BBA - Proteins and Proteomics 1866 (2018) 214–223

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43

Hao-Ru Li[a,1], Tsai-Chen Chen[a,1], Chih-Lun Hsiao[a], Lin Shi[b], Chi-Yuan Chou[b], Jie-rong Huang[a,c,*]

[a] Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan
[b] Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan
[c] Institute of Biomedical Informatics, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan

ARTICLE INFO

Keywords:
TDP-43
Self-association
Liquid-liquid phase separation
Intrinsically disordered proteins
NMR

ABSTRACT

The TAR DNA-binding protein of 43 kDa (TDP-43) has been identified as the main component of amyotrophic lateral sclerosis (ALS) cytoplasmic inclusions. The link between this proteinopathy and TDP-43′s intrinsically disordered C-terminal domain is well known, but recently also, this domain has been shown to be involved in the formation of the membraneless organelles that mediate TDP-43′s functions. The mechanisms that underpin the liquid-liquid phase separation (LLPS) of these membraneless organelles undergo remain elusive. Crucially though, these factors may be the key to understanding the delicate balance between TDP-43′s physiological and pathological functions. In this study, we used nuclear magnetic resonance spectroscopy and optical methods to demonstrate that an α-helical component in the centre (residues 320–340) of the C-terminal domain is related to the protein's self-association and LLPS. Systematically analysing ALS-related TDP-43 mutants (G298S, M337V, and Q331K) in different buffer conditions at different temperatures, we prove that this phase separation is driven by hydrophobic interactions but is inhibited by electrostatic repulsion. Based on these findings, we rationally introduced a mutant, W334G, and demonstrate that this mutant disrupts LLPS without disturbing this α-helical propensity. This tryptophan may serve as a key residue in this protein's LLPS.

1. Introduction

The transactivation response element (TAR) DNA-binding protein of 43 kDa (TDP-43) contains a folded N-terminal domain, two RNA-recognition motifs (RRM1 and RRM2), and a glycine-rich C-terminal domain. The structures of the folded N-terminal domain and the two RRMs have been determined [1–4], but its C-terminal domain (residues ~278–414) has low sequence complexity and is predicted to be intrinsically disordered (ID) [5,6]. TDP-43 has been identified as the major disease protein in frontotemporal lobar degeneration and in amyotrophic lateral sclerosis (ALS) [7]. Pathological TDP-43 forms 25-kDa hyperphosphorylated, ubiquitinated aggregates, proteolytically cleaved into C-terminal fragments [8]. This aggregation has been linked to a truncated RRM2 domain and the ID C-terminal domain [9–11]. The ID C-domain is also known to be the site of most of the familial and sporadic mutations of TDP-43 [12]. Although many studies have focused on this domain's implication in diseases, the onset of pathological TDP-43 aggregation remains unclear.

TDP-43 has also been noticed to form proteinaceous and membraneless organelles in cells [13,14]. This ability is similar to that of several recently studied proteins such as the FUS protein (an RNA binding protein with ID N-terminal domain) [15,16], DDX4 (which controls gene expression and translation) [17,18], BuGZ (a microtubule-associated zinc finger protein) [19], hnRNP A1 and A2 (nuclear ribonucleoproteins) [14,20,21], Whi3 (an RNA binding protein with polyQ expansion) [22], the NICD (the Nephrin intracellular domain) [23], and Pab1 (a poly(A)-binding protein) [24]. These “reversible aggregates” can assemble/dissolve in response to changes in the cell's environment and to maintain cellular integrity and homeostasis [25]—a process otherwise known as droplet formation [21], assemblage [26], or protein or liquid-liquid phase separation (LLPS) [27]. A common feature of these droplet-forming proteins is that they all have an intrinsically disordered region—for TDP-43, the C-terminal domain [13,28]—responsible for LLPS. The mechanisms driving LLPS vary substantially however. The presence of salt (NaCl) prevents LLPS in some cases, e.g. DDX4, hnRNP A1, and Whi3 [14,17,22], but not others, e.g. FUS and

Abbreviations: TDP-43, transactivation response element DNA-binding protein of 43 kDa; IDP, intrinsically disordered protein; LLPS, liquid-liquid phase separation; ALS, amyotrophic lateral sclerosis; HSQC, heteronuclear single-quantum coherence

* Corresponding author at: Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan.
E-mail address: jierongh@ym.edu.tw (J.-r. Huang).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bbapap.2017.10.001
Received 28 August 2017; Received in revised form 22 September 2017; Accepted 2 October 2017
Available online 04 October 2017
1570-9639/

NICD [16,23]. Most studies report LLPS as being enhanced at lower temperatures but BuGZ undergoes LLPS at higher temperatures [19]. A plausible explanation for these different LLPS behaviors is that several different forces are involved in the phase separation of the ID regions. Deriving general physicochemical rules to explain this biologically important phenomenon is therefore challenging.

Residue-specific structural information is crucial to understanding the aggregation and LLPS mechanisms of the C-terminal domain of TDP-43. However, its propensity to aggregate has meant that most studies have investigated separate fragments [5,29–38] (summarized in Supplementary Fig. S1). Only recently have two pioneering structural studies of the entire C-terminal domain been published [28,39]. The mechanisms of droplet formation, fibrillization, and aggregation are still matters of debate however [6,28,39,40]: Lim et al. identified an α-helix in the middle of the C-terminal domain at pH 4 and proposed an α-helix to β-strand transition during the fibrillization of TDP-43 [39]. On the contrary, Conicella et al. demonstrated that the helix-helix contact mediates this protein's self-association and LLPS, using detailed NMR and computational studies [28]. However, the lack of a clear pattern between the ALS-associated mutants and LLPS hinders the understanding of what the physical forces is mediating those interactions. Here, we use NMR spectroscopy and several other biophysical tools to study the self-association and droplet formation for the C-terminal domain of TDP-43. We studied three disease-related mutants at different buffer conditions and temperatures and prove that the hydrophobicity is the main force driving the formation of droplets whereas electrostatic repulsion acts against LLPS. Based on the location of ALS mutations and our hypothesis that hydrophobic interactions drive LLPS, we rationally introduced a tryptophan mutation at the position 334. This mutant does not disrupt the α-helix but abolishes LLPS, implying a potentially critical role of this residue.

2. Results

2.1. The chemical shifts of TDP-43$^{266–414}$ show an α-helix in the centre, and the disease-related mutants have similar secondary structural propensity

The aggregation-prone nature of TDP-43$^{266–414}$ hinders detailed structural characterization using solution state NMR spectroscopy. In our conditions (phosphate buffer at pH 6.5, 10 °C), 20 μM of sample is soluble for approximately four hours and even shorter at higher sample concentration. We overcome this challenge by using a generally applicable strategy, denaturation followed by titration, to assign this protein (described in Supplementary Text for detail and Fig. S2). According to the assignment, the secondary chemical shifts for residues 320–340, as predicted from the primary sequence by the PASTA [41] program (Fig. 1a), show that this region tends to form an α-helix (Fig. 1 b and c). Similarly, the program δ2D [42] predicts (based on the HN, N, Cα, Cβ, and C′ chemical shifts) an α-helical population for this domain of up to 80% (Fig. 1d). Our results agree with the previous observation [28,39], with the caveat that the buffer conditions and protein constructs are not identical in each case. Lim et al. mainly used pure water at pH 4.0 to reduce the aggregation rate [39], although they also demonstrated that similar chemical shifts (HSQC spectrum) were obtained at pH 6.8 in a 1 mM phosphate buffer. These authors only deposited NMR data for the α-helical segment (residues 307–347) in interaction with a micelle, making a direct comparison with our results meaningless. Conicella et al. [28] used conditions similar to ours (pH 6.1 versus pH 6.5 here) and obtained similar chemical shifts (BMRB accession number 26823 versus 26728 for ours). Although a few pairs of assignments are swapped [6], the discrepancies mostly occur at the N or C terminus and do not affect the conclusions between our studies.

