates the MALAT1, a non-coding RNA overexpressed in non-small cell lung cancer (NSCLC), through direct binding to MALAT1 RNA at the 3' region by RRM, whose participation is compulsory. This controls the growth, invasion, and migration of NSCLC cells [28]. Reduced tumor progression, including proliferation and metastasis, was observed upon the knockdown of TDP-43 in triple-negative breast cancer (TNBC) and RRM involvement is assured [29]. These studies suggest that targeting the TDP-43 RRM domains may, therefore, be an effective therapeutic approach for neurodegenerative diseases and cancers.

Although more is known about the TDP-43 biology and its association with neurodegenerative and cancer diseases, the development of treatments toward TDP-43 is mostly lagging behind those targeting other proteins involved in such diseases [30]. RRM and RNA complexes have long been attractive targets for small molecule inhibition targeting the RNA, not the protein [31,32]. Firstly, the aminoacridine derivative was discovered to interrupt the formation of RNA and U1A RRM1 complex [33]. Additionally, a high-throughput screening assay, based on AlphaScreen[®], technology was used to characterize DNA and RNA oligonucleotides (bt-TAR-32 and bt-TG6, respectively) binding to TDP-43 and their interaction inhibition was assessed [34]. Later, that series of 4-aminoquinoline derivatives were characterized for their capacity to modulate TDP-43 metabolism and function, whereby they bind to TDP-43, reduce its interaction with the oligonucleotide, and stimulate caspase-mediated cleavage of TDP-43 [35], but information is still lacking on the binding topology. Furthermore, some medicinal treatment reduces the TDP-43 inclusions through the autophagy pathway were discussed [36]. However, no compounds directly targeting RRM domains of TDP-43 have been uncovered to our best knowledge.

NMR spectroscopy is a powerful approach which has been extensively used by the pioneers in fragment-based drug discovery for detecting molecular interactions between the target and the fragment libraries [37–39] and to facilitate structure-based drug design [40]. Consistently, the fragment-based screening approach has been fruitful for identifying hits for the challenging protein-protein interaction

“hot-spots” [41–45]. We expect it shall be effective in the case of the shallow RNA binding pocket of TDP-43 tandem RRM.

Here, we carried out automated NMR fragment-based screening [46] to identify three hits of the tandem RRM of TDP-43. Chemical shift perturbations of the ^{15}N labeled TDP-43 tandem RRM demonstrate that these hits bind to the same site, mainly on the RRM2 domain. It has also been validated by the chemical shift perturbation experiments for TDP-43 RRM2 alone. The CSP-driven HADDOCK was used to generate the protein-hits binding mode. Collectively, our work provides a class of compounds for further hit-to-lead evolution of the TDP-43 RRM domain and paves the path for targeting protein-RNA interactions using the fragment-based approach.

2. Results

Structurally, TDP-43 tandem RRM is approximately 160 amino acids long and displays a $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ arrangement of secondary structure, with an additional β -hairpin named $\beta 3'\beta 3''$ [47] or $\beta 5$ [48,49] which is located between $\alpha 2\beta 4$, and extends the β -sheet surface to be accessible to binding by multiple RNA nucleotides. This leads to a rare RRM orientation type ($\beta 2\beta 4$) and the 14-aa linker needs to connect four β -strands instead of two [2,47]. Diverse studies revealed that TDP-43 tandem RRM can interact with both short and long single-stranded nucleic acids rich in UG/TG, either separately or collectively, to achieve high affinity and specificity [47–49]. Given the RNA recognition mode by tandem RRM, TDP-43 RRM is independent of each other in unbound form but they establish a rigid structure upon RNA binding on the flat surface β -sheet [47]. In general, this RNA-recognition pocket is much shallower than the ATP-binding sites of kinases. Hence, it poses a grand challenge for conventional high throughput screening aimed at discovering strong binders. Conversely, the fragment-based approach has proven fruitful for uncovering the initial hits, albeit at weak affinities.

NMR ligand-observed methods detect the weak protein-ligand binding by detecting changes in the characteristics of the ligand spectrum that occur upon binding to the protein. Using the ligand-based experiments, i.e., saturation transfer difference (STD) [50], water ligand observed via gradient spectroscopy (WaterLOGSY) [51], Carr–Purcell–Meiboom–Gill (CPMG) [52], and ligand-based 1D proton, we found 17 hits from the primary screening of 89 cocktails containing 10 compounds each (Figure 1a). The binders present signals while the non-binders present no signals in the STD spectra. Accordingly, the binders show inverted or a fast decay of signals in the WaterLOGSY and CPMG experiments, respectively. The combined output of these spectra enabled the identification of primary screening hits from cocktails. It is worth noting that the reference 1D proton spectra of each individual compound might be slightly different from the screening spectra as a different buffer was used to be better compatible with TDP-43 tandem RRM. The primary screening hits were further validated by the secondary screening for individual hits using the same set of NMR experiments (Figure 1b and Figure S1). The aromatic peaks of the hit are depicted as they suffer less from the interference of buffer signals. The secondary screening eliminated 13 primary hits, probably due to sample aggregation in cocktails, ambiguous selection of hits with degenerated chemical shifts, and/or spectrometer instability. Among the remaining 4 hits, hit 2 demonstrated a distinct topology relative to hits 1 and 3 (Figure 1c).

