



OPEN

Tethering-induced destabilization and ATP-binding for tandem RRM domains of ALS-causing TDP-43 and hnRNPA1

Mei Dang, Yifan Li & Jianxing Song

TDP-43 and hnRNPA1 contain tandemly-tethered RNA-recognition-motif (RRM) domains, which not only functionally bind an array of nucleic acids, but also participate in aggregation/fibrillation, a pathological hallmark of various human diseases including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), alzheimer's disease (AD) and Multisystem proteinopathy (MSP). Here, by DSF, NMR and MD simulations we systematically characterized stability, ATP-binding and conformational dynamics of TDP-43 and hnRNPA1 RRM domains in both tethered and isolated forms. The results reveal three key findings: (1) upon tethering TDP-43 RRM domains become dramatically coupled and destabilized with T_m reduced to only 49 °C. (2) ATP specifically binds TDP-43 and hnRNPA1 RRM domains, in which ATP occupies the similar pockets within the conserved nucleic-acid-binding surfaces, with the affinity slightly higher to the tethered than isolated forms. (3) MD simulations indicate that the tethered RRM domains of TDP-43 and hnRNPA1 have higher conformational dynamics than the isolated forms. Two RRM domains become coupled as shown by NMR characterization and analysis of inter-domain correlation motions. The study explains the long-standing puzzle that the tethered TDP-43 RRM1–RRM2 is particularly prone to aggregation/fibrillation, and underscores the general role of ATP in inhibiting aggregation/fibrillation of RRM-containing proteins. The results also rationalize the observation that the risk of aggregation-causing diseases increases with aging.

Aggregation/fibrillation of TDP-43 in the cytoplasm of neurons is a pathological hallmark of ~ 97% amyotrophic lateral sclerosis (ALS) and ~ 45% frontotemporal dementia-TDP (FTLD-TDP), that lack any efficacious medicine so far^{1–3}. TDP-43 is a well-known member of heterogeneous nuclear ribonucleoproteins (hnRNP) containing the folded RNA-recognition motif (RRM), which constitutes one of the most abundant domains in eukaryotes^{3–5}. Interestingly, ~ 70 human RRM-containing proteins including FUS, TDP-43 and hnRNPA1 additionally have the intrinsically disordered prion-like domains of low-complexity sequences with amino acid compositions similar to those of the prion domains in yeast responsible for driving the formation of infectious conformers^{2–4}. Remarkably, these proteins not only functionally mediate direct interactions with various nucleic acids to control both RNA processing and gene expression, but their aggregation/fibrillation is pathologically characteristic of an increasing spectrum of human diseases including ALS, frontotemporal dementia (FTD), Alzheimer's disease (AD), chronic traumatic encephalopathy, muscle regeneration/degeneration and multisystem proteinopathy (MSP)^{1–10}.

414-Residue TDP-43 contains two folded RRM domains (Fig. 1A), which are not only essential for recognizing UG-rich sequences near RNA splice sites¹¹, but also extensively demonstrated to participate in disease-causing aggregation/fibrillation in addition to its C-terminal prion-like domain^{11–19}. Remarkably, small molecules targeting the RRM domains of TDP-43 have been recently identified to reduce locomotor defects in drosophila model of ALS¹⁹. We also found that ATP, the universal energy currency, specifically binds the RRM domains of FUS and TDP-43 to inhibit their amyloid fibrillation^{20–22}. Therefore to understand the factors and mechanisms governing their aggregation/fibrillation is not only of fundamental interest, but represents an essential step for further development of therapeutic strategies/molecules to treat these diseases.

Structurally, two TDP-43 RRM domains adopt the same overall fold shared by all RRM domains, which consists of a four-stranded β -sheet and two perpendicular α -helices (Fig. 1B). Recently, we showed that aggregation/fibrillation of the single FUS RRM domain not only depends on thermodynamic stability, but also on

Department of Biological Sciences, Faculty of Science, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore. email: dbssjx@nus.edu.sg

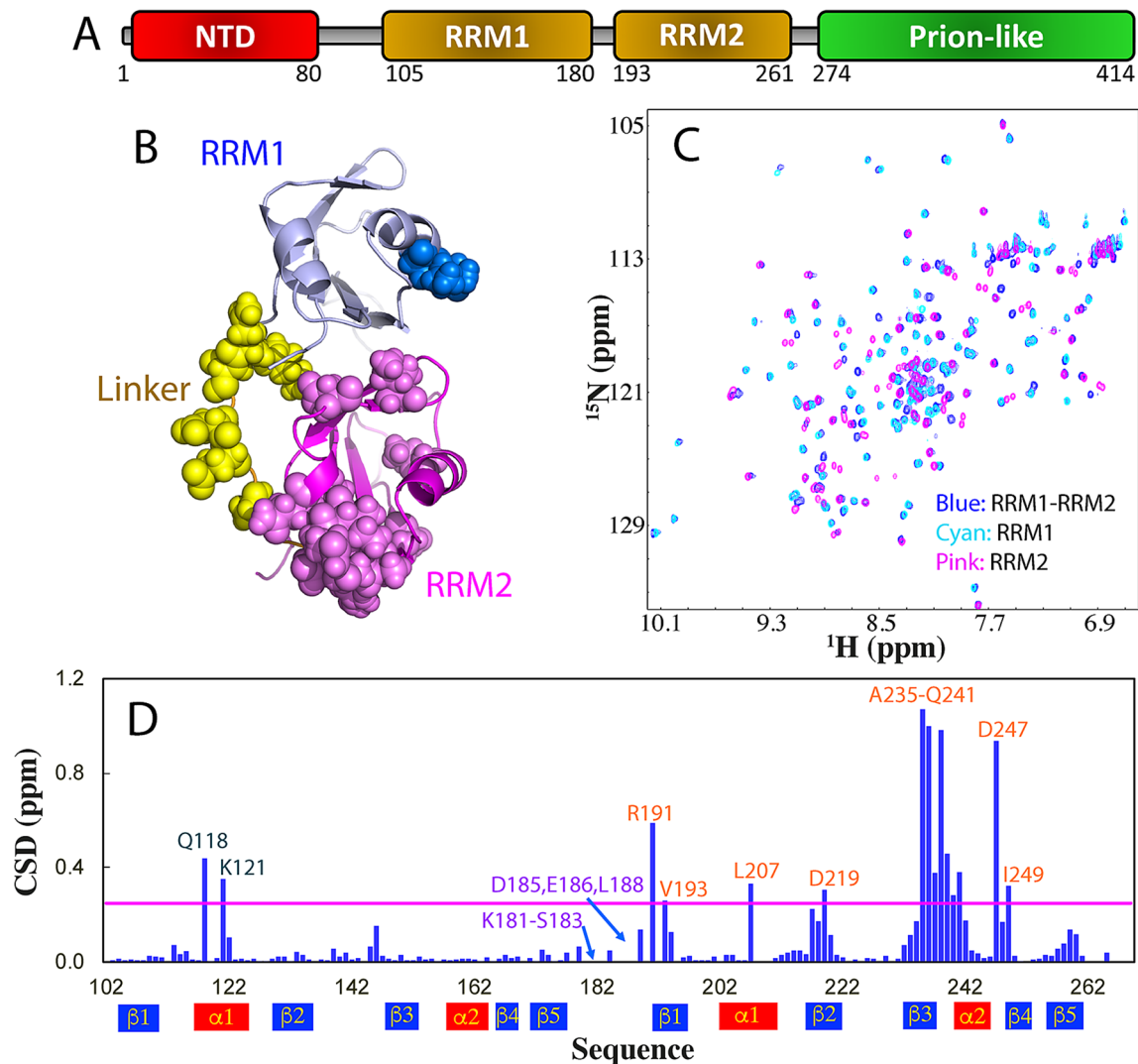


Figure 1. Dissection-induced perturbation of TDP-43 RRM domains. (A) 414-residue TDP-43 contains: N-terminal domain (NTD) over residues 1–80; two tandemly-tethered RNA recognition motifs (RRM1 and RRM2) over residues 105–261, and C-terminal prion-like domain over residues 274–414. (B) NMR structure (PDB ID of 4BS2) of the TDP-43 RRM1 (light blue) and RRM2 (pink) in which the residues are displayed in sphere for those with HSQC peaks disappeared (yellow), significantly shifted in RRM1 (blue) or in RRM2 (light pink) upon dissection. The image was prepared by PyMol2.4 (<https://pymol.org>). (C) Superimposition of ^1H - ^{15}N NMR HSQC spectra of the ^{15}N -labeled tethered RRM1–RRM2 domains (blue), the isolated RRM1 (cyan) and RRM2 (pink) proteins at 50 μM in 10 mM sodium phosphate buffer (pH 6.8) containing 150 mM NaCl and 10 mM DTT. (D) Residue-specific chemical shift difference (CSD) of the RRM1 and RRM2 domains between the tethered and isolated forms. Significantly shifted residues are labeled and displayed as spheres in (B), which are defined as those with the CSD values >0.25 (average value + one standard deviation) (pink line).

conformational dynamics²³. Briefly, the FUS RRM domain undergoes fibrillation not only due to its relatively low stability with melting temperature (T_m) of 55 $^{\circ}\text{C}$, lower than that of the human $\gamma\text{-S}$ -crystallin with T_m of 71 $^{\circ}\text{C}$ ²⁴, but also resulting from its high conformational dynamics, that allow the dynamic opening of the structure to expose its central hydrophobic regions for aggregation/fibrillation²³. Consequently, although ATP induces no enhancement of the thermal stability of the FUS RRM domain, it is sufficient to kinetically inhibit the amyloid fibrillation through weak but specific binding with a K_d of 3.8 ± 0.5 mM to a pocket within the conserved nucleic-acid-binding surfaces, which thus leads to blocking the dynamic opening of the structure¹⁹.

It has been a long-standing puzzle that the tethered RRM domains of TDP-43 were particularly prone to aggregation¹¹ while the isolated RRM1 and RRM2 domains appeared to be stable and soluble as evidenced by their NMR structures determination at high protein concentrations by RIKEN Structural Genomics/Proteomics Initiative (PDB ID of 2CQG and 1WF0). We also observed that in the buffer with pH and salt concentration close to in vivo conditions (10 mM sodium phosphate at pH 6.8 with 150 mM NaCl and 10 mM DTT), the tethered TDP-43 RRM1–RRM2 could form amyloid fibrils within several days while the isolated RRM1 and RRM2 failed even after a month²². In particular, we found that under exactly the same conditions, the tethered RRM1–RRM2 of TDP-43 has only one transition with T_m of only 49 $^{\circ}\text{C}$, lower than T_m of the isolated RRM1

(57 °C). On the other hand, the binding affinity of ATP has been characterized to be higher to the tethered RRM1 than the isolated RRM1^{21,22}.

So far, it remains completely unexplored for the relationship of thermodynamic stability, conformational dynamics and ATP-binding of RRM domains between the tethered and isolated forms. In the present study, with experimental methods including DSF and NMR as well as molecular dynamics (MD) simulations, we aimed to address this problem by characterizing the thermal stability, ATP-binding and conformational dynamics of TDP-43 and hnRNPA1 RRM domains in both tethered and isolated forms. The most unexpected finding is that upon tethering, TDP-43 RRM1 and RRM2 become highly coupled to behaving as one denaturing unit with the stability significantly reduced. By contrast, no significant destabilization was observed upon tethering of two RRM domains of hnRNPA1. Our study further showed that the tethering-induced effects mainly result from the inter-domain interactions as detected by NMR characterization and analysis of inter-domain correlation motions calculated from MD simulations. Moreover, we found that ATP can also specifically bind the hnRNPA1 RRM2 but not RRM1 domain with the affinity and complex structure highly similar to those for FUS and TDP-43 RRM domains. Intriguingly, the previous and present results together revealed that ATP bind the RRM domains of both TDP-43 and hnRNPA1 with the affinity slightly higher for the tethered than for the isolated ones. Our study provides the first mechanistic insight into the tethering-induced effects on the tandem RRM domains, and highlights the general role of ATP in inhibiting aggregation/fibrillation of RRM-containing proteins, which extensively causes various human diseases by “loss of functions” or/and “gain of function”.