We chose three disease-related mutants (two within the α-helical region: Q331K and M337V; one outside this region: G298S) to study these mutants' effects on the structural propensity. The chemical shifts of wild-type TDP-43$^{266–414}$ were used to assign the chemical shift of these mutants. These four variants show very similar HSQC spectra (Fig. 1e–k). In addition, the corresponding carbon chemical shifts (Fig. 1 l and m and Supplementary Fig. S4 a and b.) and circular dichroism (CD) spectra (Supplementary Fig. S4c–f) also show that these four variants have very similar secondary structure propensities.

2.2. The TDP-43$^{266–414}$ variants undergo LLPS at different temperatures and have different NMR signal intensity profiles

We observed that wild-type TDP-43$^{266–414}$ showed reversible LLPS when we lowered the temperature (Fig. 2 a and b; all samples returned to the original transparency either from observing the microscope image or from its turbidity). Although the structural propensity of these four variants is similar, these variants underwent LLPS under different conditions: We used buffer turbidity assays to investigate phase separation in these samples (20 μM) (Fig. 2c). The results showed that wild type TDP-43$^{266–414}$ and the G298S mutant similarly underwent LLPS below 283 K, whereas the Q331K sample was transparent at all temperatures. The M337V solution shows evidence of LLPS in between these two. We also observed that LLPS is favoured by higher protein concentrations (Fig. 2d). The LLPS of TDP-43 is previously reported but our conditions differ from Conicella et al.'s results: in their study, LLPS of this domain can only be induced when salt or RNA molecules are present [28] (see Discussion).

Before we discuss further using NMR to characterise these four variants, we emphasize the importance of our sample quality control here. Although two NMR studies on this domain have been published [28,39], contradicting observations (described above in the Introduction; the possible reasons are discussed below) may result from the different procedures of handling this aggregation-prone protein, as observed in several other well-studied systems such as α-synuclein and amyloid-β. We noticed the peak intensity profiles (intensity versus residue number) depended on the storage conditions (−80, −20, or 4 °C) and the time spent in solution before the measurements (for example Supplementary Fig. S5a). The chemical shifts did not obviously change however (Fig. S5b). It has been noticed that TDP-43 can form a hydrogel at low temperatures [39], as can several other proteins such as hnRNP A2 and FUS [43]. We found that TDP-43$^{266–414}$ turned into hydrogel even in 8 M urea when the protein concentration is high (above 500 μM). Furthermore, for proteins have similar fibril formation ability as α-synuclein [44], small amounts of aggregate or oligomeric forms can seed further aggregation. The results of TDP-43 C-terminal domain's studies are therefore dependent on the condition of the samples and the presence/absence of aggregates or stable oligomers. The focus of this report is on the behaviour of monomeric TDP-43$^{266–414}$, before it assembles irreversibly (into a hydrogel, oligomer, aggregate, or fibril). The monomeric nature of the protein was therefore controlled carefully between each biophysical measurement. Purified samples were lyophilized immediately after each HPLC run (see Methods and Supplementary Fig. S6). For each experiment, we dissolved some of the required amount of the lyophilized sample in a 10 mM pH 6.5 phosphate buffer. This was then centrifuged in an Eppendorf tube at 15,000g for 5 min to sediment any large particles present. The upper 95 vol% of the solution was then extracted and the absence of aggregates was confirmed by measuring the absorbance of the sample at ~340 nm [45]. Analytical ultracentrifugation was used to confirm the monomeric state of the protein (Supplementary Fig. S6). One-dimensional water-suppressed proton spectra were acquired before and after each NMR experiment to verify the conformation and concentration of the protein. The integrity of the protein was checked by SDS-PAGE after each experiment (Supplementary Fig. S6).

Under careful sample quality control, Fig. 3a shows the HSQC peak intensities as a function of residue number for each construct measured at five different temperatures. Although the peak intensity profiles are all alike and the intensity ratio between the mutants and the wild type is close to one at 298 K (Supplementary Fig. S7), the intensity profiles of