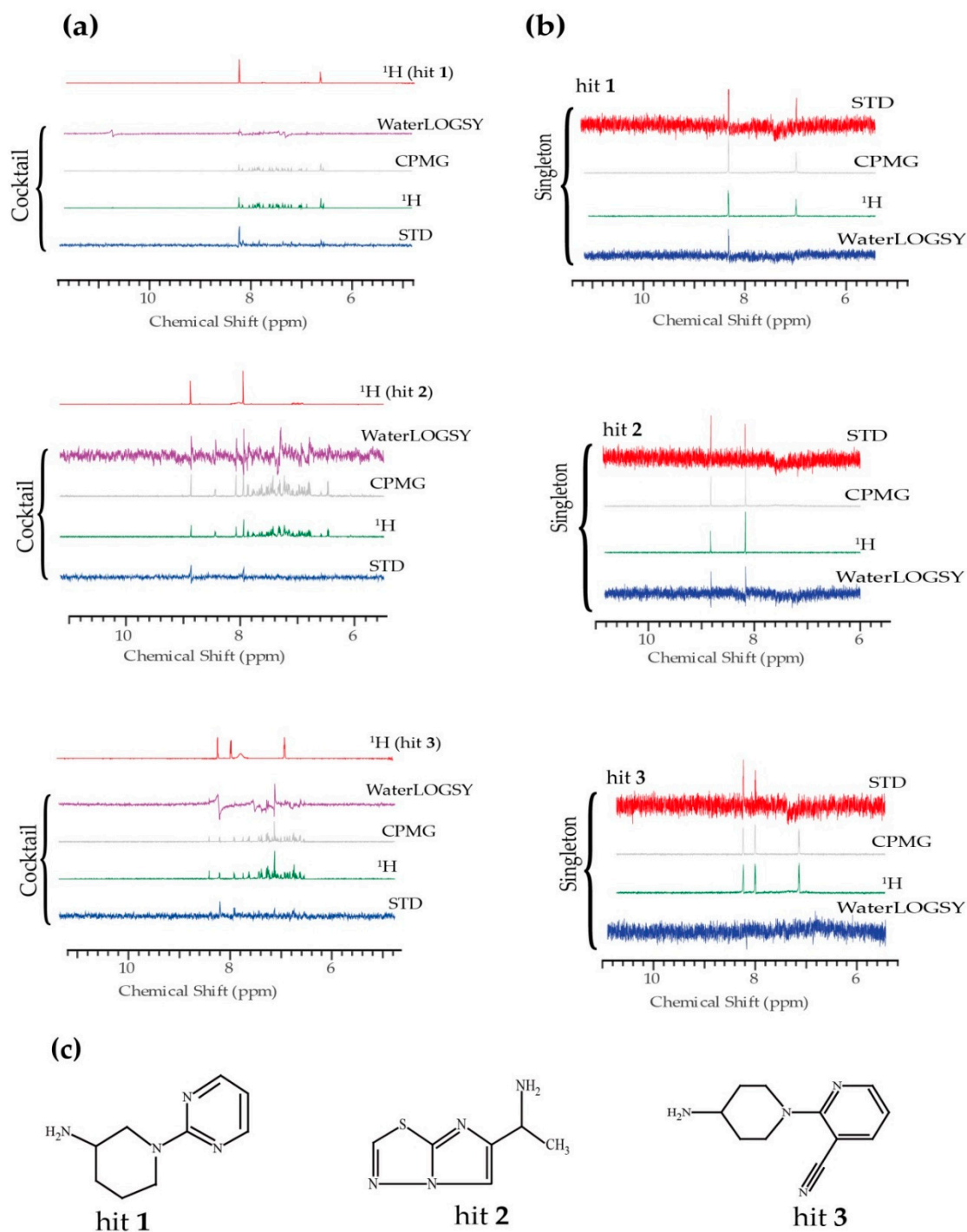


Figure 1. NMR fragment-based screening against the tandem RRM domain of TDP-43. (a) The primary screening WaterLOGSY, CPMG, ^1H and STD spectra for three representative cocktails. The ^1H reference spectrum of the respective hit is shown for comparison. (b) The secondary screening spectra for individual hit 1, 2, and 3, respectively. (c) The chemical structures of hits 1, 2, and 3.

The 4 secondary screening hits were then cross-validated using the chemical shift perturbations (CSPs) of the ^{15}N -labeled tandem RRMs of TDP-43 and 3 of them induced significant chemical shift changes of the tandem RRM (Figures 2 and 3). This approach has been extensively applied in the interrogation of protein-ligand interactions in an affinity ranging from nM to mM. As CSP is a sensitive indicator of chemical environment changes induced by ligand titration, it is particularly powerful in the detection of weak bindings. The linewidths of the amide signals of TDP-43 tandem RRM show almost

no changes upon titration of hit 1 (Table S1), which suggests that hit 1 induces no protein aggregation. This is a useful approach to remove false positives, which are commonly found in drug screening because of protein aggregation [53]. Titration of hit 1 induces dose-dependent CSPs of residues G245, E246, H256, I257, S258 (Figure 2b and Figure S2). However, the curve does not reach the saturation point, as it is limited by the weak binding affinity and the low aqueous solubility of the hit. Hence, the binding affinity of those weak binders cannot be robustly estimated from CSPs. The disturbed residues were then mapped on the surface representation of the solution structure of TDP-43 tandem RRM (PDB code: 4BS2) [47]. Residues H256, I257, S258 locate on the β_4 strand, while residues G245 and E246 bridge the α_2 and β_3 (Figure 2c).

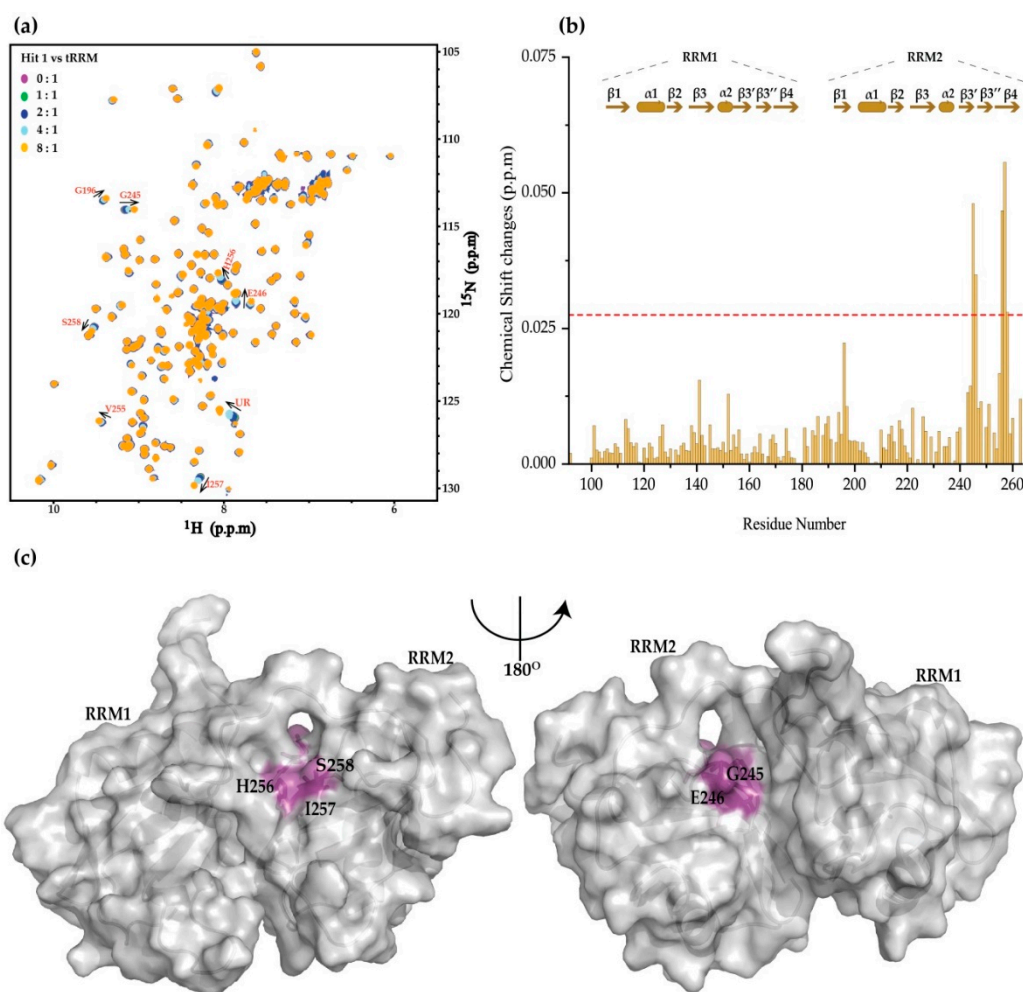


Figure 2. The binding topology of hit 1 on the tandem RRM of TDP-43 using NMR chemical shift perturbations. (a) The chemical shift perturbations of ^{15}N -labeled tandem RRM domain of TDP-43 upon titration of hit 1. The ligand/protein molar ratios are annotated. The perturbed residues are labeled and the arrows indicate the direction of chemical shift changes. UR stands for unassigned residue. (b) Chemical shift changes of the TDP-43-tandem RRM domain are at the ligand protein molar ratio of 8:1. The red horizontal dashed line represents two standard deviations above the averaged chemical shift changes of residues. (c) Surface representation of TDP-43 tandem RRM domain (PDB code: 4BS2) showing the purple-colored residues with significant chemical shift changes.