Results

Dissection-induced perturbation of TDP-43 RRM1 and RRM2 domains. Due to the presence of dynamics and aggregation at high concentrations, the structure of the tethered TDP-43 RRM1–RRM2 could not be determined in the free state by NMR or X-ray crystallography^{11,25}. As such, here we used the NMR structure (PDB ID of 4BS2) in complex with RNA¹¹, whose C α and C β chemical shifts indicative of secondary structures were shown to be very similar to those in the free state²⁵. As shown in Fig. 1A, B, TDP-43 contains two RRM1 and RRM2 domains respectively over residues 105–180 and 193–261 connected by an unstructured linker over residues 181–192. Comparison of NMR structures determined in the tethered and isolated forms (2CQG for RRM1 and 1WF0 for RRM2) revealed that RRM1 has C α atom RMSD value of 1.63 Å but RRM2 only 0.82 Å, indicating that the overall structures of both RRM1 and RRM2 domains of TDP-43 are well-folded and adopt the same RRM fold in both forms. However, it is worthwhile to point out that the relative orientation of two RRM domains in the tethered form may result from the binding to RNA although it was shown that the motion of two RRM domains was not completely independent based on the backbone relaxation measurement¹¹.

Here we cloned and expressed the tethered RRM1–RRM2 of TDP-43 (102–269), as well as its isolated RRM1 (102–191) and RRM2 (191–269). Indeed as extensively observed, the tethered RRM1–RRM2 protein was prone to aggregation even at concentrations of ~ 100 μ M but the isolated RRM1 and RRM2 proteins showed no significant aggregation at concentrations of ~ 1 mM. Nevertheless, the tethered RRM1–RRM2 protein has a well-dispersed HSQC spectrum typical of a well-folded protein at 50 μ M in 10 mM sodium phosphate buffer containing 10 mM DTT and 150 mM NaCl (pH 6.8) (Fig. 1C). Two isolated RRM1 and RRM2 proteins also have well-dispersed HSQC spectra at the same protein concentration in the same buffer. However, upon superimposing three HSQC spectra, some significant changes were identified (Fig. 1C and S1): (1) HSQC peaks of residues K181–S183, D185, E186 and L188 over the linker, which were undetectable in the tethered form became detectable in the isolated form; (2) upon dissection some HSQC peaks were largely shifted and the residues with significant chemical shift difference (CSD) are mainly located within the RRM2 domain (Fig. 1B, D). This observation strongly suggests that in the tethered form, two RRM domains of TDP-43 have dynamic inter-domain interactions. Consequently although the dissection resulted in no disruption of the overall RRM fold for both RRM1 and RRM2, it did lead to the changes of local conformations, or/and dynamics or/and chemical environments of a set of RRM residues, thus leading to significant shifts of their HSQC peaks.

Interestingly, we have further acquired HSQC spectrum of the mixture of the isolated RRM1 and RRM2 at an equal molar ratio (1:1), which is very different from that the tethered RRM1–RRM2 (Fig. S2A) but almost completely superimposable to the overlay of HSQC spectra of the isolated RRM1 and RRM2 (Fig. S2B). This observation strongly suggests that the interaction of the TDP-43 RRM1 and RRM2 is very dynamic and needs the covalent connection of two RRM domains to enhance their interaction, likely by increasing the effective concentrations.

Thermal stability and ATP-binding of the isolated TDP-43 RRM2 domain. Due to the critical role of TDP-43 RRM domains in amyloid fibrillation associated with various neurodegenerative diseases, previously their thermodynamic stability has been extensively characterized by a variety of biophysical methods which monitor distinctive probes associated with the secondary and tertiary structures. Nevertheless, despite exhaustive studies, it remains challenging for understanding the relationship between structure, stability and fibrillation. For example, previous far-UV CD studies in the low salt buffer indicated that TDP-43 RRM domains showed no complete denaturation of secondary structures even at 90 °C, implying that TDP-43 RRM domains might directly assemble into soluble β -rich oligomers from partially-unfolded intermediates¹⁴.

As the high-salt buffer with 150 mM NaCl we used triggered unacceptable noise for far-UV CD spectroscopy, we therefore conducted the thermal denaturation monitored by the intrinsic Trp UV fluorescence for the tethered TDP-43 RRM1–RRM2, isolated RRM1 and RRM2 domain, as well as their mixture at 1:1 (Fig. S3). Due to the lack of Trp residue in RRM2, no Trp fluorescence could be detected. As judged from the results, it appeared that consistent with previous studies^{14,23}, the thermal denaturation is irreversible because the spectra of the samples cooled down to 25 °C after treatment at 95 °C are different from those initial spectra at 25 °C. The

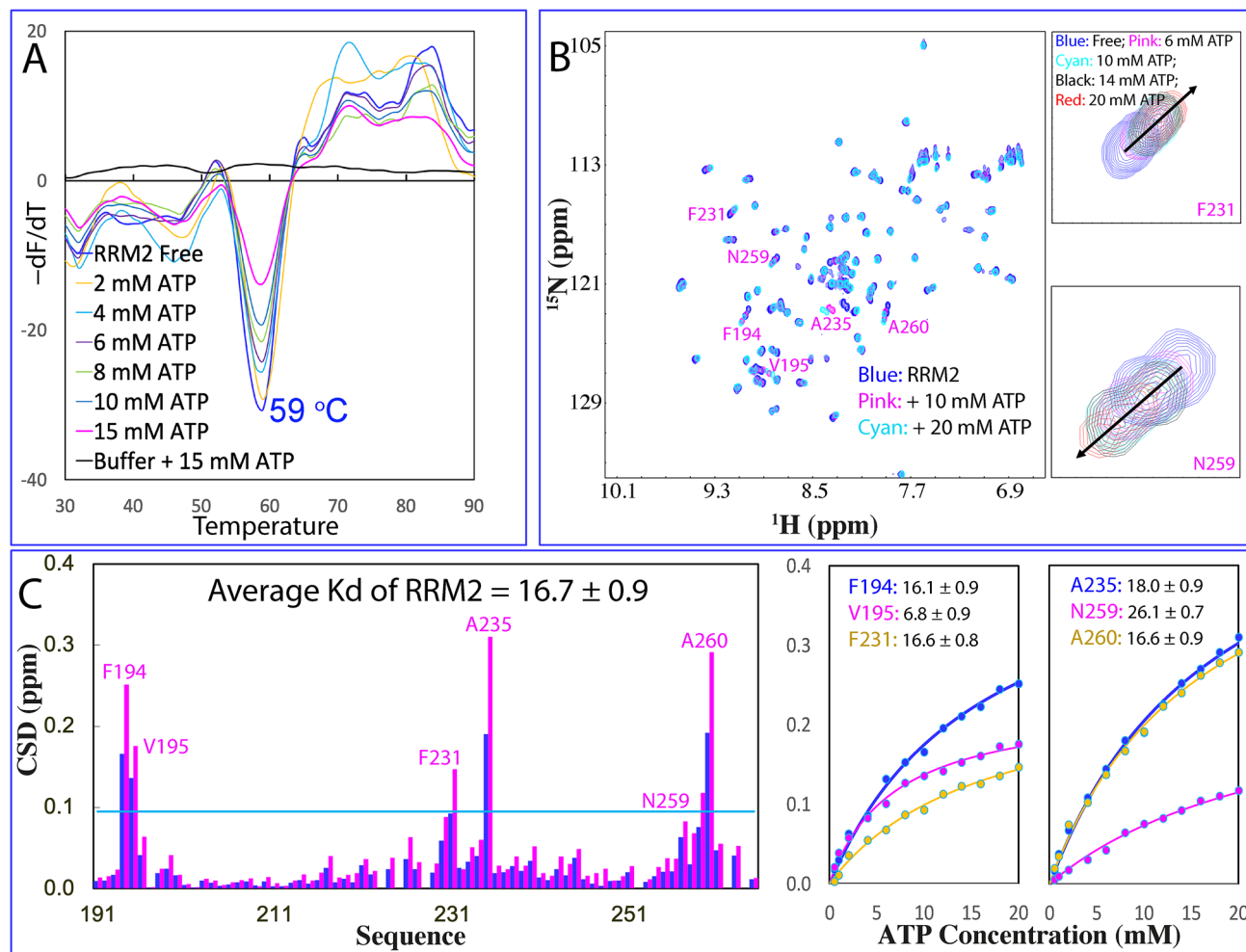


Figure 2. Thermal stability and ATP binding of the isolated TDP-43 RRM2. (A) DSF melting curves of thermal unfolding of the isolated RRM2 domain in the presence of ATP at different concentrations. (B) ^1H - ^{15}N NMR HSQC spectra of the ^{15}N -labeled RRM2 domain at 50 μM (blue) in 10 mM sodium phosphate buffer (pH 6.8) containing 150 mM NaCl and 10 mM DTT in the absence (blue), and in the presence of ATP at 10 (pink) and 20 (cyan) mM. Traces of two representative HSQC peaks, which have significant shifts but are not severely overlapped with other peaks. For clarity, only peaks at five ATP concentrations are shown: in the free state (blue); in the presence of ATP at 6 mM (pink); 10 mM (cyan); 14 mM (black); and 20 mM (red). (C) Residue-specific chemical shift difference (CSD) of the isolated RRM2 in the presence of ATP at 10 mM (blue) and 20 mM (pink). Significantly shifted residues are labeled, which are defined as those with the CSD values at 20 mM ATP > 0.1 (average value + one standard deviation) (cyan line). Fitting of the residue-specific dissociation constant (Kd): experimental (dots) and fitted (lines) values for the chemical shift differences induced by addition of ATP at 0.5, 1.0, 2.0, 4.0, 6.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0 and 20.0 mM.

melting temperature (T_m) was estimated to be 55 $^\circ\text{C}$ for TDP-43 RRM1. Noticeably, the tethered RRM1–RRM2 sample has a T_m of 45 $^\circ\text{C}$ while the mixed RRM1 and RRM2 sample shows a T_m of 55 $^\circ\text{C}$. As RRM2 has no Trp intrinsic fluorescence, this observation suggests that only upon covalent connection to RRM2, RRM1 is significantly destabilized with a T_m decrease of 10 $^\circ\text{C}$.

Previously we have characterized the thermal stability and ATP binding of the tethered RRM1–RRM2²¹ and isolated RRM1²² of TDP-43 by differential scanning fluorimetry (DSF) which reports the increase of binding of the fluorescent dye due to the exposure of hydrophobic patches in the partially-unfolded intermediates which might directly assemble into β -rich oligomers. Interestingly the tethered RRM1–RRM2 has only one thermal unfolding transition with T_m of 49 $^\circ\text{C}$, which increased to 54 $^\circ\text{C}$ with addition of ATP. The isolated RRM1 also has only one thermal unfolding transition but with T_m of 57 $^\circ\text{C}$, which increased to 60 $^\circ\text{C}$ with addition of ATP. To understand this unexpected observation, here we measured the thermal stability of the isolated RRM2 under the same conditions. Interestingly, the isolated RRM2 has also only one thermal unfolding transition with T_m of 59 $^\circ\text{C}$, which is not affected by the addition of ATP even up to 15 mM (Fig. 2A). This result suggests that upon tethering, the RRM1 and RRM2 domains of TDP-43 became significantly coupled to acting as one unfolding unit as well as thermodynamically destabilized. Strikingly, the results by intrinsic Trp fluorescence and DSF are in general consistent although they monitor fundamentally different probes.