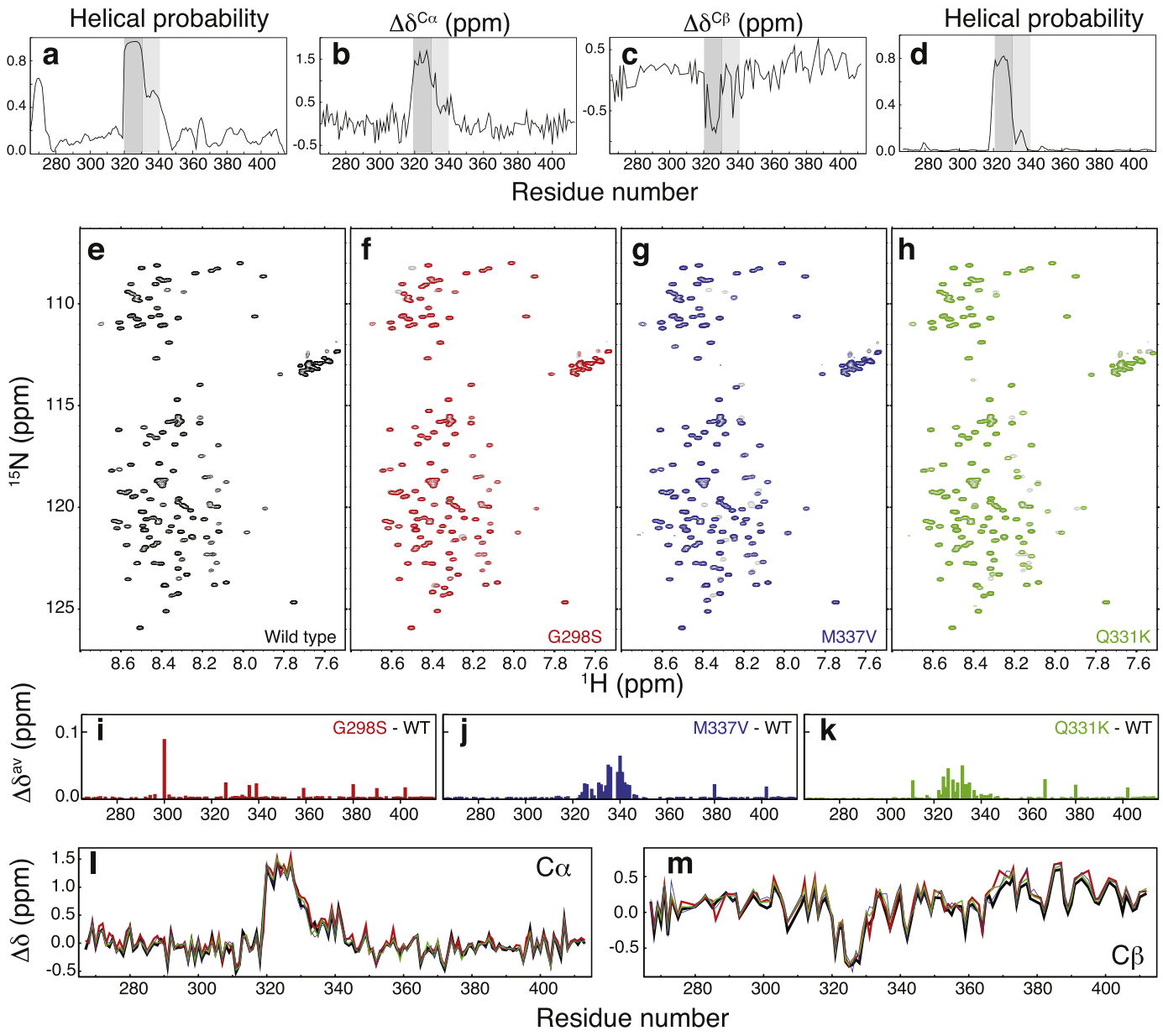

Fig. 1. Secondary chemical shifts and comparison of different mutants of TDP-43$^{266–414}$. (a) Residue-by-residue α-helical propensity of TDP-43$^{266–414}$ as predicted by the program PASTA [41]. (b,c) Secondary chemical shifts of (b) the Cα and (c) the Cβ atoms in TDP-43$^{266–414}$. (d) Residue-by-residue α-helical propensity of TDP-43$^{266–414}$ as predicted by the program δ2D [42] using the H^N, N, Cα, Cβ, and C′ chemical shifts. The regions corresponding to residues with a strong (residues 320–330) or weak (residues 331–340) helical propensity are respectively shaded in dark or light grey. ^{15}N-^{1}H HSQC NMR spectra of (e) the wild type and (f–h) the mutants (f) G298S, (g) M337V, and (h) Q331K. (i–k) The corresponding residue-by-residue average ^{15}N and ^{1}H chemical shift differences ($\Delta\delta^{av}$) between the wild type and the mutants. The spectrum of the wild type is also shown (in grey) beneath those of the mutants. (l) Cα and (m) Cβ secondary chemical shifts in the presence of 0.8 M urea. This amount of urea has little effect on this protein's secondary structure propensity (Supplementary Fig. S3 and S4).

the four constructs vary differently with temperature. We separated the peak intensities into three segments: N: residues 266–315; M: residues 318–338; and C: residues 351–414. We analysed the averaged intensity of these three segments and noticed that they all have similar relative values at the same temperature (Supplementary Fig. S8). Indeed, the intensity profiles of the mutants can be scaled to match that of wild type TDP-43$^{266–414}$ using a single factor for each mutant (Fig. 3b).

Although the signal intensity profiles between the Q331K variant and the wild type are very different at 278 K, their similar transverse relaxation rates (R_2) (Fig. 3c, the bottom panels for the wild type and the Q331K mutant) suggest a similar dynamic property, regardless of the presence of LLPS. The overall HSQC peak intensity of the wild-type is weaker at low temperatures because the droplet cannot be detected by NMR (because of its large molecular weight); in other words, only the molecules remaining in the monomeric or weakly populated multimeric form contribute to the detected NMR cross-peak intensity (also see Fig. 4c for a schematic illustration). The R_2s of the wild type in 40 μM at 288 K and 278 K (under such conditions the protein undergoes LLPS, Fig. 2d) show similar values comparing to the 20 μM sample (Fig. 3c), reinforcing that the presence of droplets has little effect on the dynamics of the protein molecules free from LLPS; the exchange rate is expected to be on a seconds timescale [25] between droplets and free molecules (i.e. monomers or transient multimers), and thus the exchange effect is negligible. Therefore, the overall decrease in peak intensity is greater for the wild type and the G298S mutant than for the other variants suggests that this decrease is because the number of molecules free from LLPS is reduced and fewer molecules contribute to the intensity in the HSQC spectrum.

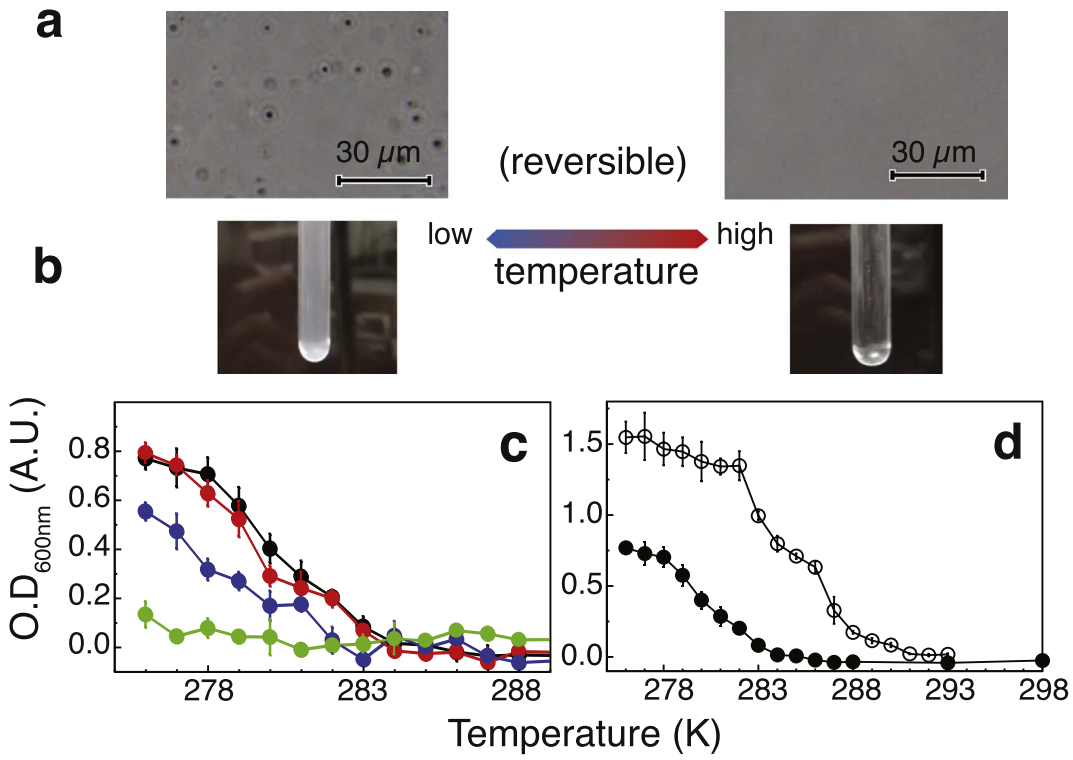

Fig. 2. Low temperature condition induces LLPS of TDP-43$^{266–414}$. (a) Microscopic images and (b) photographs showing the reversible liquid-liquid phase separation of wild type TDP-43$^{266–414}$. (c) Optical density at 600 nm ($O.D_{600nm}$) as a function of temperature for 20 μM samples of wild type (black), G298S (red), M337V (blue), and Q331K (green) variants. (d) $O.D_{600nm}$ as a function of temperature of 20 μM (solid black circle) and 40 μM (open black circle) wild type.

2.3. Concentration-dependent signal intensities indicate self-association

The peak intensities from the M segment, correspondent to the α-helical structure, of the four TDP-43$^{266–414}$ variants are similar to those of the C and N parts at 298 K, but decrease markedly more with decreasing temperature than those of the latter two sections (Fig. 3a). This is not due to the α-helix "melts" at higher temperatures because the carbon secondary chemical shifts at 298 K are very similar to those measured at 283 K (Supplementary Fig. S9 a and b), as are the populations of the helical conformation calculated from CD data (Supplementary Fig. S9 e and f). In addition, the temperature coefficients for the α-helical segment are lower in all four constructs (< 4.6 ppb/K) than those obtained for the other regions, indicating that there is a hydrogen bonding network in the central helical section (Supplementary Fig. S9 c and d) [46].

The decrease with decreasing temperature of the NMR peak intensities from the M segment hints at a kinetic effect, i.e. an exchange of different states at equilibrium. Our data indicate that this exchange occurs in the fast regime: no extra peaks appear during titration from a strongly acidic urea environment to physiological conditions and all the peaks are traceable (Fig. S2). Furthermore, the decrease in peak intensity for the M segment (Fig. 3 a and b) is consistent with the expected decrease in the exchange rate with reduced temperature. However, we could not exclude the effect that the conformation (e.g. multimeric form) with slower dynamics is also populated at lower temperatures.