Consistently, hits 2 and 3 titrations also point to the same binding topology in the tandem RRM of TDP-43 (Figure 3). For example, hit 2 perturbed residues G245, H256, and I257 (Figure 3a,c), while hit 3 induced significant CSPs for residues G245, E246, H256, and I257 (Figure 3b,d). The similarity of the

binding pattern of the three hits suggests that weak but specific binders were successfully identified using the NMR fragment-based screening.

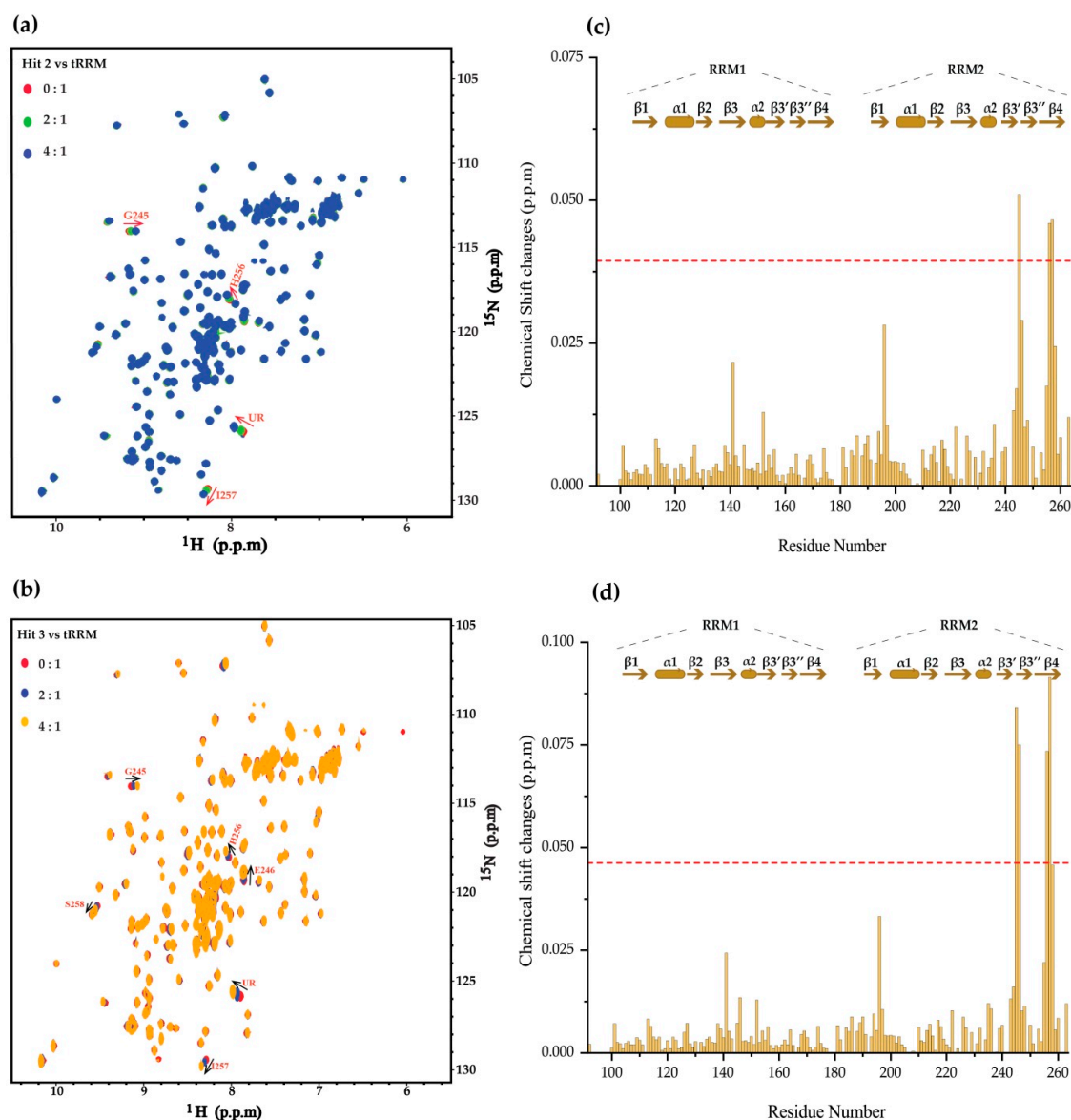


Figure 3. Chemical shift perturbations of tandem RRM upon binding of hit 2 and 3. (a,b) The chemical shift perturbations of TDP-43 tandem RRM domain induced by titration of hit 1 and 2, respectively. Annotated are the hits: Protein molar ratios. UR stands for unassigned residue. (c,d) Residue-by-residue chemical shift changes of tandem RRM at the hit/protein molar ratio of 8:1 for compound 2 and 3, respectively. The red dashed lines represent two standard deviations above the averaged chemical shift changes of residues.

Having confirmed that 3 different hits bind on the same site of the TDP-43 RRM2 domain, we further investigated whether RRM2 alone is sufficient for ligand binding. Hit 2 was thus titrated to the ^{15}N -labeled RRM2 domain of TDP-43 (Figure 4a). Consequently, the residues G245, on loop bridging the $\alpha 2$ and $\beta 3'$, H256, and I257, located on $\beta 4$ -strand, were perturbed (Figure 4b). Those residues were mapped on the surface representation of the TDP-43 RRM2 [49] domain in complex with a single-stranded DNA (Figure 4c). The hit binds to the same sites of either TDP-43 tandem RRM or RRM2 alone. That is to say, TDP-43 RRM2 is the main contributor for ligand binding and should be considered as the target for follow-up hit-to-lead evolutions.