On the other hand, in the tethered form, the K_d value of RRM1 binding to ATP is 2.6 ± 0.3 mM²¹ while the K_d of the isolated RRM1 is 3.9 ± 0.8 mM²². Although in the tethered form, RRM2 also has an ATP-binding pocket but with much lower affinity than RRM1, with K_d of 13.9 ± 0.9 mM. Here our NMR titrations showed that isolated RRM2 is also able to bind ATP to induce large shifts of a set of HSQC peaks (Fig. 2B) with the overall pattern of the perturbed residues highly similar to that of in the tethered form but with slightly larger K_d (16.7 ± 0.9 mM) (Fig. 2C). It is worthwhile to note that ATP binds to both RRM1 and RRM2 domains of TDP-43 with the slightly higher affinity in the tethered form than those in the isolated form.

Dissection-induced perturbation of hnRNPA1 RRM1 and RRM2 domains. To understand whether the destabilization observed on the tethered TDP-43 RRM1 and RRM2 domains is unique to TDP-43 or also applicable to other RRM-containing proteins. Here we decided to further characterize the tethered and isolated tandem RRM domains of 320-residue hnRNPA1, which contains two RRM domains over residues 15–90, 106–179 respectively (Fig. 3A). Previously, the NMR structures of the tethered RRM1–RRM2²⁶ as well as isolated RRM1 and RRM2 domains have been determined by NMR²⁷. The NMR results indicate that very different from what was observed in the crystal structure, the relative orientation of two RRM domains in the free state in fact already resembles to that in complex with nucleic acids²⁶. The RRM1 domain has C α atom RMSD value of only 0.68 Å while the RRM2 domain has RMSD of 0.62 Å between the tethered and isolated forms, implying the RRM1 and RRM2 domains adopt the same fold regardless of being tethered or isolated. Interestingly, the linker for the hnRNPA1 RRM1 and RRM2 domains is not completely unstructured as that of TDP-43 but has a short helix over residues 91–96 (Fig. 3B).

Here we cloned and expressed the tethered hnRNPA1 RRM1–RRM2 (5–184), as well as its isolated RRM1 (5–95) and RRM2 (94–184). Different from what was observed on the TDP-43 RRM domains, the tethered RRM1–RRM2 is highly soluble and has a well-dispersed HSQC spectrum typical of a well-folded protein at 50 μ M (Fig. 3C), while two isolated RRM1 and RRM2 proteins are also highly soluble and have well-dispersed HSQC spectra at the same protein concentration in the same buffer conditions. Upon superimposing three HSQC spectra, some significant shifts of HSQC peaks were observed (Fig. 3C,D). Detailed analysis revealed (Fig. S4): (1) CSD values of HSQC peaks upon dissection of hnRNPA1 RRM1–RRM2 (with the largest < 0.3) are much less than those (with the largest close to 1.2 ppm) upon dissection of TDP-43 RRM1–RRM2 (Fig. 1D). (2) residues with large shifts are located on RRM1, linker and RRM2 (Fig. 3B). This observation suggests that in the tethered form, two RRM domains of hnRNPA1 also have some dynamic inter-domain interactions to some degree. On the other hand, we have also acquired HSQC spectrum of the mixture of the hnRNPA1 isolated RRM1 and RRM2 at 1:1, which is also different from that the tethered RRM1–RRM2 (Fig. S5A) but almost completely superimposable to the overlay of HSQC spectra of the isolated RRM1 and RRM2 (Fig. S5B). This observation indicates that similar to what was observed on TDP-43 RRM1 and RRM2 domains (Fig. S2), the interaction of the hnRNPA1 RRM1 and RRM2 is also very dynamic and the covalent connection of both RRM domains is needed to increase their effective concentrations.

Thermal stability and ATP-binding of the tethered and isolated hnRNPA1 RRM domains. We also carried out the thermal denaturation monitored by the intrinsic Trp UV fluorescence for the tethered hnRNPA1 RRM1–RRM2, isolated RRM1 and RRM2 domain, as well as their mixture at 1:1 (Fig. S6). Due to the lack of Trp residue in RRM2, no Trp fluorescence could be detected. T_m was estimated to be 58 °C for RRM1, while the tethered RRM1–RRM2 as well as the mixed RRM1 and RRM2 samples also show similar T_m values of ~ 58 °C. This observation suggests that different from what was observed on TDP-43 RRM domains (Fig. S3), the covalent connection of hnRNPA1 RRM2 to RRM1 has no significantly destabilization of RRM1.

We further characterized the thermal stability and ATP binding of the tethered and isolated RRM domains of hnRNPA1. The tethered RRM1–RRM2 has only one thermal unfolding transition with T_m of 55 °C, which increased to 58 °C with addition of ATP (Fig. 4A). Intriguingly, the isolated RRM1 without ATP has two thermal unfolding transitions with T_m of 51 and 57 °C respectively. With addition of ATP up to 15 mM, the transition at 51 °C disappeared and only the transition at 57 °C retained. This implies that ATP has a capacity in shifting the equilibrium of different conformations, as we recently observed that ATP could enhance the stability without detectable binding by NMR on the ALS-causing C71G mutant of profilin-1 most likely by dynamically interacting with the exposed hydrophobic patches or/and even mediating the hydration shell of proteins²⁸. On the other hand, the isolated RRM2 has one thermal unfolding transition with T_m of 60 °C, which increased to 63 °C with the addition of ATP. The results together suggest that the tethering of hnRNPA1 RRM1 and RRM2 domain led to no significant destabilization.

We further characterized the binding of ATP to the tethered and isolated RRM domains. As shown in Fig. 4B, ATP induced large shifts of many HSQC peaks of the tethered RRM1–RRM2, and detailed analysis revealed that the residues with significant shifts are located on the RRM2 domain except for Arg88, Val90 and Ser91 within RRM1 (Fig. 5A). This was further confirmed by the ATP titrations on the isolated RRM1 and RRM2 domains (Fig. 4B). ATP even with concentrations up to 20 mM only triggered large shifts of two residues Val90 and Ser91 of RRM1 but induced significant shifts of a large set of peaks of the isolated RRM2 domain. Furthermore, the overall patterns of shifted residues of RRM2 are very similar in both tethered and isolated forms (Fig. 5A, B).

With the same method we previously used to characterize the ATP-binding to the FUS and TDP-43 RRM domains^{20–22,29}, we determined K_d value of the ATP binding to hnRNPA1 RRM2 to be 4.9 and 7.7 mM respectively for the tethered and isolated forms (Fig. 5A, B). The values are very similar to those for FUS RRM, TDP-43 RRM1 as well as for a non-canonical helix-only RNA-binding domain³⁰ of hnRNPA1 Q (K_d of 3.1 mM). Interestingly, as observed on TDP-43 RRM domains, the ATP binding affinity to the RRM domains of hnRNPA1 is also higher for the tethered form than for the isolated form.

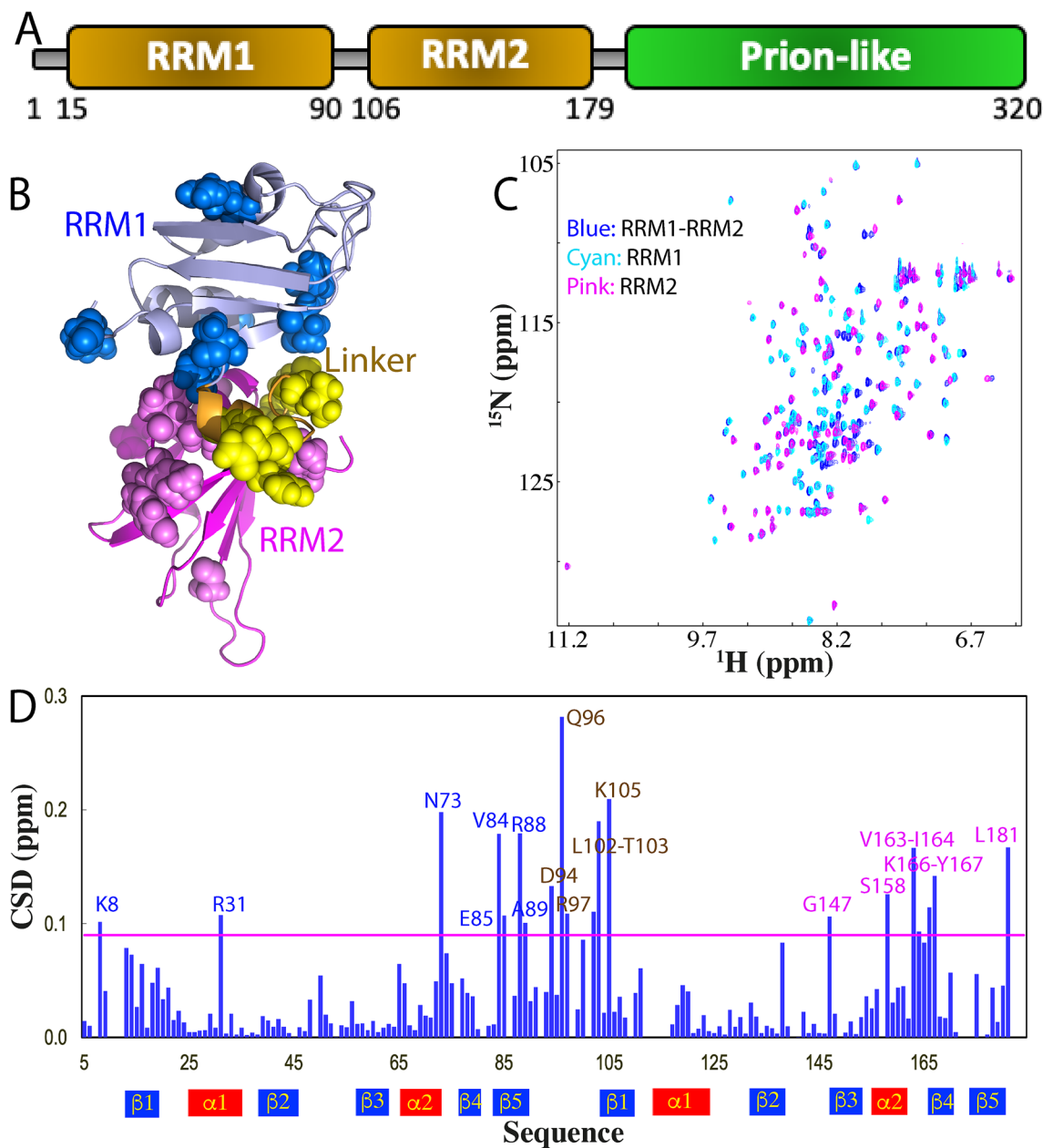


Figure 3. Dissection-induced perturbation of hnRNPA1 RRM domains. (A) 320-residue hnRNPA1 contains: two RNA recognition motifs (RRM1 and RRM2) over residues 1–179, and C-terminal prion-like domain over residues 180–320. (B) NMR structure (PDB ID of 2LYV) of the hnRNPA1 RRM1 (light blue) and RRM2 (pink) in which the residues are displayed in spheres for those with HSQC peaks significantly shifted in RRM1 (blue) or in RRM2 (light pink) upon dissection. The image was prepared by PyMol2.4 (<https://pymol.org>). (C) Superimposition of ^1H - ^{15}N NMR HSQC spectra of the ^{15}N -labeled tethered RRM1–RRM2 domains (blue), the isolated RRM1 (cyan) and RRM2 (pink) domains at 50 μM in 10 mM sodium phosphate buffer (pH 6.8) containing 150 mM NaCl and 10 mM DTT. (D) Residue-specific chemical shift difference (CSD) of the RRM1 and RRM2 domains between the tethered and isolated forms. Significantly shifted residues are labeled and displayed as spheres, which are defined as those with the CSD values >0.09 (average value + one standard deviation) (pink line).