This decrease in intensity can be explained by a helix-coil transition or by monomer-multimer self-association. The intensity ratio profiles measured for the wild type at concentrations of 20 and 40 μM (Fig. 4a and Supplementary Fig. S10) show that a helix-coil transition is not the dominant exchange mechanism: according to Le Châtelier's principle, since a helix-coil transition equilibrium has the same stoichiometric coefficient, it should not be affected by the sample concentration; in other words, the intensity ratios in Fig. 4a would be flat if a helix-coil transition dominated. On the contrary, at 298 and 293 K, although the increase in intensity of the peaks from the N and C segments matches the doubling of the protein concentration, the peaks from the M region strengthen much less (Fig. 4a). This can be explained by the higher protein concentration inducing a shift in the equilibrium toward the self-associated multimeric form (illustrated in Fig. 4c, upper panels.). There is furthermore a minor but systematic difference (< 0.010 ppm) between the average chemical shifts measured for the helical segment in the high- and low-concentration samples (Fig. 4b and Supplementary Fig. S11). This also indicates that some of the monomeric and multimeric forms exchange on a fast timescale. These results can be interpreted as follows (Fig. 4c): At high temperatures, the wild type TDP-43$^{266–414}$ are self-associated, the more so the higher the concentration is. At intermediate temperatures, self-association increases at low concentrations but droplets form at high protein concentrations (Fig. 4c, middle panels), which explains the lower-than-expected intensity ratios in Fig. 4a and in agreement with the results in Fig. 2d. At low temperature, the wild type phase separates at both concentrations (20 and 40 μM). Increasing the concentration only increases the proportion of protein in the droplet conformation: the amount of NMR-detectable protein does not change. Accordingly, NMR peak intensities, chemical shifts (Fig. 4 a and b, the 278 K panels), and R_2 (Fig. 3c) are similar at both concentrations (Fig. 4c, lower panels).

On the contrary instance of the Q331K mutant, which does not undergo LLPS at low temperatures (Fig. 2c), the concentration-dependency of the peak intensities is as expected: the increase in peak intensity as the protein concentration is doubled is around two for the N and C segments and slightly less for the M region (Fig. 4d and Supplementary Fig. S10), with small chemical shift perturbations observed in the latter for all five temperatures (Fig. 4e and Supplementary Fig. S11). The scenario for this instance is illustrated in Fig. 4f.

Our conclusion from the intensity ratio between different protein concentrations agrees the previous observation using intermolecular paramagnetic relaxation enhancement methods in their wild type

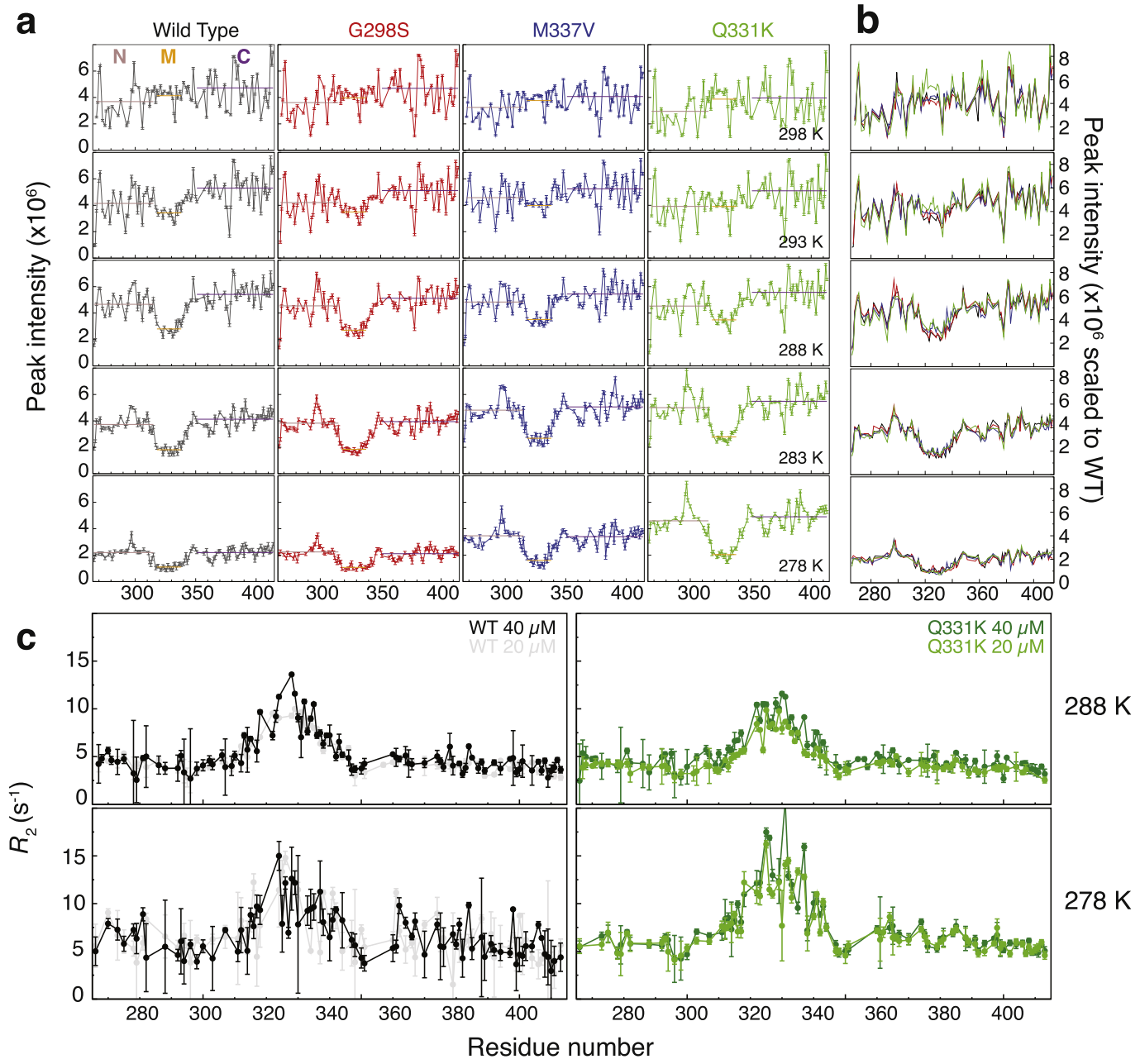

Fig. 3. The NMR peak intensity profiles and R_2s of TDP-43$^{266–414}$. (a) ^{15}N–^{1}H HSQC NMR peak intensity profiles of wild type TDP-43$^{266–414}$ (grey), and G298S (red), M337V (blue), and Q331K (green) variants (temperature: 278–298 K, protein concentration: 20 μM). The average intensity for the segments labelled N, M, and C, appear as brown, orange, and purple horizontal lines respectively. (b) The peak intensity profiles of different constructs scaled to the wild type's profile using the respective average intensity ratios. (c) Transverse relaxation rate constants (R_2) for wild type TDP-43$^{266–414}$ and the Q331K mutant at 20 and 40 μM and 278 and 288 K.

instance [28]. However, the direct comparison of intensity ratio may be a preferable way for accessing self-association for these aggregation-prone proteins. Introducing a point mutation for the hydrophobic nitroxide spin label (such as MTSL) [47] may affect the observed phenomenon, especially for those proteins sensitive to self-assembly. Furthermore, we have also demonstrated here that the Q331K mutant still self-associates (Fig. 4 d and e) regardless of its reduced LLPS tendency (Fig. 2c) from the intensity ratio analysis.

2.4. Hydrophobicity and electrostatic repulsive forces mediate LLPS

We used the Q331K mutant, which cannot undergo LLPS at lower temperatures (Fig. 2c), as an example to show that LLPS can be induced with adding 300 mM NaCl (Fig. 5a). Similarly for the wild type: the higher the concentration of salt is, the more droplets form (Fig. 5b). Because salt screens electrostatic interactions, these results suggest that electrostatic repulsion inhibits the formation of droplets.