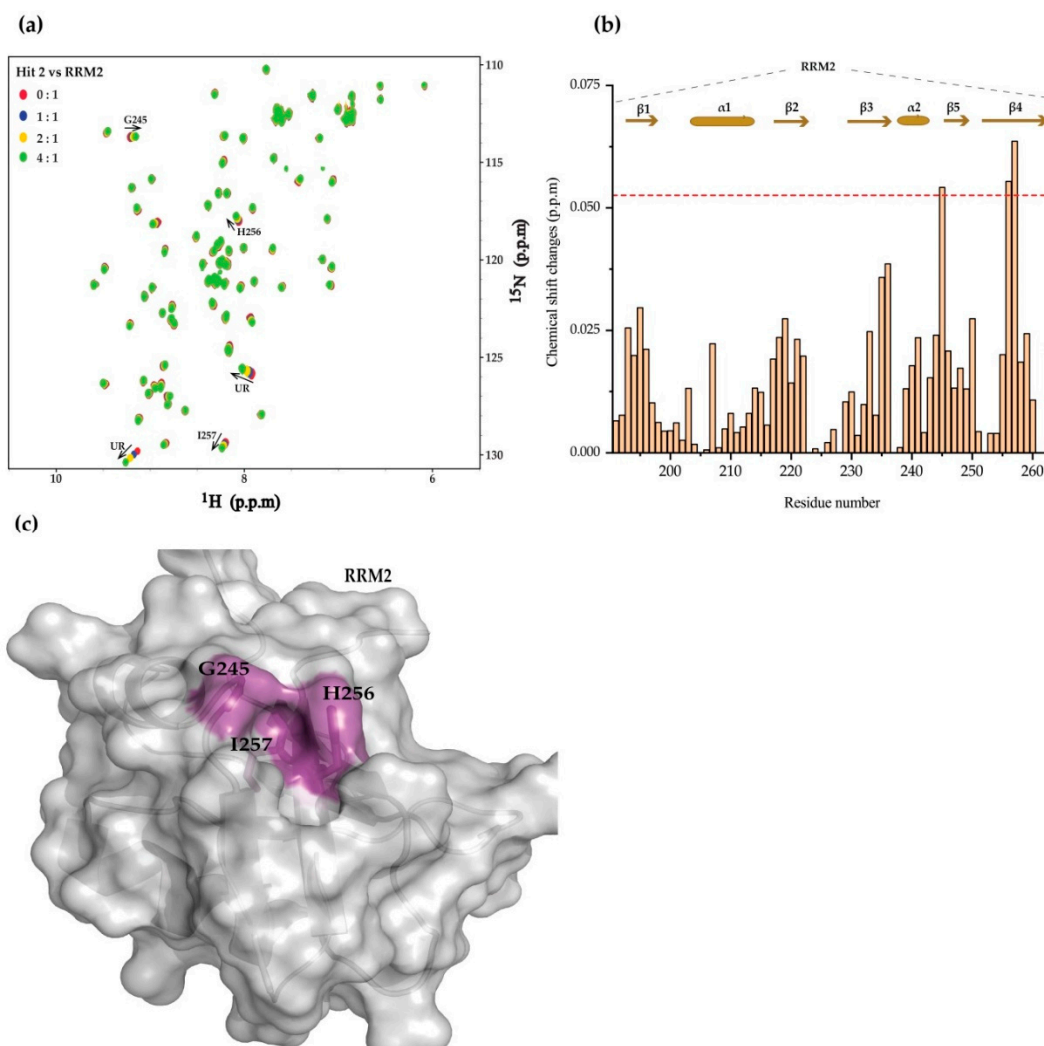


Figure 4. Chemical shift perturbations of the TDP-43-RRM2 domain upon hit 2 titration. (a) The chemical shift perturbations of the RRM2 domain of TDP-43 by hit 2 titration. (b) Chemical shift changes of the TDP-43 RRM2 domain residues at a hit 2; protein molar ratio of 4:1. The red dashed line represents two standard deviations above the averaged chemical shift changes of residues. (c) Residues (colored in purple) undergo significant chemical shift changes and are mapped on the surface representation of TDP-43-RRM2 domain (PDB code: 1WF0).

We further compared the small molecule binding topology with the nucleic acid recognition sites of the TDP-43 RRM domain. In TDP-43 tandem RRMs, 10 out of 12 nucleotides of the AUG12 RNA (GUGUGAAUGAAU) interact with RRM1 and RRM2 (PDB code: 4BS2) [47]. Among them, the first five (G₁U₂G₃U₄G₅) nucleotides are accommodated on the RRM1 β -sheet and the following two nucleotides (A₆A₇) act as a connector between two RRMs, while the next three nucleotides (U₈G₉A₁₀) lie on the RRM2. The U₈ nucleotide of RNA is recognized on S258 (β 4) through hydrogen bonds, on the backbone carbonyl oxygen of N259 (β 4), and the backbone amide of E261 from the C-terminus [47]. Comparatively, all three hits have perturbed some residues located on the β 4-strand, hits 1 and 3 specifically disturbed S258 (β 4). This also interacts with the U₈ nucleotide in tandem RRM (Figure 5a). Furthermore, the RRM2 residues D247 (loop α 2- β 3') and I249 (β 3') are involved in inter-RRM interactions upon RNA binding on the tandem RRM of TDP-43. This study revealed that their nearby residues, G245 and E246 (loop α 2- β 3'), display higher chemical shift perturbations induced by the hits binding (Figures 2b and 3c,d).

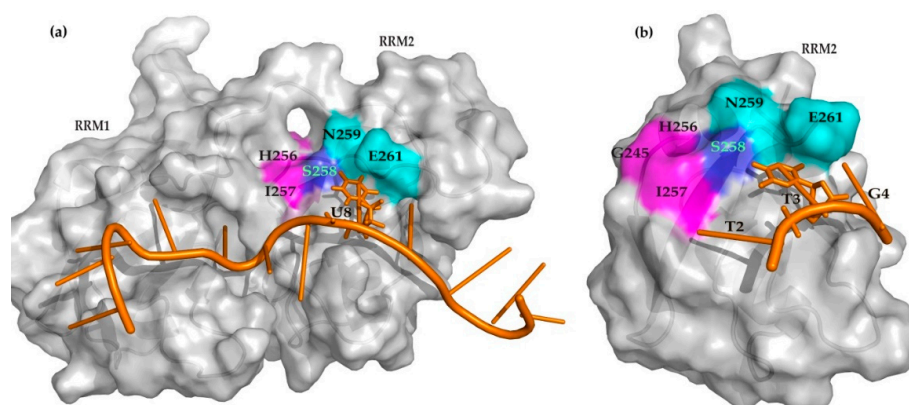


Figure 5. Comparison of binding sites of nucleic acids and hits on TDP-43. **(a)** Surface representation of TDP-43 tandem RRMs in complex with AUG12 RNA (orange cartoon), where residues interact with the U₈ nucleotide (stick) and hits are highlighted in cyan and magenta, respectively. Residue S258 (blue) interacts with both U₈ and hit 1. **(b)** Surface representation of TDP-43 RRM2 in complex with ssDNA (PDB code: 3D2W) using the same coloring scheme.

Accordingly, the crystal structure of TDP-43 RRM2 in complex with ss-DNA 5'-GTTGAGCGTT-3' (PDB entry: 3D2W) reveals that only three 5' end nucleotides (T2, T3, G4) make extensive contacts with β -sheet residues of RRM2, whereby T3 particularly contacts with S258, Asn259, and Glu261 through hydrogen bonds [49], while in our study the residues H256 and I257, nearby the S258 (β 4), have been perturbed upon hit binding on the single RRM2 (Figure 5b). This suggests that the fragment screening hits bind to a proximal site for RNA/DNA recognition, thus new hits can be designed using a fragment grow strategy to block the DNA/RNA recognition capability of TDP-43 RRM2.