Visualization of the ATP-RRM2 complex of hnRNPA1. Because of the extremely low binding affinity with K_d of $\sim\text{mM}$, it is impossible to determine the three-dimensional structure of the ATP-RRM2 complexes by the classic methods of NMR spectroscopy or X-ray crystallography. So here to visualize the complex structure, we utilized the NMR-binding derived constraints to guide the molecular docking with the well-established HADDOCK program³¹, as we extensively conducted before on the non-classic ATP-protein complexes^{20–22,30}.

Figure 5C presents the lowest-energy docking structure of the ATP-RRM2 complex of hnRNPA1. Overall, this structure is very similar to those of the ATP-RRM1 and ATP-RRM2 complexes of TDP-43 in which ATP

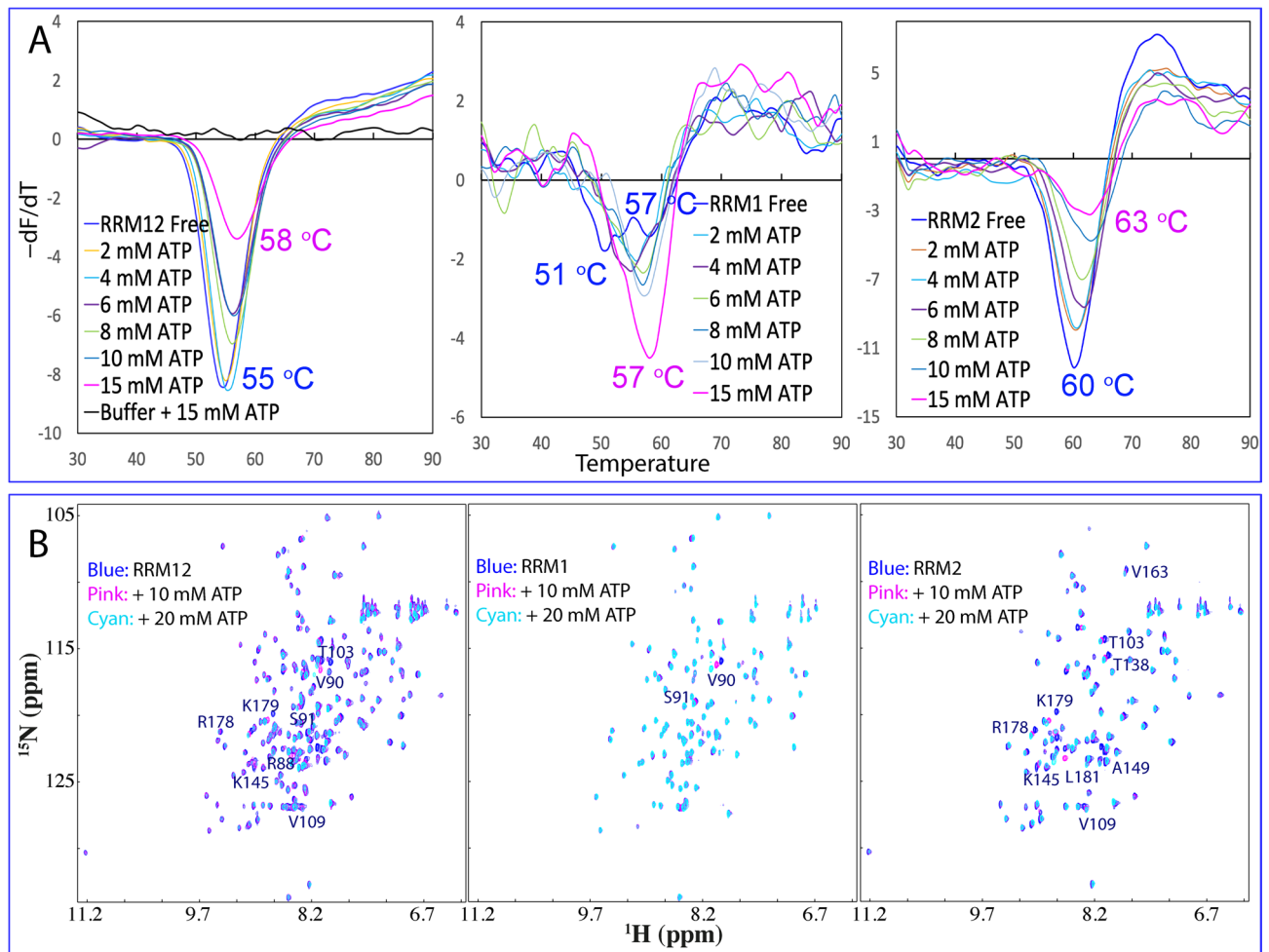


Figure 4. Thermal stability and ATP binding of the hnRNPA1 RRM domains. **(A)** DSF melting curves of thermal unfolding of the tethered RRM1–RRM2, as well as isolated RRM1 and RRM2 domains in the presence of ATP at different concentrations. **(B)** ^1H - ^{15}N NMR HSQC spectra of the ^{15}N -labeled tethered RRM1–RRM2, as well as isolated RRM1 and RRM2 domains at 50 μM (blue) in 10 mM sodium phosphate buffer (pH 6.8) containing 150 mM NaCl and 10 mM DTT in the absence (blue), and in the presence of ATP at 10 (pink) and 20 (cyan) mM.

occupies a pocket within the conserved surfaces of RRM domains for binding various nucleic acids (Fig. 5D). A close examination reveals that in the ATP–RRM2 complex of hnRNPA1, the aromatic purine ring of ATP has close contacts with the positively-charged surface constituted by the side chains of both RRM2 Arg178 and Lys179 likely to establish π -cation interactions on the one hand, as well as with Phe108 and Phe148 likely by establishing π - π interactions on the other hand (Fig. 5E). Furthermore, the NH of the purine ring of ATP forms a hydrogen bond with the backbone oxygen of Leu181, while the α - and β -phosphate oxyanions of ATP form other two hydrogen bonds respectively with the backbone nitrogen atoms of Gly114 and Gly147 (Fig. 5F).

So far, we have studied the binding complexes of ATP to FUS and TDP-43 RRM domains, as well as non-classic AcD domains which are all capable of binding nucleic acids. ATP appears to always occupy the pockets within their interfaces utilized for binding nucleic acids. Although ATP has no complete binding pocket on the RRM1 domain of hnRNPA1, the three residues with large shifts of HSQC peaks induced by adding ATP are also located within the conserved surfaces for RRM domains to bind nucleic acids (Fig. 5G).

Dynamic behaviours of the TDP-43 tethered and isolated RRM domains. To understand the dynamic basis underlying the coupling of the tethered RRM1–RRM2 of TDP-43, we conducted molecular dynamics (MD) simulations for the tethered as well as isolated RRM1 and RRM2 of TDP-43 with three parallel 50-ns simulations for each constructs. Molecular dynamics simulation is a powerful tool which can not only provide insights into the conformational dynamics that underlies protein functions, but also detect long-range inter-domain correlation motions^{32–34} as we previously showed on other proteins^{35–38}.

I of Fig. 6A presents the root-mean-square deviations (RMSD) of the Ca atoms averaged over three trajectories for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink). The tethered RRM1–RRM2 of TDP-43 has larger RMSD value ($4.32 \pm 0.39 \text{ \AA}$) than those of the isolated RRM1 ($3.12 \pm 0.35 \text{ \AA}$)