Because the middle α-helical motif in TDP-43$^{266–414}$ is known as the hydrophobic patch [38] (Supplementary Fig. S1), it is possible that the observed LLPS is driven by hydrophobicity. To test this hypothesis, we added mild amount (80 mM) of urea, and observed that no droplet can form under such condition (Fig. 5c). Urea is a common compound used to denature proteins via disrupting its hydrophobic core [48,49]. Although the detailed mechanism on how urea unfolds protein is still under debate on the interaction and hydrogen bonding with water molecule or protein itself directly or indirectly [50–52], mild amount of urea can be used to disrupt protein-protein hydrophobic interaction or increase amyloid's solubility [53]. Furthermore, we confirmed that

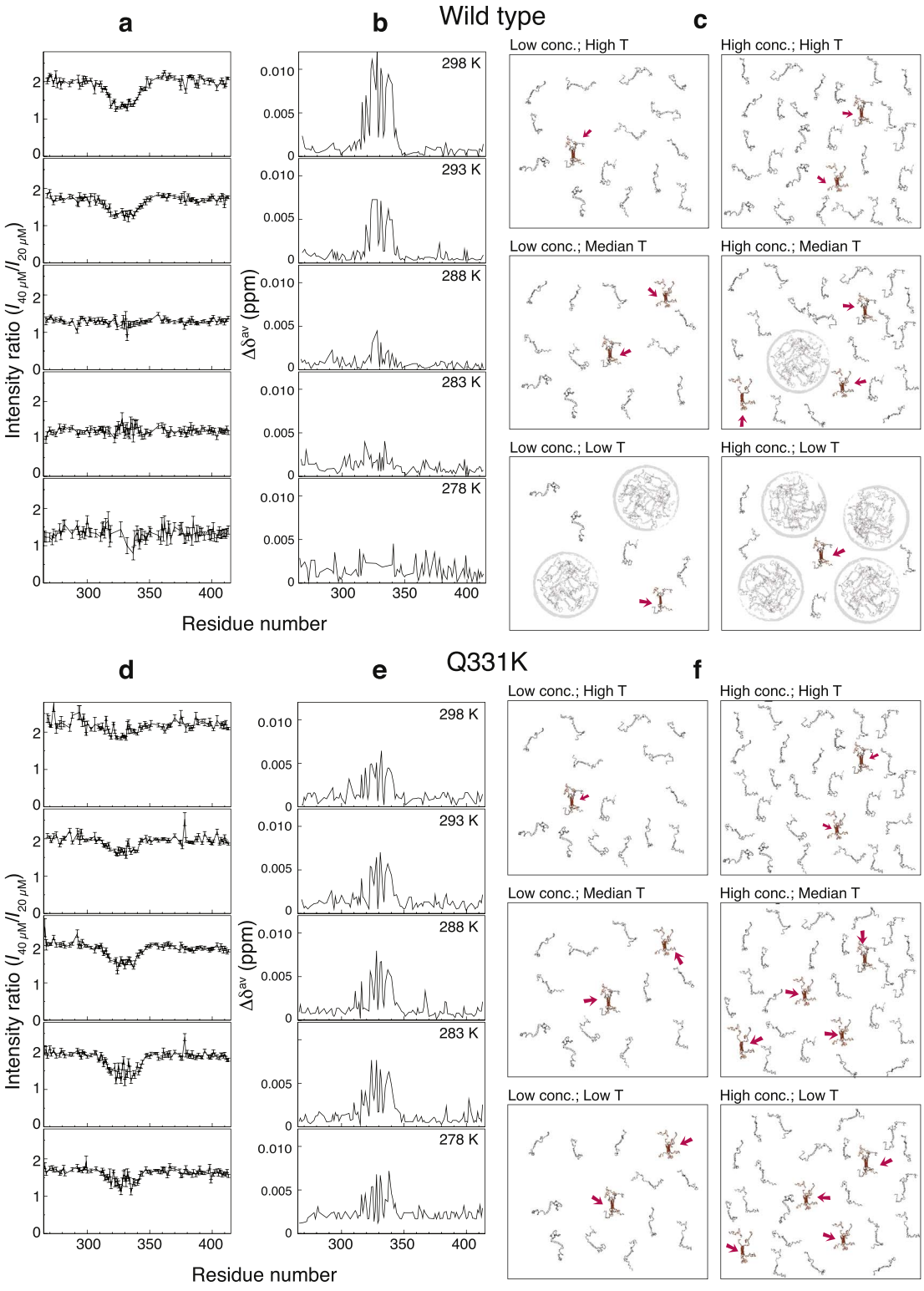

Fig. 4. The NMR peak intensity profiles and chemical shifts of TDP-43$^{266–414}$ vary with the protein concentration. Residue-by-residue (a,d) ratio of peak intensities and (b,e) average chemical shift difference measured between 40 and 20 μM samples of (a,b) wild type TDP-43$^{266–414}$ and (d,e) the Q331K mutant, at (from top to bottom) 298, 293, 288, 283, and 278 K. (c,f) Schematic illustration of the proposed models for self-association and liquid-liquid phase separation for (c) wild type TDP-43$^{266–414}$ and (f) the Q331K mutant. The self-associated form is indicated with red arrows. The LLPS form is shown in large grey circle.

adding 80 mM urea does not affect the α-helix in the middle of TDP-43$^{266–414}$ based on secondary chemical shift analysis (Supplementary Fig. S3); the concentration-dependent intensity ratio remains, indicating the protein still self-associates under this condition (Supplementary S12). We also used 5% of 1,6-hexanediol, an alcohol used to disrupt weak hydrophobic interaction [54], and observed no LLPS at all temperatures measured (Fig. 5c). This result agrees a recent study using the same compound to disrupt TDP-43 associated granules in neuron cells [55].