To further characterize the binding mode, a data-driven approach, HADDOCK [54], was used to model the tandem RRM-hit 1 complex structure. Residues G245, E246, H256, I257, and S258 were defined as active ones in the binding site. Among the docking poses generated by HADDOCK, the best-fit ones were filtered out based on CSP and STD restraints [41,55,56]. One representative docking pose (Figure 6) indicates that hit 1 forms a hydrogen bond with the side chain of S258 and the aromatic ring of hit 1 is proximal to residues G245, E246, H256, and I257. These docking poses pave the path for following structure-guided hit-to-lead evolution.

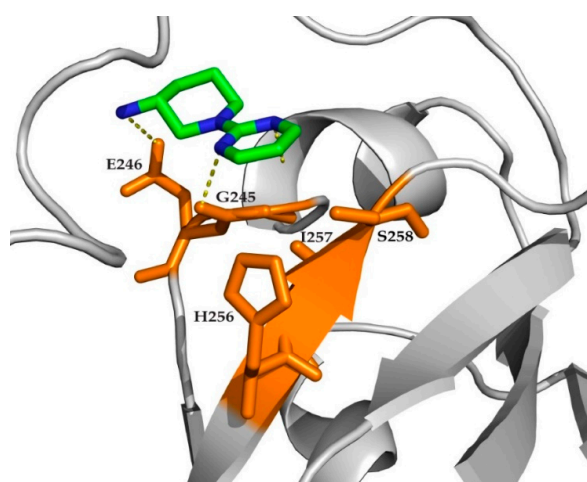


Figure 6. The representative docking model of hit 1 in consistency with experimental CSP and STD restraints. Hit 1 (green color) in the binding site of tandem RRM (PDB: 4bs2) where the carbonyl hydrogen is oriented toward G245, while the side chain hydrogen interacts with E246 residue of tandem RRM. Other active residues (orange sticks), H256, I257, and S258 are located in proximal of the hit 1.

3. Discussion

Proteins containing RRM domains function in important aspects of the posttranscriptional regulation of gene expression, mRNA maturation, and other RNA processing machinery. These proteins perform their diverse roles depending on the dual ability to recognize RNA and to interact with other proteins by using their RRM domain [31]. As TDP-43 is closely correlated with neurodegenerative and cancerous diseases [29,57], the RRM domain of TDP-43 becomes an attractive therapeutic target. However, there is no direct inhibitor targeting the RRM discovered to date.

We uncovered three small molecules binding to the tandem RRM domain of TDP-43 by using NMR fragment-based screening techniques. The NMR spectroscopy, one of a plethora of biophysical methods, is particularly powerful to detect even ultra-weak protein-ligand interactions. Accordingly, chemical shift perturbations observed in the heteronuclear single-quantum coherence (HSQC) spectra or the linewidth analysis of the small molecules allow the determination of binding affinity [58,59]. This is sometimes recalcitrant, as the titration to saturation point may be infeasible in case of weak binding affinities and low aqueous solubility of compounds.

NMR is extensively applied in fragment-based lead discovery [60]. The central idea is to screen a small library (500–2000 molecules) of low-molecular-weight compounds (110–250 Da), as their low complexity enhances the probability of matched interactions between the target and these fragment compounds. The reasonable hit rate indicates the druggability of the TDP-43 tandem RRM domain.

Although the 4-aminoquinolines molecules have been discovered through high throughput screening against the full-length TDP-43 [34], the enlightenment on binding site is still lacking. TDP-43 contains two RNA-binding RRM domains and the C-terminal low complexity domain, which may form liquid–liquid phase separation as a reservoir of mRNAs. Here, it is essential to determine the small molecule binding topology on TDP-43. The tandem RRM of TDP-43 is composed of a canonical RRM arrangement ($\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$), with an additional β -hairpin ($\beta 3'\beta 3''$ or $\beta 5$) found between $\alpha 2$ and $\beta 4$ which extends the β -sheet surface for RNA recognition [2,47,49]. The binding topology of our fragment screening hits and CSP-guided HADDOCK modeling reveal a ligand-binding “hot spot” of TDP-43 RRM2, proximal to H256, I257, and S258. Interestingly, these residues are also close to the RRM1 and RRM2 interface. The previous study proposed that both RRM domains are indispensable for achieving the greater binding affinity between the TDP-43 and nucleic acids [49]. Since this “hot spot” is partially overlapped with the RNA/DNA recognition site, it directs the following structure-guided hit-to-lead evolution against TDP-43 tandem RRM domains.

4. Materials and Methods

4.1. Cloning, Expression, and Protein Purification

The tandem RRM domain of TDP-43 (residues 101–269) was synthesized by GENEWIZ (Suzhou, China) and sub-cloned into the pET22b vector (GE Healthcare, Shanghai, China) with the His₆ tag. The RRM2 domain was amplified from the tandem RRM construct and then sub-cloned into the pET22b vector (GE Healthcare, Shanghai, China) with the His₆ tag. The constructs were transformed into *Escherichia coli* BL21 and cultivated in 1 L LB media, incubated at 37 °C. The proteins were expressed at 16 °C after induction by 0.5 mM isopropyl β -D-thiogalactosidase (IPTG) for 20 h. The bacteria were harvested by centrifugation (5000 rpm, 10 min), resuspended in lysis buffer (25 mM Tris, 500 mM NaCl at pH 7.5), and then lysed by sonication. The cell lysates were centrifuged (13,000 rpm, 30 min). The collected supernatant was purified on a column filled with Nickel-chelated resin (QIAGEN, Shanghai, China). The impurities were washed out using a buffer (25 mM Tris, 1 M NaCl at pH 7.5) containing a linear gradient of 20–40 mM imidazole, then the same buffer containing 500 mM imidazole was used to elute out the target proteins. All proteins were further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 75 column (GE Healthcare, Shanghai, China). The target proteins were confirmed by SDS-PAGE.

For ^{15}N -labeled proteins, the cells were first cultured in 1 L LB media, harvested when A_{600} reached 1.0 and then transferred to 1 L M9 media containing $^{15}\text{NH}_4\text{Cl}$. The cells were induced by 0.4 mM IPTG to express the proteins (tandem RRM and RRM2 domains). The purified proteins were concentrated in PBS buffer plus the 5 mM DTT at pH 7.5.