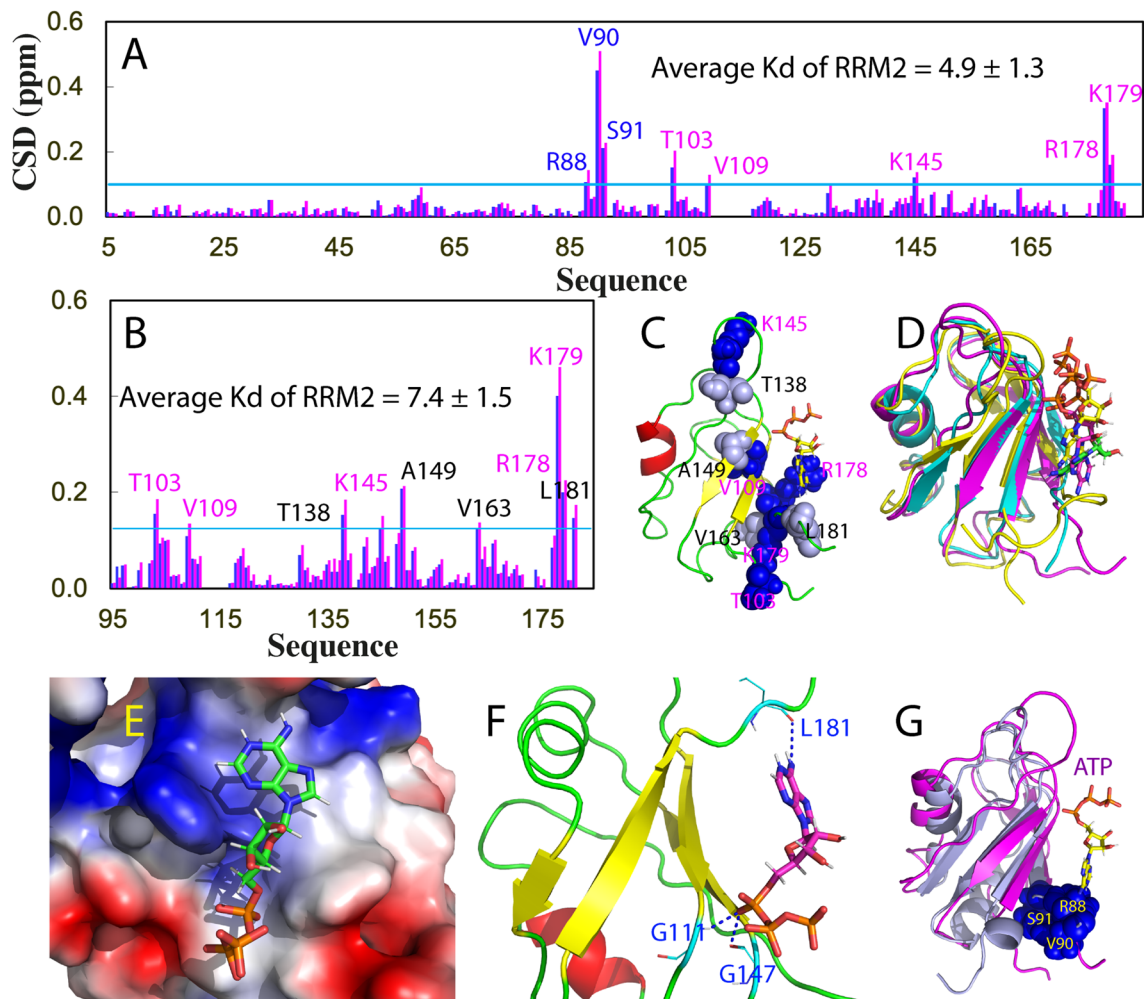


Figure 5. ATP binding of the tethered RRM1–RRM2, as well as isolated RRM2 of hnRNPA1. **(A)** Residue-specific chemical shift difference (CSD) of the tethered hnRNPA1 RRM1–RRM2 in the presence of ATP at 10 mM (blue) and 20 mM (pink). Significantly shifted residues are labeled, which are defined as those with the CSD values at 20 mM ATP > 0.1 (average value + one standard deviation) (cyan line). **(B)** Residue-specific chemical shift difference (CSD) of the isolated RRM2 in the presence of ATP at 10 mM (blue) and 20 mM (pink). Significantly shifted residues are labeled, which are defined as those with the CSD values at 20 mM ATP > 0.125 (average value + one standard deviation) (cyan line). The shifted residues identical to those in the tethered form are labeled in pink while the additionally shifted residues are in black. **(C)** The lowest energy docking model of the ATP–RRM2 complex. The structure of hnRNPA1 RRM2 is displayed in ribbon, while ATP is in sticks. The nine residues with significant CSD values are displayed in spheres and labeled. **(D)** Superimposition of three complexes of ATP: with hnRNPA1 RRM2 (pink), TDP-43 RRM1 (yellow) and TDP-43 RRM2 (cyan). **(E)** The ATP–RRM2 complex of hnRNPA1 with the RRM2 structure displayed in the electrostatic potential surface and ATP in sticks. **(F)** The ATP–RRM complex of hnRNPA1 showing three hydrogen bonds (in blue dotted lines) between ATP and RRM2 atoms. **(G)** Superimposition of the hnRNPA1 RRM1 and ATP–RRM2 complex with the three significantly shifted RRM1 residue (Arg88, Val90 and Ser91) displayed in spheres. The image was prepared by PyMol2.4 (<https://pymol.org>).

and RRM2 (2.72 ± 0.32 Å). Furthermore, we calculated RMSD values by overlaying only RRM1 (104–179) and RRM2 (191–261) in both tethered and isolated forms with the exclusion of the unstructured N- and C-termini. RRM1 (104–179) in the tethered form still has larger RMSD value (4.39 ± 0.41 Å) than RRM1 (104–179) in the isolated form (2.51 ± 0.23 Å) (II of Fig. 6A). Similarly, RRM2 (191–261) in the tethered form also has larger RMSD value (3.48 ± 0.37 Å) than RRM2 (104–179) in isolated form (1.74 ± 0.14 Å) (III of Fig. 6A).

Figure 6B presents the structure snapshots in the first MD simulations for the tethered RRM1–RRM2 as well as isolated RRM1 and RRM2, clearly indicating that the structures of the tethered RRM1–RRM2 are more fluctuating than those of the isolated RRM1 or RRM2, completely consistent with the RMSD results. Similar dynamic behaviours are also reflected by the root-mean-square fluctuations (RMSF) of the C α atoms averaged over three trajectories (Fig. 7). As shown in Fig. 7A, while the residue-specific fluctuations of RRM1 are very similar in both tethered and isolated forms, those of RRM2 in the tethered form are higher than those in the isolated forms. As a consequence, the majority of the residues with significant differences of RMSF between tethered and isolated

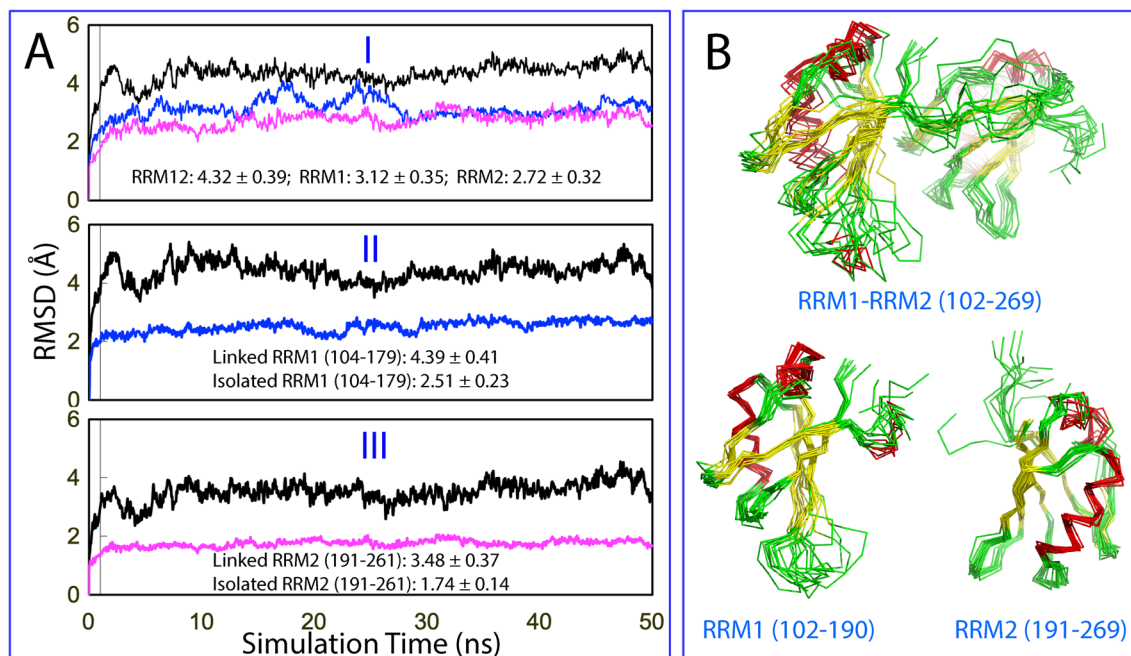


Figure 6. Overall dynamic behaviors of the TDP-43 RRM domains. **(A)** Average root-mean-square deviations (RMSD) of the Ca atoms over three independent 50-ns MD simulations for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink) (I). For the RRM1 in the tethered form (black) and isolated form (blue) with the unstructured N- and C-termini excluded for calculation (II); as well as for the RRM2 in the tethered form (black) and isolated form (pink) with the unstructured N- and C-termini excluded for calculation (III). **(B)** Structure snapshots in the first MD simulation with one structure for each 5 ns. The image was prepared by PyMol2.4 (<https://pymol.org>).

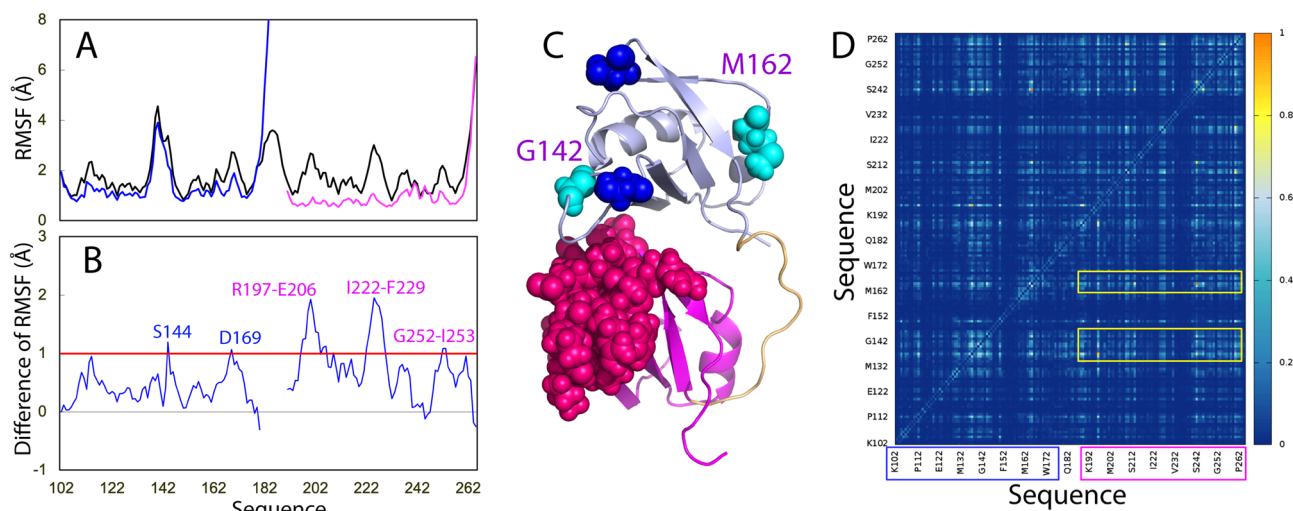


Figure 7. Residue-specific dynamic behaviors of the TDP-43 RRM domains. **(A)** Averaged root-mean-square fluctuations (RMSF) of the Ca atoms computed over three independent MD simulations for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink). **(B)** The difference of the average RMSF between the tethered RRM1–RRM2, and isolated RRM1/RRM2. The residues are labeled for those with significant difference of RMSF (> 1.0 ; average + one STD). **(C)** Structure of the TDP-43 RRM1–RRM2 with the residues displayed in spheres which show significant difference of RMSF between the tethered and isolated forms. The image was prepared by PyMol2.4 (<https://pymol.org>). **(D)** Mutual information matrix calculated from three parallel MD simulation data of the tethered TDP-43 RRM1–RRM2 by MutInf, with residues having significant inter-domain correlation motions highlighted by yellow boxes. The RRM1 sequence is indicated by blue box and RRM2 sequence in pink box.

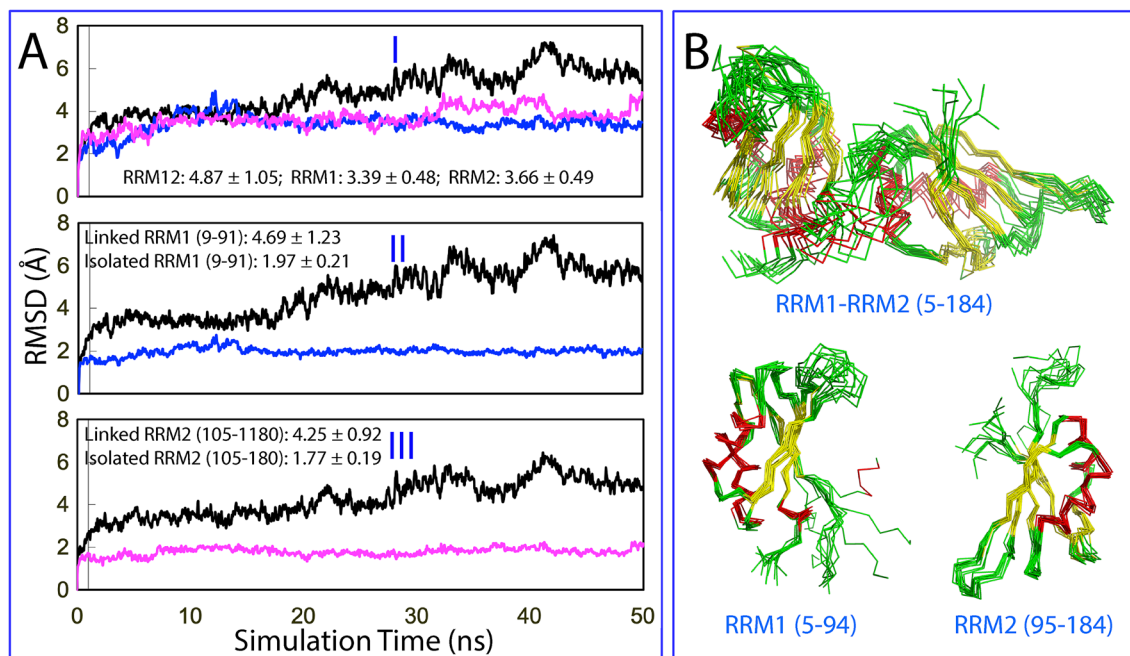


Figure 8. Overall dynamic behaviors of the hnRNPA1 RRM domains. **(A)** Average root-mean-square deviations (RMSD) of the Ca atoms over three independent 50-ns MD simulations for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink) (I). for the RRM1 in the tethered form (black) and isolated form (blue) with the unstructured N- and C-termini excluded for calculation (II); as well as for the RRM2 in the tethered form (black) and isolated form (pink) with the unstructured N- and C-termini excluded for calculation (III). **(B)** Structure snapshots in the first MD simulation with one structure for each 5 ns. The image was prepared by PyMol2.4 (<https://pymol.org>).

forms are located in RRM2 (Fig. 7B,C), which is only in general consistent with the dissection-induced effects as detected by NMR (Fig. 1D). Nevertheless, the involved residues detected by NMR HSQC and MD could not be exactly the same as two methods report different probes and time scales.