Hydrophobicity drives the formation of droplets whereas the

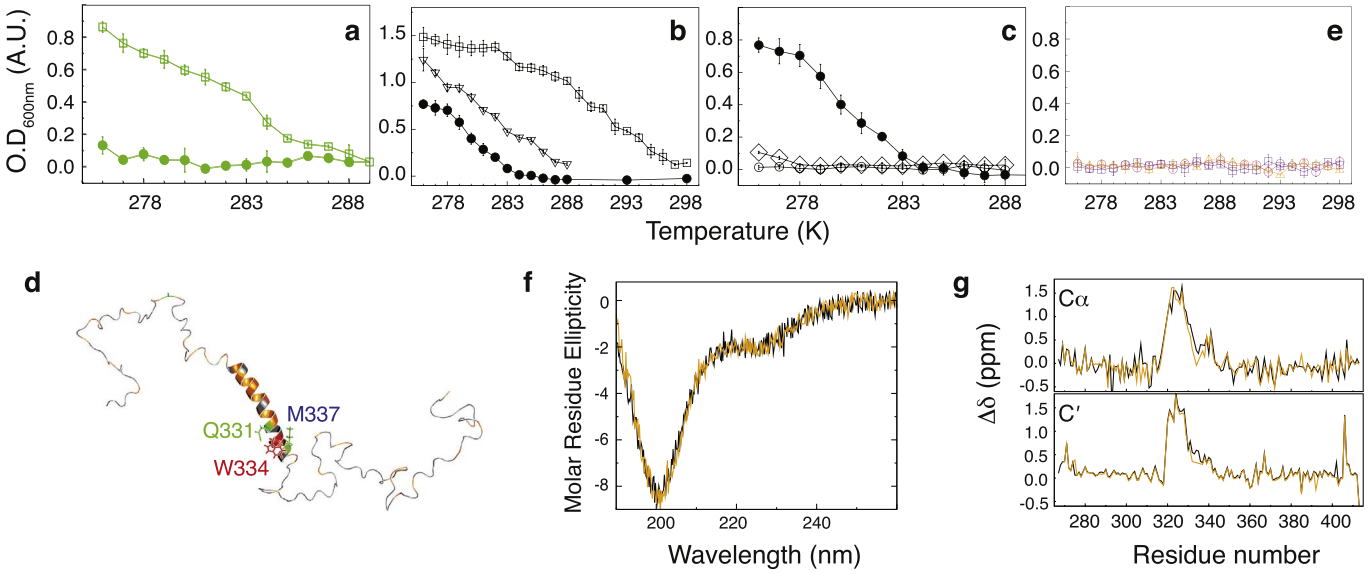

Fig. 5. The formation of TDP-43$^{266–414}$ droplets is governed by electrostatic and hydrophobic interactions. Optical density at 600 nm as a function of temperature for (a) the Q331K construct in the absence (green circles) and presence of NaCl (300 mM, open squares); (b) the wild type in the absence (black circles) and presence of NaCl (100 mM, open triangles; 300 mM, open squares); (c) the wild type in the absence (black circles), and presence of urea (80 mM, open diamonds), or in the presence of 1,6-hexanediol (5% open circles). (d) A cartoon molecular model shows that W334 (red stick) is close to M331 and Q331 (green stick). Orange represents hydrophobic residues (ala, val, phe, pro, met, ile, and leu). (e) Optical density at 600 nm as a function of temperature for 20 μM W334G mutant (orange), 40 μM W334G mutant (purple) samples or 20 μM W334G mutants with 300 mM NaCl (magenta). (f) The CD spectrum of the wild type (black) and the W334G mutant (orange) (20 μM, 288 K). (g) The secondary chemical shifts for Cα (upper panel) and C′ (lower panel) for the wild type (black) and the W334G mutant (orange).

electrostatic repulsive force inhibits this behaviour. TDP-43$^{266–414}$ has six positively and three negatively charged residues; the reason that the Q331K mutant is more difficult to form droplet is probably due to an extra positively charged residue that increases the net charge. The mildly reduced LLPS of the M337V variant (Fig. 2c) might be explained from its spatial position. From a putative structural model based on the secondary chemical shifts, we noticed that a tryptophan, the only aromatic amino acid in this α-helical region, is close to the position 331 and 337 (Fig. 5d). We thus hypothesize that this strong hydrophobic amino acid might be critical for LLPS of TDP-43$^{266–414}$; LLPS ability of M337V might be affected through disrupting its interference in between this tryptophan. Accordingly, we replaced a glycine at this position (W334G) and found that this mutant abolishes LLPS in those favoured conditions (high protein concentration, high salt buffer, or low temperature) (Fig. 5e). We also noticed that the self-association of this variant is significantly reduced but not totally abolished as observed from the intensity ratio between the 40 μM and 20 μM samples (Supplementary Fig. S14). It has been reported that this α-helical region is essential for the droplet formation of TDP-43, as LLPS is not observed when the α-helix is removed [13]. Our results on the contrary show that the α-helix remains in the W334G mutant from CD spectrum and secondary chemical shift analysis (Fig. 5 f and g). The study of the W334G variant strengthens the importance of hydrophobic driving force and hints a potentially important role of this residue.

3. Discussion

We used a generally applicable assignment strategy to obtain the NMR chemical shifts of the aggregation-prone TDP-43$^{266–414}$ (Fig. S2). Secondary chemical shift analysis shows that the middle part of this domain (~residues 320–340) has α-helical propensity, and disease-related mutants within this region (M337V and Q331K) have little effect on the secondary structure (Fig. 1). These mutants and the wild-type underwent LLPS at different temperatures however (Fig. 2). Our observation of LLPS at low temperatures (Figs. 2 and 3) contradicts to Conicella et al.'s results: in their study, LLPS could only occur in the presence of high concentration of salt or RNA molecules [28]. This apparent difference may stem from the different pH (6.1 versus 6.5 of ours) and buffer conditions used or from the design of the constructs. Indeed, a lower pH condition favours the net positive charge of this domain (increasing its repulsive interaction; this protein's pI is 10.78), in agreement with our model. Furthermore, a trace amount of urea, even left 1% from 8 M urea for dissolving inclusion bodies during the protein purification using a fast desalting procedure, may also result in the same, as we have demonstrated in Fig. 5c (note that urea was completely removed using our protocol and sample quality control (see above and Methods section)). It is noteworthy in this context that LLPS also occurs in a construct prepared without a hexahistidine tag (Supplementary Fig. S13; at pH 6.5, the net charge of hexahistidine is close to zero, as indicated also by simulations [40]).

For the wild type TDP-43$^{266–414}$ and the G298S mutant, droplet formation is accompanied by an overall decrease in NMR peak intensities (Fig. 3a) but without detectable line-broadening, as confirmed by NMR relaxation measurements (Fig. 3c). These results indicate that the droplet form does not affect the NMR observables because only the free molecules (not in the droplet) are detected and thus the overall NMR intensity decreases (Fig. 3a and see Fig. 4 c and d for an illustrative explanation). The weakening of the peaks in the middle segment of TDP-43$^{266–414}$ (part M, Fig. 3) is due to self-association (Fig. 4). The presence of LLPS leads the rest of the protein to be much diluted and thus to be less self-associated than those in the original sample (Fig. 4a–c).

It was concluded that the ALS-associated TDP-43 mutants disrupt its self-assembly or phase separation ability [28,36,39]. Those studies focused on the mutants located within or around the α-helix region. There are about 50 ALS-associated TDP-43 mutants reported to date spreading over its C-terminal domain [56] without a distinct pattern. To deduce this diverse mutational effect into one or two factors is still wait for further investigations as we have demonstrated here an example (G298S) that its similar biophysical observables to the wild type imply a different link to the disease in addition to the relation to LLPS (Figs. 2c, 3a, and Supplementary Fig. S7).

Using different buffer conditions, we prove that the droplet formation is a balance between hydrophobic and electrostatic forces (Fig. 5).

The Q331K mutant, with an increased net charge (from + 3 to + 4) of the molecule, has less LLPS ability because of the increased electrostatic repulsive force (Fig. 2c and 5a). On the other hand, the W334G variant, with the highly hydrophobic residue replaced, has a severely abolished LLPS tendency (but its secondary structure propensity remains) underpinning the importance of the hydrophobic driving force (Fig. 5e). Furthermore, this tryptophan may be crucial in LLPS. The Trp-334 is in between a serine and a glycine in the primary sequence (Supplementary Fig. S1), a typical [G/S]-[F/Y/W]-[G/S] motif for forming LLPS [16,20,43]. This tryptophan might also play a role in LLPS crosslinking. The protein's multivalency, however, is a key for forming the higher-order protein complex [25,57] and thus it is expected that the Trp-334 will not be the only residue involved in the formation of droplet. There are still six other such motifs in TDP-43$^{266–414}$; any of them may also contribute to LLPS at certain level. Extensive mutagenesis studies in different combination of these sites will elucidate more details on how LLPS forms.

In conclusion, this study shows that TDP-43$^{266–414}$ self-associates through its central α-helical domain. We demonstrated that the LLPS of this domain could be induced at a low temperature instead of in the presence of other molecules for coacervation, and we prove that LLPS of TDP-43$^{266–414}$ is via hydrophobicity. The only tryptophan located within the α-helical domain may be a key for LLPS. Droplet formation is associated with the protein's environment (salt concentration, temperature); sample conditions thus govern whether the protein undergoes a phase transition or not. It has very recently been shown that the TDP-43 granules that form in neuron cells are also specific to particular environments [55]. Our finding of the physical mechanisms governing TDP-43$^{266–414}$ LLPS may therefore shed light on the function of these granules and how they are related to neurodegenerative diseases.

4. Material and methods

4.1. DNA plasmids

The cDNA was derived from plasmid encoded human TDP-43 (provided by the Genomic Research Center, National Yang-Ming University). The gene encoding wild-type TDP-43 from residue 266 to 414 was inserted into a pET21a + -based vector with the restriction enzymes *Bam*HI and *Xho*I. The T7 tag on the vector was replaced with six histidines [6]. Point mutations (G298S, M337V, Q331K, W334G) were introduced using the appropriate primer design. All constructs were verified by DNA sequencing.