4.2. NMR Fragment-Based Screening

All NMR fragment screening experiments were carried out at 25 °C using an Agilent 700 MHz spectrometer equipped with a 96 well auto-sampler and a 5 mm cryoprobe. During the primary screening, the ligand-based NMR spectra (STD, WaterLOGSY, CPMG, and 1D ^1H) were acquired against the 890 fragments library (ChemBridge, San Diego, CA, USA) as described previously in detail [46,61]. Those fragments were distributed in 89 cocktails, composed of 10 compounds each, at a final concentration of 0.4 mM. These cocktails were incubated with protein (10 μM) in sodium phosphate (50 mM, pH 7.5), NaCl (200 mM), dithiothreitol (5 mM), and D_2O (50%). To further confirm the identified primary hits, secondary screening was individually carried out for single hits using the same buffer and NMR experimental settings. We then automatically processed and visualized the primary and secondary data with our ACD/Labs scripts, as previously described [46].

4.3. NMR Chemical Shift Perturbation

NMR HSQC spectra were acquired at 25 °C on either an Agilent 700MHz spectrometer equipped with a cryoprobe or an Agilent 500MHz spectrometer equipped with a room temperature probe. The ^{15}N -labeled proteins (0.1 mM or 0.2 mM), in PBS buffer (50 mM, pH 7.5), containing NaCl (200 mM), dithiothreitol (5 mM), and D_2O (10%) were titrated by small molecules stocked in DMSO at a concentration of 200 mM, using a series of hit/protein molar ratios of 0.0, 0.5, 1.0, 2.0, 4.0, and 8.0 for TDP-43 tandem RRM and 0.0, 0.5, 1.0, 2.0, and 4.0 for RRM2, respectively. Spectra were processed in NMRpipe and analyzed with Sparky. The chemical shift changes ($\Delta\delta$) relative to the free form of protein were defined as follows:

$$\Delta\delta = \sqrt{(\delta_{1\text{H}})^2 + (0.2\delta_{15\text{N}})^2}, \quad (1)$$

where $\delta_{1\text{H}}$ and $\delta_{15\text{N}}$ are the chemical shift differences of the ^1H and ^{15}N dimensions, respectively. We referred to the following chemical shift assignments previously deposited in the Biological Magnetic Resonance Data Bank: RRM1 (BMRB Entry 18765), RRM2 (BMRB Entry 19922), and tandem RRM (BMRB Entry 19290). All structures figures were prepared by Pymol (DeLano Scientific, LLC, Palo Alto, CA, USA).

4.4. Molecular Docking

HADDOCK is an information-driven docking technique used for modeling biomolecule structures by using experimental or predictive restraints [54,62]. The CSPs, obtained from the NMR HSQC titration data, were used both as HADDOCK restraints and for defining the protein active residues. The tandem RRM structure (PDB: 4bs2) served as the starting structure, while the hit 1 PDB file was generated by the PRODRG [63]. The docking calculations were done by the HADDOCK web server and clustered 186 structures in 16 clusters according to the RMSD threshold of 2 Å.

4.5. Linewidth Analysis

The NMR HSQC spectra at molar ratios of 0:1 and 8:1 (hit/protein) were processed using the same NMRpipe script, e.g., 2-fold zero-filling, Fourier transformation, and phase corrections. The spectra were then analyzed, with randomly selected peaks, using Sparky. After peak integration, the linewidth, i.e., the full width at half the peak height, was automatically estimated by Sparky.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/13/3230/s1>.

Author Contributions: Conceptualization and formal analysis, G.N. and K.R.; resources, J.G., Y.L., M.L., S.A.M., R.M., F.L., J.Z., J.W., and Y.S.; writing-original draft, G.N.; writing-review and editing, G.N. and K.R.; supervision, Y.S., and K.R.; funding acquisition, J.W., Y.S., and K.R.

Funding: We thank the financial support by grants from the Ministry of Science and Technology of China (2016YFA0500700), Strategic Priority Research Program of the Chinese Academy of Sciences (XDA12020355 and XDPB10), National Natural Science Foundation of China (21703254, 21807095, 21874123 and U1632153), and Fundamental Research Funds for the Central Universities (WK2070080002 and WK2060190086).

Acknowledgments: Part of our NMR work was performed at the National Facility for Protein Sciences Shanghai and the High Magnetic Field Laboratory, Chinese Academy of Sciences. We thank Jiuyang Liu and Na Wang for their technical and experimental support. We also thank the University of Science and Technology of China which hosted Gilbert Nshogoza as a Ph.D. student financially supported by the Chinese Scholarship Council.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ALS	Amyotrophic lateral sclerosis
ATG7	Autophagy-related protein 7
BMRB	Biological Magnetic Resonance Data Bank
CPMG	Carr–Purcell–Meiboom–Gill
CSPs	Chemical shift perturbations
FBS	Fragment-based screening
FTLD	Frontotemporal lobar degeneration
HADDOCK	High Ambiguity Driven biomolecular DOCKing
HDAC6	Histone deacetylase 6
hNFL	human low molecular weight neurofilament
HSQC	Heteronuclear single-quantum coherence
ITC	Isothermal titration calorimetry
NMR	Nuclear Magnetic Resonance
RNP	Ribonucleoproteins
RRM	RNA Recognition Motifs
SGs	Stress granules
SNHG12	Small nucleolar RNA host gene 12
SPR	Surface Plasmon Resonance
STD	Saturation Transfer difference
TDP-43	Transactive response DNA-binding Protein 43
TNBC	Triple-negative breast cancer
TRIM16	Tripartite motif-containing protein 16
WaterLOGSY	Water ligand observed via gradient spectroscopy

References

- Venter, J.C. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [[CrossRef](#)] [[PubMed](#)]
- Afroz, T.; Cienikova, Z.; Clery, A.; Allain, F.H. One, Two, Three, Four! How Multiple RRM's Read the Genome Sequence. *Methods Enzym.* **2015**, *558*, 235–278. [[CrossRef](#)]
- Clery, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290–298. [[CrossRef](#)] [[PubMed](#)]
- Kielkopf, C.L.; Lucke, S.; Green, M.R. U2AF homology motifs: Protein recognition in the RRM world. *Genes Dev.* **2004**, *18*, 1513–1526. [[CrossRef](#)] [[PubMed](#)]
- Buratti, E.; Brindisi, A.; Giombi, M.; Tisminetzky, S.; Ayala, Y.M.; Baralle, F.E. 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 exon 9 splicing. *J. Biol. Chem.* **2005**, *280*, 37572–37584. [[CrossRef](#)] [[PubMed](#)]