MutInf represents an entropy-based approach to analyze ensembles of protein conformers, such as those from molecular dynamics simulations by using internal coordinates and focusing on dihedral angles. In particular, this approach is particularly applicable for those in which conformational changes are subtle³³. Briefly, this approach utilizes second-order terms from the configurational entropy expansion, called the mutual information, to identify pairs of residues with correlated conformations, or correlated motions. Figure 7D shows the normalized correlation motion matrix of the tethered RRM1–RRM2 of TDP-43. Interestingly, the correlation motions exist not only within RRM1 or RRM2 domain, but also between two domains. In particular, the RRM1 residues around Gly142 and Met162 have extensive correlation motions with many RRM2 residues. Furthermore, the strength of the inter-domain correlation motions has no significant difference from that of the intra-domain motions, suggesting that the tethered RRM1–RRM2 indeed behaves as a coupled dynamic unit. This rationalizes the DSF results that the tethered RRM1–RRM2 of TDP-43 only has one thermal unfolding transition although the isolated RRM1 and RRM2 have their own transitions with very different T_m values.

Dynamic behaviours of the hnRNPA1 tethered and isolated RRM domains. We also conducted 50-ns molecular dynamics (MD) simulations for the tethered as well as isolated RRM1 and RRM2 of hnRNPA1 with three parallel simulations for each constructs. I of Fig. 8A presents the root-mean-square deviations (RMSD) of the Ca atoms averaged over the three trajectories for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink). Interestingly, the tethered RRM1–RRM2 of hnRNPA1 also has larger RMSD value (4.87 ± 1.05 Å) than those of RRM1 (3.39 ± 0.48 Å) and RRM2 (3.66 ± 0.49 Å). Furthermore, although the unstructured N- and C-termini were not included for calculation, the RRM1 (9–91) in the tethered form still has larger RMSD value (4.69 ± 1.23 Å) than RRM1 (9–91) in the isolated form (1.97 ± 0.21 Å) (II of Fig. 8A). Similarly, RRM2 (105–180) in the tethered form also has larger RMSD value (4.25 ± 0.92 Å) than RRM2 (105–180) in the isolated form (1.77 ± 0.19 Å) (III of Fig. 8A).

Figure 8B presents the structure snapshots in the first MD simulations for the tethered RRM1–RRM2 as well as isolated RRM1 and RRM2 of hnRNPA1, showing that the structures of the tethered RRM1–RRM2 are indeed more fluctuating than those of the isolated RRM1 and RRM2, completely consistent with the RMSD results. Noticeably, different from those observed for the TDP-43 RRM domains (Fig. 7), the residue-specific RMSF values of both RRM1 and RRM2 of hnRNPA1 in the tethered forms are larger than those of isolated RRM1 and RRM2 (Fig. 9A). Consequently, the residues with significant differences of RMSF between the tethered and

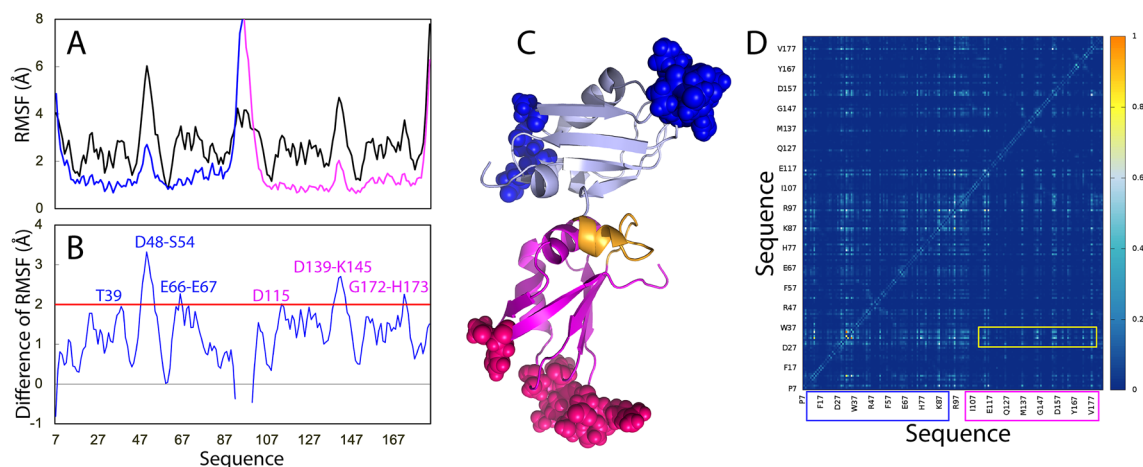


Figure 9. Residue-specific dynamic behaviors of the hnRNPA1 RRM domains. **(A)** Averaged root-mean-square fluctuations (RMSF) of the Ca atoms computed over three independent MD simulations for the tethered RRM1–RRM2 (black), as well as isolated RRM1 (blue) and RRM2 (pink). **(B)** The difference of the average RMSF between the tethered RRM1–RRM2, and isolated RRM1/RRM2. The residues are labeled for those with significant difference of RMSF (> 2.0 ; average + one STD). **(C)** Structure of the hnRNPA1 RRM1–RRM2 with the residues displayed in spheres which show significant difference of RMSF between the tethered and isolated states. The image was prepared by PyMol2.4 (<https://pymol.org>). **(D)** Mutual information matrix calculated from three parallel MD simulation data of the tethered hnRNPA1 RRM1–RRM2 by MutInf, with residues having significant inter-domain correlation motions highlighted by yellow box. The RRM1 sequence is indicated by blue box and RRM2 sequence in pink box.

isolated forms are located on both RRM1 and RRM2 (Fig. 9B,C), which is also in general consistent with the dissection-induced perturbation as detected by NMR (Fig. 3D).

Figure 9D presents the normalized correlation motion matrix of the tethered RRM1–RRM2 of hnRNPA1. Although the correlation motions still exist between two domains, the inter-domain correlation motions are mainly between the RRM1 residues Arg31–Ser32 and RRM2 residues. The strength of the correlation motions for the hnRNPA1 RRM1–RRM2 is weaker than that observed for the TDP-43 RRM1–RRM2 (Fig. 7D). This suggests that the coupling of hnRNPA1 RRM1 and RRM2 domains might be weaker than that for TDP-43 RRM1 and RRM2 domains.

Discussion

Protein aggregation/fibrillation has been now established to be the universal hallmark of an increasing spectrum of human diseases beyond neurodegenerative diseases, which also include cardiac dysfunction, eye cataract, degeneration of muscle and bone, as well as aging down to *E. coli*^{24,39–50}. Out of various factors that modulate aggregation/fibrillation of the folded proteins, two key determinants are thermodynamic stability and conformational dynamics. Nevertheless, despite exhaustive studies, the relationship between thermodynamic stability and conformational dynamics still remains largely elusive. In human genome, many RRM-containing proteins such as TDP-43 and hnRNPA1 have two tethered RRM domains, which have been demonstrated to play a key role in the disease-causing aggregation/fibrillation. Although the previous observations imply that the tethered TDP-43 RRM domains are particularly prone to aggregation/fibrillation, so far there has been no systematic study to understand the underlying mechanisms, which, however, are of both fundamental and therapeutic interest.

In the present study, by both experimental and computational approaches we conducted a systematic study to characterize the thermal stability and ATP-binding by DSF and NMR, followed by MD simulations to assess the conformational dynamics of two RRM domains of TDP-43 and hnRNPA1 in both tethered and isolated forms. Very unexpectedly, the results showed that the isolated TDP-43 RRM1 and RRM2 domains have T_m of 57 and 59 °C respectively, while the tethered form has only one denaturation transition with T_m significantly reduced to only 49 °C. This set of results indicates that the tethering induced the significant coupling of two RRM domains of TDP-43, as well as dramatic destabilization. Intriguingly, no significant destabilization was observed for the tethering of two RRM domains of hnRNPA1. The results thus underscore the extreme complexity of the tethering-induced effects even for the tandem RRM domains of the different members within the same hnRNP protein family. In a general context, the dramatic destabilization for the tethered TDP-43 RRM domains is also very unusual because previously the tethering was demonstrated to have either no significant effect or even to stabilize the tethered domains. Only recently it was found that the tethering may also destabilize the protein domains^{51–54}, as exemplified by the ubiquitination-induced destabilization of the modified proteins, which was proposed to function to facilitate their degradation⁵⁴. Interestingly, the tethering-induced destabilization has been proposed to evolve from the trade-offs for functions⁵³. In this regard, it is of great interest in the future to define the functional role of the unique destabilization for tandem TDP-43 RRM domains. Nevertheless, this unexpected destabilization certainly contributes to the unusually high tendency of TDP-43 in aggregation/fibrillation, which has been well established to lead to a variety of human diseases by “loss of function” or/and “gain of function”.

The tethering-induced effects for TDP-43 and hnRNPA1 RRM domains appear to result mainly from their inter-domain connection. Residue-specific NMR data showed that the dissection-induced perturbation for TDP-43 RRM domains is much more profound than that for hnRNPA1 RRM domains, while MD simulations further revealed that the inter-domain correlation motions of TDP-43 RRM domains are more extensive than those for hnRNPA1 RRM domains. Therefore, experimental and simulation results together rationalize the observation that the tethered TDP-43 RRM domains could become highly coupled with only one denaturation transition with T_m much lower than those of their isolated ones. On the other hand, even only based on our previous studies^{35–38,55}, the conformational dynamics appear to operate through a global network and the perturbation by a mutation or binding without altering the average structure will trigger extensive reorganization of the whole network. Therefore, it still remains an extreme challenge in the future to integrate the results of thermodynamic stability and conformational dynamics to understand why the inter-domain connection significantly destabilize the two RRM domains of TDP-43 but lead to even a slight stabilizing effect on the RRM1 domain of hnRNPA1.