4.2. Protein expression and purification

The constructs were expressed using the same protocol as described elsewhere [6], except for Q331K, which was expressed at 37 °C for 6 h. The constructs were purified as follows. A pellet of inclusion bodies from cell lysate was dissolved in 20 mM Tris buffer with 8 M urea at pH 8.0. The dissolved solution was filtered with a 0.45 μM filter and then loaded onto a nickel-charged immobilized metal-ion affinity chromatography column (Qiagen, Inc.). The column was washed with ten times volume of 10 mM sodium phosphate buffer at pH 6.5 with 10 mM imidazole to remove non-binding proteins and urea. The target protein was eluted with the same buffer but with an additional 500 mM imidazole. The eluted sample was acidified using trifluoroacetic acid (down to pH ~3) and loaded into a C4 reverse phase column (Thermo Scientific, Inc) and eluted with an increasing gradient of acetonitrile mixed with triple-distilled water using an HPLC system. The eluted sample was frozen with liquid nitrogen and lyophilized immediately. The lyophilized sample was stored in a drying cabinet (relative humidity, ~20%) until it was used.

The required amount of protein was weighed and dissolved in 10 mM phosphate buffer at pH 6.5 with protease inhibitor cocktail (Roche, Inc). The concentration was also confirmed via the sample's absorbance at 280 nm measured using a NanoDrop spectrometer (Thermo Scientific, Inc.).

4.3. NMR experiments

Chemical shifts were assigned using standard HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB experiments acquired with non-uniform sampling [58,59]. ^{15}N-edited HSQC spectra were recorded using the standard pulse sequence with WATERGATE solvent suppression [60,61]. A one-dimensional proton spectrum was recorded with an improved WATERGATE solvent saturation scheme [62] before and after each NMR experiment to verify the sample's integrity. ^{15}N-transverse relaxation rates were measured using the standard experiments [63] using a series of relaxation time delays (17.2, 34.3, 51.5, 68.6, (85.8 in some cases), and 102.9 ms). The peak intensities were obtained via automated fitting using non-linear lineshapes. The resulting datasets were then fitted using exponential functions using the nmrPipe modelExp script [64]. All the HSQC spectra and dynamics experiments were recorded using a Bruker AVIII 600 MHz spectrometer with a cryogenic probe. The assignment experiments were acquired using Bruker AVIII 600 MHz or AVANCE 800 MHz spectrometers.

4.4. NMR data analysis

The collected data were processed using NMRPipe [64]. The assignments were made using Sparky [65]. Secondary chemical shift analysis was performed using Kjaergaard et al.'s database of random-coil shifts [66]. The non-linear line shape modeling (nlinLS) function in NMRPipe, including a Lorenz-to-Gauss window function, was applied to all HSQC spectra for peak intensity analysis, as described in the NMRPipe tutorial (https://spin.niddk.nih.gov/bax/software/NMRPipe/doc2new/). The average chemical shift difference ($\Delta\delta^{av}$) was calculated using:

$$\Delta\delta^{av} = \sqrt{\frac{(\Delta\delta_H)^2 + \left(\frac{1}{5}\Delta_N\right)^2}{2}}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the differences in chemical shift between two ^{1}H-^{15} N HSQC spectra for the amide proton and nitrogen, respectively.

4.5. Turbidity measurements

The turbidity of the protein samples was quantified by measuring the light transmittance at 600 nm using a JASCO V550 UV/VIS spectrophotometer. The spectrophotometer was temperature-controlled using a water bath. For each measurement (at each temperature), the sample was equilibrated in the temperature-controlled water bath for 5 min. Ten scans were accumulated for each measurement and measurements at each temperature point were repeated three times to estimate the associated errors. All measurements were returned to the beginning temperature to confirm the reversibility of droplet formation.

4.6. Circular dichroism spectroscopy

Circular dichroism spectra were recorded using a JASCO J-810 spectropolarimeter. Data were collected between 190 nm and 260 nm with an interval of 0.1 nm. Ten measurements were co-added for each data point. A 1.0 mm cuvette was used. The resulting curves were fitted using the program CDNN (Applied Photophysics, Inc.) to estimate the secondary structure population.

Transparency Document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

The authors thank Mr. W.-R. Chu for preliminary work on this project, Professors Shang-Te Danny Hsu, Joseph Jen-Tse Huang (Academia Sinica), and Ta-Hsien Lin (National Yang-Ming University) for helpful comments, the Core Facility for Protein Structural Analysis in Academia Sinica and the Instrumentation Center in National Taiwan University for access to NMR spectrometers, and Professor Won-Jing Wang (National Yang-Ming University) for the help of using the inverted microscope. This work was supported by the Ministry of Science and Technology of Taiwan (104-2113-M-010-001-MY2, 106-2113-M-010-005-MY2).

Author contributions statement

H.R.L., T.C.C., C.L.H., and L.S. conducted the experiments. H.R.L., T.C.C., C.L.H., L.S., C.Y.C., and J.R.H. analysed the results. H.R.L. and J.R.H. conceived the idea. J.R.H. obtained the funding and wrote the paper. All authors reviewed the manuscript.

Additional Information

The authors declare no competing financial interest.