6. Mercado, P.A.; Ayala, Y.M.; Romano, M.; Buratti, E.; Baralle, F.E. Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res.* **2005**, *33*, 6000–6010. [[CrossRef](#)] [[PubMed](#)]
7. Banks, G.T.; Kuta, A.; Isaacs, A.M.; Fisher, E.M. TDP-43 is a culprit in human neurodegeneration, and not just an innocent bystander. *Mamm. Genome* **2008**, *19*, 299–305. [[CrossRef](#)]
8. Strong, M.J.; Volkening, K.; Hammond, R.; Yang, W.C.; Strong, W.; Leystra-Lantz, C.; Shoesmith, C. TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol. Cell. Neurosci.* **2007**, *35*, 320–327. [[CrossRef](#)]
9. Ayala, Y.M.; Misteli, T.; Baralle, F.E. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3785–3789. [[CrossRef](#)]
10. Weskamp, K.; Barmada, S.J. TDP43 and RNA instability in amyotrophic lateral sclerosis. *Brain Res.* **2018**, *1693*, 67–74. [[CrossRef](#)]
11. Wilson, C.A.; Dugger, B.; Dickson, D.W.; Wang, D.S. TDP-43 in aging and Alzheimer's disease. *Int. J. Clin. Exp. Pathol.* **2011**, *4*, 147–155. [[PubMed](#)]
12. Sun, M.; Yamashita, T.; Shang, J.; Liu, N.; Deguchi, K.; Liu, W.; Ikeda, Y.; Feng, J.; Abe, K. Acceleration of TDP43 and FUS/TLS protein expressions in the preconditioned hippocampus following repeated transient ischemia. *J. Neurosci. Res.* **2014**, *92*, 54–63. [[CrossRef](#)] [[PubMed](#)]
13. Josephs, K.A. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol.* **2014**, *127*, 811–824. [[CrossRef](#)] [[PubMed](#)]
14. 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. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006**, *314*, 130–133. [[CrossRef](#)] [[PubMed](#)]
15. Bentmann, E.; Neumann, M.; Tahirovic, S.; Rodde, R.; Dormann, D.; Haass, C. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* **2012**, *287*, 23079–23094. [[CrossRef](#)] [[PubMed](#)]
16. Liu-Yesucevitz, L.; Bilgutay, A.; Zhang, Y.J.; Vanderweyde, T.; Citro, A.; Mehta, T.; Zaarur, N.; McKee, A.; Bowser, R.; Sherman, M.; et al. Tar DNA binding protein-43 (TDP-43) associates with stress granules: Analysis of cultured cells and pathological brain tissue. *PLoS ONE* **2010**, *5*, e13250. [[CrossRef](#)]
17. Chen, Y.; Cohen, T.J. Aggregation of the nucleic acid-binding protein TDP-43 occurs via distinct routes that are coordinated with stress granule formation. *J. Biol. Chem.* **2019**.
18. Ash, P.E.; Zhang, Y.J.; Roberts, C.M.; Saldi, T.; Hutter, H.; Buratti, E.; Petrucelli, L.; Link, C.D. Neurotoxic effects of TDP-43 overexpression in *C. elegans*. *Hum. Mol. Genet.* **2010**, *19*, 3206–3218. [[CrossRef](#)]
19. Li, Y.; Ray, P.; Rao, E.J.; Shi, C.; Guo, W.; Chen, X.; Woodruff, E.A., 3rd; Fushimi, K.; Wu, J.Y. A *Drosophila* model for TDP-43 proteinopathy. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3169–3174. [[CrossRef](#)]
20. Voigt, A.; Herholz, D.; Fiesel, F.C.; Kaur, K.; Muller, D.; Karsten, P.; Weber, S.S.; Kahle, P.J.; Marquardt, T.; Schulz, J.B. TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS ONE* **2010**, *5*, e12247. [[CrossRef](#)]
21. Bhandare, V.V.; Ramaswamy, A. The proteinopathy of D169G and K263E mutants at the RNA Recognition Motif (RRM) domain of tar DNA-binding protein (tdp43) causing neurological disorders: A computational study. *J. Biomol. Struct. Dyn.* **2018**, *36*, 1075–1093. [[CrossRef](#)] [[PubMed](#)]
22. Fang, H.Y.; Chen, S.B.; Guo, D.J.; Pan, S.Y.; Yu, Z.L. Proteomic identification of differentially expressed proteins in curcumin-treated MCF-7 cells. *Phytomedicine* **2011**, *18*, 697–703. [[CrossRef](#)] [[PubMed](#)]
23. Kim, P.Y.; Tan, O.; Liu, B.; Trahair, T.; Liu, T.; Haber, M.; Norris, M.D.; Marshall, G.M.; Cheung, B.B. High TDP43 expression is required for TRIM16-induced inhibition of cancer cell growth and correlated with good prognosis of neuroblastoma and breast cancer patients. *Cancer Lett.* **2016**, *374*, 315–323. [[CrossRef](#)] [[PubMed](#)]
24. Nan, Y.N.; Zhu, J.Y.; Tan, Y.; Zhang, Q.; Jia, W.; Hua, Q. Staurosporine induced apoptosis rapidly downregulates TDP-43 in glioma cells. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 3575–3579. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, Y.L.; Li, Y.D.; Guo, F.J. Expression of TDP43 in cervical cancer. *Int. J. Clin. Exp. Pathol.* **2016**, *9*, 1467–1473.