Mysteriously, all cells maintain very high ATP concentrations of 2–12 mM, much higher than those required for its previously-known functions^{56–58}, although the majority of ATP needs to be produced by very complex supramolecular machineries embedded in membranes⁵⁶. Only recently, it was decoded that ATP with concentrations > 5 mM acts to hydrothermally dissolve liquid–liquid phase separation (LLPS), aggregation/fibrillation⁵⁷, which appears to operate at a proteome-wide scale⁵⁸. We further found that by weak but specific binding with K_d of ~ mM, ATP is also able to biphasically modulate LLPS of intrinsically disordered domains^{59,60}. Remarkably, ATP can inhibit amyloid fibrillation not only for the tethered TDP-43 RRM1–RRM2 by enhancing its thermal stability²¹, but also for the single FUS RRM domain by blocking the dynamic opening of the RRM fold without alternation of its thermal stability²⁰. Here, again we found that ATP can specifically bind the hnRNPA1 RRM2 domain in both tethered and isolated contexts with the affinity and complex structure very similar to those of the FUS and TDP-43 RRM domains. Moreover ATP can even enhance the thermal stability of the hnRNPA1 RRM1–RRM2 domains. Interestingly, our previous and current results together suggest that the affinities of the ATP binding to the tethered forms of both TDP-43 and hnRNPA1 RRM domains are slightly higher than those to the isolated forms. This may be mainly due to the higher conformational dynamics of the tethered RRM domains than those of the isolated forms as uncovered by MD simulations, which thus provides ATP the higher dynamic accessibility to the binding pockets in the tethered RRM domains.

Previous studies implied that TDP-43 RRM domains might start to assemble into amyloid structures without needing the complete denaturation¹⁴. We also showed that the relatively high conformational dynamics of FUS RRM domain appear to be sufficient to allow the opening of the RRM fold, thus leading to aggregation/fibrillation. Furthermore, the ATP binding to FUS RRM even without enhancing its stability is sufficient to kinetically block the conformational opening. Therefore, the results here that ATP can bind the conserved pockets of TDP-43 and hnRNPA1 RRM domains again highlight that ATP may play a general role in preventing the pathological aggregation/fibrillation of the RRM-containing proteins containing more than one RRM domain.

In summary, in the present study we showed that upon tethering, two TDP-43 RRM domains become highly coupled but dramatically destabilized with the T_m reduction of ~ 8 °C. On the other hand, no significant destabilization occurs for the tethering of two hnRNPA1 RRM domains. Mechanistically, the tethering-induced effects appear to mainly result from the inter-domain connection between two RRM domains as reflected by NMR and MD simulation results. Moreover, we showed that ATP can specifically bind TDP-43 and hnRNPA1 RRM domains with the affinity to the tethered forms slightly higher than to the isolated forms. Results together thus suggest that ATP, the universal energy currency, may also play a general role in preventing aggregation/fibrillation of RRM-containing proteins, which has been extensively identified to cause an increasing spectrum of human diseases beyond neurodegenerative diseases. Therefore, our results imply a potential mechanism to rationalize the observation that upon being aged, the risk of protein aggregation-causing diseases increases most likely also because ATP concentrations gradually reduce in all cells during aging^{53,54}.

Methods

Cloning, expression and purification of the tethered and isolated RRM domains of TDP-43 and hnRNPA1. Previously the expression vectors of TDP-43 RRM1–RRM2 (102–269), RRM1 (102–191) and RRM2 (191–269) have been constructed and their expression and purification were established^{21,22,61}. On the other hand, the expression vector of the tethered RRM1–RRM2 (5–184) of hnRNPA1 were purchased from a local company (Bio Basic Asia Pacific Pte Ltd), which was subsequently dissected into the isolated RRM1 (5–95) and RRM2 (94–184) by PCR with designed primers.

The six recombinant proteins were expressed in *E. coli* BL21 cells with IPTG induction at 20 °C overnight. They were found all in the supernatant, and therefore were purified by a Ni^{2+} -affinity column (Novagen) under native conditions. Subsequently the on-gel cleavage by thrombin was conducted and the eluted fractions containing the RRM proteins were further purified by a heparin column to remove nucleic acids followed by FPLC purification with either a Superdex-75 or a Superdex-200 column.

Here we followed our previous protocol to generate isotope-labeled proteins for NMR studies^{20–24,61}. Briefly, the bacteria were grown in M9 medium with addition of $(^{15}NH_4)_2SO_4$ for ^{15}N -labeling. The protein concentrations were determined by the UV spectroscopic method in the presence of 8 M urea, under which the molar extinct coefficient at 280 nm of a protein can be calculated by adding up the contribution of Trp, Tyr, and Cys residues^{59,62}.

ATP was purchased from Sigma-Aldrich with the same catalog numbers as previously reported. $MgCl_2$ was added into ATP for stabilization by forming the ATP-Mg complex^{20–22,57}. The fluorescent dye SYPRO Orange (S5692-50UL) was purchased from Sigma-Aldrich. The protein samples, as well as ATP, were all prepared in 10 mM sodium phosphate buffer containing 10 mM DTT and 150 mM NaCl with a final pH adjusted to 6.8.

Determination of thermal stability by fluorescence spectroscopy and DSF. To monitor the thermal denaturation by intrinsic Trp fluorescence for the tethered RRM1–RRM2, isolated RRM1 and RRM2 domains as well as their mixture at 1:1 of both TDP-43 and hnRNPA1 at 10 μ M in 10 mM sodium phosphate buffer containing 10 mM DTT and 150 mM NaCl (pH 6.8), their emission spectra of UV fluorescence were acquired on a Jasco J-1500 spectropolarimeter from 25 to 95 $^{\circ}$ C at a 5-degree interval with the excitation wavelength at 280 as previously described and the fluorescence intensity was reported in arbitrary unit²³.

Furthermore, as we previously showed^{20–24}, ATP triggered very high non-specific noise in CD spectroscopy and quenched the intrinsic fluorescence of exposed Trp residues, here again we used differential scanning fluorimetry (DSF) as we previously reported to determine the thermodynamic stability of RRM1–RRM2, RRM1 and RRM2 domains of TDP-43 and hnRNPA1 at 10 μ M in 10 mM sodium phosphate buffer containing 10 mM DTT and 150 mM NaCl (pH 6.8) with addition of ATP at different concentrations.

DSF experiments were performed using the CFX384 Touch Real-Time PCR Detection System from BIO-RAD, following the SYBR green melting protocol to obtain T_m value^{21–24}. Briefly, in a single well of a 384-well PCR plate, a 10 μ l reaction solution was placed, which contains the RRM12, or RRM1 or RRM2 domain at 10 μ M, ATP at different concentrations, and 10 \times SYPRO Orange in 10 mM sodium phosphate buffer containing 150 mM NaCl (pH 6.8). The program in Real-Time PCR instrument was set to be SYBR green and run temperature scan from 30 to 90 $^{\circ}$ C with the increment of 1 $^{\circ}$ C per minute. Upon completion, the obtained thermal unfolding curves were displayed as the first derivatives (dF/dT) by the RT-PCR software Bio-Rad CFX Manager 3.0.

NMR characterizations of the ATP binding. All NMR experiments were acquired at 25 $^{\circ}$ C on an 800 MHz Bruker Avance spectrometer equipped with pulse field gradient units and a shielded cryoprobe as we described previously^{20–24}. For NMR HSQC titration studies of the interactions of RRM1–RRM2, RRM1 and RRM2 with ATP, two dimensional ^1H - ^{15}N NMR HSQC spectra were collected on the ^{15}N -labelled samples at a protein concentration of 50 μ M in 10 mM sodium phosphate buffer containing 10 mM DTT and 150 mM NaCl (pH 6.8) at 25 $^{\circ}$ C in the absence and in the presence of ATP at concentrations of 0.5, 1.0, 2.0, 4.0, 6.0, 8.0, 10.0, 12.0, 14.0, 16.0, 18.0, and 20.0 mM.

Calculation of CSD and data fitting to obtain K_d . Sequential assignments of the tethered and isolated RRM1 and RRM2 domains of TDP-43 and hnRNPA1 were achieved based on the previously deposited chemical shifts: the tethered TDP-43 RRM1–RRM2 domain (BMRB ID of 19290 and 27613), the isolated RRM1 (BMRB ID of 18765) and RRM2 (BMRB ID of 19922), as well as the tethered hnRNPA1 RRM1–RRM2 domain (BMRB ID of 18728).

Due to the fast exchange with bulk water, or/and μ s–ms dynamics, or/and overlap for some, particularly loop residues, the 168-residue TDP-43 RRM1–RRM2 containing 6 Pro residues had 142 peaks detected and assigned. The 90-residue TDP-43 RRM1 containing 3 Pro residues had 82 peaks detected and assigned, while the 79-residue TDP-43 RRM2 containing 3 Pro residues had 72 peaks detected and assigned. The 184-residue hnRNPA1 RRM1–RRM2 containing 6 Pro residues had 155 peaks detected and assigned. The 96-residue hnRNPA1 RRM1 containing 5 Pro residues had 86 peaks detected and assigned, while the 90-residue hnRNPA1 RRM2 containing 1 Pro residues had 76 peaks detected and assigned.

To calculate chemical shift difference (CSD), the HSQC spectra were superimposed for the ^{15}N -labelled RRM1–RRM2, RRM1 and RRM2 domains collected in the absence and in the presence of ATP at different concentrations. Subsequently, the shifted HSQC peaks could be identified and further assigned to the corresponding RRM residues based on the sequential assignments. The chemical shift difference (CSD) was calculated by an integrated index calculated by the following formula:

$$\text{CSD} = \left((\Delta^1\text{H})^2 + (\Delta^{15}\text{N})^2 / 4 \right)^{1/2}.$$

In order to obtain residue-specific dissociation constant (K_d), we fitted the shift traces of the residues with significant shifts of HSQC peaks (CSD > average + STD), by using the one binding site model^{20–23,29} with the following formula:

$$\text{CSD}_{\text{obs}} = \text{CSD}_{\text{max}} \left\{ ([\text{P}] + [\text{L}] + K_d) - \frac{([\text{P}] + [\text{L}] + K_d)^2 - 4[\text{P}][\text{L}]}{2[\text{P}]} \right\}^{1/2}$$

Here, [P] and [L] are molar concentrations of RRM domains and ligands (ATP) respectively.

Molecular docking. The structure model of the ATP–RRM2 complex of hnRNPA1 was constructed by use of HADDOCK software^{20–22,31}, which makes use of CSD data to derive the docking with various degrees of flexibility. Briefly, five residues of hnRNPA1 RRM2 with significant CSD values were set to be active residues and HADDOCK docking procedure for the complexes was performed at three stages: (1) randomization and rigid body docking; (2) semi-flexible simulated annealing; (3) flexible explicit solvent refinement. The ATP–RRM2 structure with the lowest energy score were selected for the detailed analysis and displayed by Pymol (The PyMOL Molecular Graphics System, Version 0.99rc6 Schrödinger, LLC).

Molecular dynamics (MD) simulations. For MD simulations, the NMR structures of RRM1–RRM2 of TDP-43 (PDB ID of 4BS2) and hnRNPA1 (PDB ID of 2LYV) were selected as the tethered models while their isolated RRM1 and RRM2 models were obtained by dissecting the two structures into the isolated RRM domains.

The simulation setting was previously reported^{36–38}. Briefly, the simulation cell is a periodic cubic box with a minimum distance of 10 Å between the protein and the box walls to ensure the protein would not directly

interact with its own periodic image given the cutoff. The water molecules, described using the TIP3P model, were filled in the periodic cubic box for the all atom simulation. To neutralize the system, some Na⁺ and Cl⁻ ions were randomly placed far away from the surface of the protein.