26. Zeng, Q.; Cao, K.; Liu, R.; Huang, J.; Xia, K.; Tang, J.; Chen, X.; Zhou, M.; Xie, H.; Zhou, J. Identification of TDP-43 as an oncogene in melanoma and its function during melanoma pathogenesis. *Cancer Biol.* **2017**, *18*, 8–15. [[CrossRef](#)] [[PubMed](#)]
27. Liu, X.; Zheng, J.; Xue, Y.; Qu, C.; Chen, J.; Wang, Z.; Li, Z.; Zhang, L.; Liu, Y. Inhibition of TDP43-Mediated SNHG12-miR-195-SOX5 Feedback Loop Impeded Malignant Biological Behaviors of Glioma Cells. *Mol. Nucleic Acids* **2018**, *10*, 142–158. [[CrossRef](#)]
28. Guo, F.; Jiao, F.; Song, Z.; Li, S.; Liu, B.; Yang, H.; Zhou, Q.; Li, Z. Regulation of MALAT1 expression by TDP43 controls the migration and invasion of non-small cell lung cancer cells in vitro. *Biochem. Biophys. Res. Commun.* **2015**, *465*, 293–298. [[CrossRef](#)]
29. Ke, H.; Zhao, L.; Zhang, H.; Feng, X.; Xu, H.; Hao, J.; Wang, S.; Yang, Q.; Zou, L.; Su, X.; et al. Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with SRSF3. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3426–E3435. [[CrossRef](#)]
30. Boxer, A.L.; Gold, M.; Huey, E.; Gao, F.B.; Burton, E.A.; Chow, T.; Kao, A.; Leavitt, B.R.; Lamb, B.; Grether, M.; et al. Frontotemporal degeneration, the next therapeutic frontier: Molecules and animal models for frontotemporal degeneration drug development. *Alzheimers Dement.* **2013**, *9*, 176–188. [[CrossRef](#)]
31. Varani, G.; Nagai, K. RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 407–445. [[CrossRef](#)] [[PubMed](#)]
32. Mei, H.Y.; Cui, M.; Heldsinger, A.; Lemrow, S.M.; Loo, J.A.; Sannes-Lowery, K.A.; Sharmeen, L.; Czarnik, A.W. Inhibitors of protein-RNA complexation that target the RNA: Specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. *Biochem.-Us* **1998**, *37*, 14204–14212. [[CrossRef](#)] [[PubMed](#)]
33. Gayle, A.Y.; Baranger, A.M. Inhibition of the U1A–RNA Complex by an Aminoacridine Derivative. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2839–2842. [[CrossRef](#)]
34. Cassel, J.A.; Blass, B.E.; Reitz, A.B.; Pawlyk, A.C. Development of a novel nonradiometric assay for nucleic acid binding to TDP-43 suitable for high-throughput screening using AlphaScreen technology. *J. Biomol. Screen.* **2010**, *15*, 1099–1106. [[CrossRef](#)] [[PubMed](#)]
35. Cassel, J.A.; McDonnell, M.E.; Velvadapu, V.; Andrianov, V.; Reitz, A.B. Characterization of a series of 4-aminoquinolines that stimulate caspase-7 mediated cleavage of TDP-43 and inhibit its function. *Biochimie* **2012**, *94*, 1974–1981. [[CrossRef](#)] [[PubMed](#)]
36. Vidal, R.L.; Matus, S.; Bargsted, L.; Hetz, C. Targeting autophagy in neurodegenerative diseases. *Trends Pharm. Sci* **2014**, *35*, 583–591. [[CrossRef](#)] [[PubMed](#)]
37. Hajduk, P.J.; Bures, M.; Praestgaard, J.; Fesik, S.W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem* **2000**, *43*, 3443–3447. [[CrossRef](#)] [[PubMed](#)]
38. Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. NMR-based screening in drug discovery. *Q Rev. Biophys.* **1999**, *32*, 211–240. [[CrossRef](#)]
39. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534. [[CrossRef](#)]
40. Fesik, S.W. NMR structure-based drug design. *J. Biomol. NMR* **1993**, *3*, 261–269. [[CrossRef](#)]
41. Liu, J.; Gao, J.; Li, F.; Ma, R.; Wei, Q.; Wang, A.; Wu, J.; Ruan, K. NMR characterization of weak interactions between RhoGDI2 and fragment screening hits. *Biochim. Et Biophys. Acta (Bba)-Gen. Subj.* **2017**, *1861*, 3061–3070. [[CrossRef](#)] [[PubMed](#)]
42. Liu, J.; Li, F.; Bao, H.; Jiang, Y.; Zhang, S.; Ma, R.; Gao, J.; Wu, J.; Ruan, K. The polar warhead of a TRIM24 bromodomain inhibitor rearranges a water-mediated interaction network. *FEBS J.* **2017**, *284*, 1082–1095. [[CrossRef](#)] [[PubMed](#)]
43. Liu, J.; Zhang, S.; Liu, M.; Liu, Y.; Nshogoza, G.; Gao, J.; Ma, R.; Yang, Y.; Wu, J.; Zhang, J.; et al. Structural plasticity of the TDRD3 Tudor domain probed by a fragment screening hit. *FEBS J.* **2018**, *285*, 2091–2103. [[CrossRef](#)] [[PubMed](#)]
44. Ruan, K.; Gao, J.; Ma, R. NMR in Fragment Based Lead Discovery. *Chin. J. Magn. Reson.* **2012**, *29*, 163–181.
45. Tang, H.; Nshogoza, G.; Liu, M.; Liu, Y.; Ruan, K.; Ma, R.; Gao, J. Identification of Novel Hits of the NSD1 SET Domain by NMR Fragment-Based Screening. *Chin. J. Magn. Reson.* **2018**, *36*, 148–154. [[CrossRef](#)]
46. Gao, J.; Ma, R.; Wang, W.; Wang, N.; Sasaki, R.; Snyderman, D.; Wu, J.; Ruan, K. Automated NMR fragment based screening identified a novel interface blocker to the LARG/RhoA complex. *PLoS ONE* **2014**, *9*, e88098. [[CrossRef](#)] [[PubMed](#)]

47. Lukavsky, P.J.; Daujotyte, D.; Tollervey, J.R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F.E.; Damberger, F.F.; Allain, F.H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1443–1449. [[CrossRef](#)]
48. Kuo, P.H.; Chiang, C.H.; Wang, Y.T.; Doudeva, L.G.; Yuan, H.S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **2014**, *42*, 4712–4722. [[CrossRef](#)]
49. Kuo, P.H.; Doudeva, L.G.; Wang, Y.T.; Shen, C.K.; Yuan, H.S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **2009**, *37*, 1799–1808. [[CrossRef](#)]
50. Viegas, A.; Manso, J.o.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88*, 990–994. [[CrossRef](#)]
51. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **2001**, *21*, 349–359. [[CrossRef](#)] [[PubMed](#)]
52. Hajduk, P.J.; Olejniczak, E.T.; Fesik, S.W. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J. Am. Chem. Soc.* **1997**, *119*, 12257–12261. [[CrossRef](#)]
53. Zega, A. NMR Methods for Identification of False Positives in Biochemical Screens. *J. Med. Chem.* **2017**, *60*, 9437–9447. [[CrossRef](#)] [[PubMed](#)]
54. Van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastiris, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **2015**, *428*, 720–725. [[CrossRef](#)] [[PubMed](#)]
55. Aguirre, C.; ten Brink, T.; Cala, O.; Guichou, J.F.; Krimm, I. Protein–ligand structure guided by backbone and side-chain proton chemical shift perturbations. *J. Biomol. Nmr.* **2014**, *60*, 147–156. [[CrossRef](#)] [[PubMed](#)]
56. Aguirre, C.; ten Brink, T.; Guichou, J.F.; Cala, O.; Krimm, I. Comparing Binding Modes of Analogous Fragments Using NMR in Fragment-Based Drug Design: Application to PRDX5. *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
57. Harrison, A.F.; Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **2017**, *474*, 1417–1438. [[CrossRef](#)]
58. Shortridge, M.D.; Hage, S.H.; Harbison, G.S.; Powers, R. Estimating Protein-Ligand Binding Affinity using High-Throughput Screening by NMR. *J. Comb. Chem.* **2008**, *10*, 948–958. [[CrossRef](#)]
59. Mashalidis, E.H.; Sledz, P.; Lang, S.; Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* **2013**, *8*, 2309–2324. [[CrossRef](#)]
60. Gossert, A.D.; Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 82–125. [[CrossRef](#)]
61. Wang, N.; Li, F.; Bao, H.; Li, J.; Wu, J.; Ruan, K. NMR Fragment Screening Hit Induces Plasticity of BRD7/9 Bromodomains. *ChemBioChem* **2016**, *17*, 1456–1463. [[CrossRef](#)] [[PubMed](#)]
62. Wassenaar, T.A.; Van Dijk, M.; Loureiro-Ferreira, N.; Van Der Schot, G.; De Vries, S.J.; Schmitz, C.; Van Der Zwan, J.; Boelens, R.; Giachetti, A.; Ferella, L.; et al. WeNMR: Structural Biology on the Grid. *J. Grid Comput.* **2012**, *10*, 743–767. [[CrossRef](#)]
63. Schuttelkopf, A.W.; van Aalten, D.M. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Cryst. D Biol. Cryst.* **2004**, *60*, 1355–1363. [[CrossRef](#)] [[PubMed](#)]