Three independent 50-ns MD simulations were performed for each of six constructs: namely the tethered RRM1–RRM2, isolated RRM1 and RRM2 of TDP-43 as well as of hnRNPA1 by the program GROMACS⁶³ with the AMBER-03 all-atom force field⁶⁴. The long-range electrostatic interactions were treated using the fast particle-mesh Ewald summation method, with the real space cutoff of 9 Å and a cutoff of 14 Å was used for the calculation of van der Waals interactions. The temperature during simulation was kept constant at 300 K by Berendsen's coupling. The pressure was held at 1 bar. The isothermal compressibility was $4.5 \times 10^{-5} \text{ bar}^{-1}$. The time step was set as 2 fs. Prior to MD simulations, all the initial structures were relaxed by 1000 steps of energy minimization using steepest descent algorithm, followed by 100 ps equilibration with a harmonic restraint potential applied to all the heavy atoms of the proteins.

Analysis of correlation motions. To identify the coupling of RRM1 and RRM2 domains of both TDP-43 and hnRNPA1, we calculated their correlation matrices of three independent simulations of the tethered RRM1–RRM2 by MutInf⁵⁴. MutInf represents an entropy-based method to analyze ensembles of protein conformers, including those from MD simulations by using internal coordinates and focusing on dihedral angles. This method is particularly applicable for subtle conformational changes by utilizing second-order terms of the configurational entropy expansion called mutual information to pinpoint the pairs of residues with correlated conformations, or correlated motions, in an equilibrium ensemble. Here, the normalized matrix values were used and the threshold value was set up to 0.3 for determining the pairs of highly correlated residues.

Received: 2 August 2020; Accepted: 21 December 2020

Published online: 13 January 2021

References

1. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
2. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
3. Harrison, A. F. & Shorter, J. RNA-binding proteins With Prion-Like domains in health and disease. *Biochem J.* **474**, 1417–1438 (2017).
4. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: Insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).
5. Cléry, A. M., Blatter, M. & Allain, F. H. RNA recognition motifs: Boring?. *Not. Quite Curr. Opin. Struct. Biol.* **18**, 290–298 (2008).
6. Josephs, K. A. *et al.* TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol.* **127**, 811–824 (2014).
7. Nelson, P. T. *et al.* Limbic-predominant age-related TDP-43 encephalopathy (LATE): Consensus working group report. *Brain* **142**, 1503–1527 (2019).
8. Vogler, T. O. *et al.* TDP-43 and RNA form amyloid-like Myo-granules in regenerating muscle. *Nature* **563**, 508–513 (2018).
9. Palomo, V. *et al.* TDP-43: A key therapeutic target beyond amyotrophic lateral sclerosis. *ACS Chem. Neurosci.* **10**, 1183–1196 (2019).
10. Lim, L. *et al.* ALS-causing mutations significantly perturb the self-assembly and interaction with nucleic acid of the intrinsically disordered prion-like domain of TDP-43. *PLoS Biol.* **14**, e1002338 (2016).
11. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449 (2013).
12. Garnier, C. *et al.* Zinc binding to RNA recognition motif of TDP-43 induces the formation of amyloid-like aggregates. *Sci. Rep.* **7**, 6812 (2017).
13. Rabdano, S. O. *et al.* Onset of disorder and protein aggregation due to oxidation-induced intermolecular disulfide bonds: Case study of RRM2 domain from TDP-43. *Sci Rep.* **7**, 11161 (2017).
14. Zacco, E. *et al.* The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.* **11**, 372 (2018).
15. Tavella, D., Zitzewitz, J. A. & Massi, F. Characterization of TDP-43 RRM2 partially folded states and their significance to ALS Pathogenesis. *Biophys. J.* **115**, 1673–1680 (2018).
16. Guenther, E. L. *et al.* Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol.* **25**, 311–319 (2018).
17. Agrawal, S. *et al.* RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Sci. Rep.* **9**, 6171 (2019).
18. Prakash, A. *et al.* Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *J. Biomol. Struct. Dyn.* **37**, 178–194 (2019).
19. François-Moutal, L. *et al.* Small molecule targeting TDP-43's RNA recognition motifs reduces locomotor defects in a Drosophila Model of amyotrophic lateral sclerosis (ALS). *ACS Chem. Biol.* **14**, 2006–2013 (2019).
20. Kang, J., Lim, L. & Song, J. ATP binds and inhibits the neurodegeneration-associated fibrillization of the FUS RRM domain. *Commun. Biol.* **2**, 223 (2019).
21. Dang, M. *et al.* ATP is a cryptic binder of TDP-43 RRM domains to enhance stability and inhibit ALS/AD-associated fibrillation. *Biochem. Biophys. Res. Commun.* **522**, 247–253 (2020).
22. Dang, M. & Song, J. ALS-causing D169G mutation disrupts the ATP-binding capacity of TDP-43 RRM1 domain. *Biochem. Biophys. Res. Commun.* **24**, 459–464 (2020).
23. Lu, Y., Lim, L. & Song, J. RRM domain of ALS/FTD-causing FUS characteristic of irreversible unfolding spontaneously self-assembles into amyloid fibrils. *Sci. Rep.* **7**, 1043 (2017).
24. He, Y., Kang, J. & Song, J. ATP antagonizes the crowding-induced destabilization of the human eye-lens protein γ S-crystallin. *Biochem. Biophys. Res. Commun.* **526**, 1112–1117 (2020).

25. Scott, D. D., Francois-Moutal, L., Kumirov, V. K. & Khanna, M. ¹H, ¹⁵N and ¹³C backbone assignment of apo TDP-43 RNA recognition motifs. *Biomol. NMR Assign.* **13**, 163–167 (2019).
26. Barraud, P. & Allain, F. H. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: How the relative orientation between RRM influences the nucleic acid binding topology. *J. Biomol. NMR* **55**, 119–138 (2013).
27. Beusch, I. *et al.* Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *Elife* **6**, e25736 (2017).
28. Kang, J., Lim, L. & Song, J. ATP induces protein folding, inhibits aggregation and antagonizes destabilization by effectively mediating water-protein-ion interactions, the heart of protein folding and aggregation. *BioRxiv* <https://doi.org/10.1101/2020.06.21.163758> (2020).
29. Williamson, M. P. *Prog Nucl Magn Reson Spectrosc.* **73**, 1 (2013).
30. He, Y. *et al.* ATP binds nucleic-acid-binding domains beyond RRM fold. *Biochem. Biophys. Res. Commun.* **522**, 826–831 (2020).
31. de Vries, S. J. *et al.* HADDOCK versus HADDOCK: New Features and performance of HADDOCK20 on the CAPRI targets. *Proteins* **69**, 726–733 (2007).
32. Karplus, M. & McCammon, J. A. Dynamics of proteins: Elements and function. *Annu. Rev. Biochem.* **52**, 263–300 (1983).
33. Dodson, G. G., Lane, D. P. & Verma, C. S. Molecular simulations of protein dynamics: New windows on mechanisms in biology. *EMBO Rep.* **9**, 144–150 (2008).
34. McClendon, C. L. *et al.* Quantifying correlations between allosteric sites in thermodynamic ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009).
35. Lim, L. *et al.* Structurally- and dynamically-driven allostery of the chymotrypsin-like proteases of SARS, dengue and zika viruses. *Prog. Biophys. Mol. Biol.* **143**, 52–66 (2019).
36. Shi, J. *et al.* Dynamically-driven Inactivation of the catalytic machinery of the SARS 3C-like protease by the N214A mutation on the extra domain. *PLoS Comput. Biol.* **7**, e1001084 (2011).
37. Lua, S. *et al.* Structural, stability, dynamic and binding properties of the ALS-causing T46I mutant of the hVAPB MSP domain as revealed by NMR and MD simulations. *PLoS ONE* **6**, e27072 (2011).
38. Lim, L. *et al.* Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284–T285–I286/A mutations on the extra domain. *PLoS ONE* **9**, e101941 (2014).
39. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
40. Eisele, Y. S. *et al.* Targeting protein aggregation for the treatment of degenerative diseases. *Nat. Rev. Drug. Discov.* **14**, 759–780 (2015).
41. Willis, M. S. & Patterson, C. Proteotoxicity and cardiac dysfunction—Alzheimer’s disease of the heart?. *N. Engl. J. Med.* **368**, 455–464 (2013).
42. Meyer, T. & Voigt, N. In search for novel functions of adenosine 5'-triphosphate (ATP) in the heart. *Cardiovasc. Res.* **113**, e59–e60 (2017).
43. Greiner, J. V. & Glonek, T. Hydrotropic function of ATP in the crystalline lens. *Exp. Eye Res.* **190**, 107862 (2020).
44. He, Y., Kang, J. & Song, J. Cataract-causing G18V eliminates the antagonization by ATP against the crowding-induced destabilization of human gS-crystallin. *Biochem Biophys. Res. Commun.* <https://doi.org/10.1016/j.bbrc.2020.07.070> (2020).
45. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
46. Hartl, F. U. Cellular homeostasis and aging. *Annu. Rev. Biochem.* **85**, 1–4 (2016).
47. Walther, D. M. Widespread proteome remodeling and aggregation in aging *C. elegans*. *Cell* **161**, 919–932 (2013).
48. Ben-Zvi, A., Miller, E. A. & Morimoto, R. I. Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl. Acad. Sci. USA* **106**, 14914–14919 (2009).
49. David, D. C. *et al.* Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol.* **8**, e1000450 (2010).
50. Lindner, A. B. *et al.* Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. USA* **105**, 3076–3081 (2008).
51. Levy, Y. Protein assembly and building blocks: Beyond the limits of the LEGO brick metaphor. *Biochemistry* **56**, 5040–5048 (2017).
52. Vishwanath, S. *et al.* Same but not alike: Structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLoS Comput. Biol.* **14**, e1006008 (2018).
53. Bigman, L. S. & Levy, Y. Proteins: Molecules defined by their trade-offs. *Curr. Opin. Struct. Biol.* **60**, 50–56 (2020).
54. Morimoto, D. *et al.* Ubiquitylation directly induces fold destabilization of proteins. *Sci. Rep.* **6**, 39453 (2016).
55. Lim, L. *et al.* Curcumin allosterically inhibits the dengue NS2B-NS3 protease by disrupting its active conformation. *ACS Omega* **5**, 25677–25686 (2020).
56. Lehninger’s Principles of Biochemistry, 5th Edition. W.H. Freeman and Company, New York, pp. 502-503 (2005)
57. Patel, A. *et al.* ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).
58. Sridharan, S. *et al.* Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat. Commun.* **10**, 1155 (2019).
59. Kang, J., Lim, L. & Song, J. ATP enhances at low concentrations but dissolves at high concentrations liquid-liquid phase separation (LLPS) of ALS/FTD-causing FUS. *Biochem. Biophys. Res. Commun.* **504**, 545–551 (2018).
60. Kang, L. *et al.* A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids. *PLoS Biol.* **17**, e3000327 (2019).
61. Wei, Y. *et al.* Inter-domain Interactions of TDP-43 as Decoded by NMR. *Biochem Biophys Res Commun.* **473**, 614–619 (2016).
62. Pace, C. N. *et al.* How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**, 2411–2423 (1995).
63. Hess, B. *et al.* GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
64. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

Acknowledgements

This work was supported by Ministry of Education (MOE), Singapore Tier 1 Grant R-154-000-B45-114 and R-154-000-B92-114 to Jianxing Song.

Author contributions

J.S. and M.D. conceived and designed the experiments. M.D. Y.L. and J.S. performed the research, analyzed the data. M.D. and J.S. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80524-6>.

Correspondence and requests for materials should be addressed to J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021