References

[1] P.H. Kuo, et al., Structural insights into TDP-43 in nucleic-acid binding and domain interactions, Nucleic Acids Res. 37 (6) (2009) 1799–1808.
[2] F. He, et al., Solution Structure of Rrm Domain in Tar Dna-binding Protein-43, (2004) (Unpublished PDB: 1WFO).
[3] M. Mompean, et al., The TDP-43 N-terminal domain structure at high resolution, FEBS J. 283 (7) (2016) 1242–1260.
[4] H. Qin, et al., TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA, Proc. Natl. Acad. Sci. U. S. A. 111 (52) (2014) 18619–18624.
[5] A.K. Chen, et al., Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis, J. Am. Chem. Soc. 132 (4) (2010) 1186–1187.
[6] T.C. Chen, et al., The nearest-neighbor effect on random-coil NMR chemical shifts demonstrated using a low-complexity amino-acid sequence, Protein Pept. Lett. 23 (11) (2016) 967–975.
[7] M. Neumann, et al., Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, Science 314 (5796) (2006) 130–133.
[8] L.M. Igaz, et al., Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies, J. Biol. Chem. 284 (13) (2009) 8516–8524.
[9] Y.T. Wang, et al., The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates, J. Biol. Chem. 288 (13) (2013) 9049–9057.
[10] Y.C. Huang, et al., Inhibition of TDP-43 aggregation by nucleic acid binding, PLoS One 8 (5) (2013) e64002.
[11] C. Capitini, et al., TDP-43 inclusion bodies formed in bacteria are structurally amorphous, non-amyloid and inherently toxic to neuroblastoma cells, PLoS One 9 (1) (2014) e86720.
[12] E.B. Lee, V.M. Lee, J.Q. Trojanowski, Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration, Nat. Rev. Neurosci. 13 (1) (2012) 38–50.
[13] H.B. Schmidt, R. Rohatgi, In vivo formation of vacuolated multi-phase compartments lacking membranes, Cell Rep. 16 (5) (2016) 1228–1236.
[14] A. Molliex, et al., Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization, Cell 163 (1) (2015) 123–133.
[15] A. Patel, et al., A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation, Cell 162 (5) (2015) 1066–1077.
[16] K.A. Burke, et al., Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II, Mol. Cell 60 (2) (2015) 231–241.
[17] T.J. Nott, et al., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles, Mol. Cell 57 (5) (2015) 936–947.
[18] T.J. Nott, T.D. Craggs, A.J. Baldwin, Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters, Nat. Chem. 8 (6) (2016) 569–575.
[19] H. Jiang, et al., Phase transition of spindle-associated protein regulate spindle apparatus assembly, Cell 163 (1) (2015) 108–122.
[20] Y. Lin, et al., Formation and maturation of phase-separated liquid droplets by RNA-binding proteins, Mol. Cell 60 (2) (2015) 208–219.
[21] S. Xiang, et al., The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei, Cell 163 (4) (2015) 829–839.
[22] H. Zhang, et al., RNA controls PolyQ protein phase transitions, Mol. Cell 60 (2) (2015) 220–230.
[23] C.W. Pak, et al., Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, Mol. Cell 63 (1) (2016) 72–85.
[24] J.A. Riback, et al., Stress-triggered phase separation is an adaptive, evolutionarily tuned response, Cell 168 (6) (2017) 1028–1040 (e19).
[25] H. Wu, M. Fuxreiter, The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules, Cell 165 (5) (2016) 1055–1066.
[26] J.A. Toretsky, P.E. Wright, Assemblages: functional units formed by cellular phase separation, J. Cell Biol. 206 (5) (2014) 579–588.
[27] L. Guo, J. Shorter, It's raining liquids: RNA tunes viscoelasticity and dynamics of membraneless organelles, Mol. Cell 60 (2) (2015) 189–192.
[28] A.E. Conicella, et al., ALS mutations disrupt phase separation mediated by alpha-helical structure in the TDP-43 low-complexity C-terminal domain, Structure 24 (9) (2016) 1537–1549.
[29] L.L. Jiang, et al., Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation, Sci Rep 6 (2016) 23928.
[30] R.Y. He, et al., Characterization and real-time imaging of the FTLD-related protein aggregation induced by amyloidogenic peptides, Chem. Commun. (Camb.) 51 (41) (2015) 8652–8655.
[31] M. Mompean, et al., Structural characterization of the minimal segment of TDP-43 competent for aggregation, Arch. Biochem. Biophys. 545 (2014) 53–62.
[32] M. Mompean, et al., Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43, J. Phys. Chem. Lett. 6 (13) (2015) 2608–2615.
[33] A. Saini, V.S. Chauhan, Delineation of the core aggregation sequences of TDP-43 C-terminal fragment, Chembiochem 12 (16) (2011) 2495–2501.
[34] A. Saini, V.S. Chauhan, Self-assembling properties of peptides derived from TDP-43 C-terminal fragment, Langmuir 30 (13) (2014) 3845–3856.
[35] L. Zhu, et al., An ALS-mutant TDP-43 neurotoxic peptide adopts an anti-parallel beta-structure and induces TDP-43 redistribution, Hum. Mol. Genet. 23 (25) (2014) 6863–6877.
[36] W. Guo, et al., An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity, Nat. Struct. Mol. Biol. 18 (7) (2011) 822–830.
[37] C.S. Sun, et al., The influence of pathological mutations and proline substitutions in TDP-43 glycine-rich peptides on its amyloid properties and cellular toxicity, PLoS One 9 (8) (2014) e103644.
[38] L.L. Jiang, et al., Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion, J. Biol. Chem. 288 (27) (2013) 19614–19624.
[39] L. Lim, 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 (1) (2016) e1002338.
[40] M. Mompean, et al., Electrostatic repulsion governs TDP-43 C-terminal domain aggregation, PLoS Biol. 14 (4) (2016) e1002447.
[41] I. Walsh, et al., PASTA 2.0: an improved server for protein aggregation prediction, Nucleic Acids Res. 42 (2014) W301–7 (Web Server issue).
[42] C. Camilloni, et al., Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts, Biochemistry 51 (11) (2012) 2224–2231.
[43] M. Kato, et al., Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels, Cell 149 (4) (2012) 753–767.
[44] C.A. Ross, M.A. Poirier, Protein aggregation and neurodegenerative disease, Nat. Med. 10 (Suppl) (2004) S10–7.
[45] B. Raynal, et al., Quality assessment and optimization of purified protein samples: why and how? Microb. Cell Factories 13 (2014) 180.
[46] T. Cierpicki, J. Otlewski, Amide proton temperature coefficients as hydrogen bond indicators in proteins, J. Biomol. NMR 21 (3) (2001) 249–261.
[47] T. Mittag, J.D. Forman-Kay, Atomic-level characterization of disordered protein ensembles, Curr. Opin. Struct. Biol. 17 (1) (2007) 3–14.
[48] C. Tanford, Isothermal unfolding of globular proteins in aqueous urea solutions, J. Am. Chem. Soc. 86 (10) (1964) 2050-&.
[49] R. Zangi, R. Zhou, B.J. Berne, Urea's action on hydrophobic interactions, J. Am. Chem. Soc. 131 (4) (2009) 1535–1541.
[50] L.B. Sagle, et al., Investigating the hydrogen-bonding model of urea denaturation, J. Am. Chem. Soc. 131 (26) (2009) 9304–9310.
[51] J.A. Schellman, The thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond, C. R. Trav. Lab. Carlsberg Chim. 29 (14–15) (1955) 223–229.
[52] J.R. Huang, et al., Sequence-specific mapping of the interaction between urea and unfolded ubiquitin from ensemble analysis of NMR and small angle scattering data, J. Am. Chem. Soc. 134 (9) (2012) 4429–4436.
[53] J.R. Kim, et al., Urea modulation of beta-amyloid fibril growth: experimental studies and kinetic models, Protein Sci. 13 (11) (2004) 2888–2898.
[54] S.S. Patel, et al., Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex, Cell 129 (1) (2007) 83–96.
[55] P.P. Gopal, et al., Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons, Proc. Natl. Acad. Sci. U. S. A. 114 (12) (2017) E2466–E2475.
[56] Y. Sun, A. Chakrabartty, Phase to phase with TDP-43, Biochemistry 56 (6) (2017) 809–823.
[57] P. Li, et al., Phase transitions in the assembly of multivalent signalling proteins, Nature 483 (7389) (2012) 336–340.
[58] S.G. Hyberts, et al., Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling, J. Biomol. NMR 52 (4) (2012) 315–327.
[59] S.G. Hyberts, et al., FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation, J. Biomol. NMR 45 (3) (2009)

283–294.
[60] M. Piotto, V. Saudek, V. Sklenar, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, J. Biomol. NMR 2 (6) (1992) 661–665.
[61] G. Bodenhausen, D.J. Ruben, Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy, Chem. Phys. Lett. 69 (1980) 185–189.
[62] M. Liu, et al., Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy, J. Magn. Reson. 132 (1) (1998) 125–129.
[63] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, Biochemistry 28 (23) (1989) 8972–8979.
[64] F. Delaglio, et al., NMRPipe: A multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR 6 (3) (1995) 277–293.
[65] T.D. Goddard, D.G. Kneller, Sparky 3, University of California, San Francisco, 2005.
[66] M. Kjaergaard, S. Brander, F.M. Poulsen, Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, J. Biomol. NMR 49 (2) (2011) 139–149.

Supplementary Data

The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43

Hao-Ru Li[1,†], Tsai-Chen Chen[1,†], Chih-Lun Hsiao[1], Lin Shi[2], Chi-Yuan Chou[2], and Jie-rong Huang[1,3,*]

[1]Institute of Biochemistry and Molecular Biology, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan
[2]Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan
[3]Institute of Biomedical Informatics, National Yang-Ming University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan
[*]jierongh@ym.edu.tw
[+]these authors contributed equally to this work

Supplementary Text

Assignment Strategy

We used TDP-43 as a model protein with low sequence complexity in our previous work on the nearest-residue effect on NMR chemical shifts[1]. The chemical shifts for the denatured and physiological states were deposited in the Biological Magnetic Resonance Bank (BMRB, accession numbers: 26816 and 26728, respectively). Our assignment strategy is unpublished and generally applicable to other aggregation-prone proteins so we describe it here in more detail. In the first step, the chemical shifts are assigned in 10 mM glycine buffer at pH 2.5 in the presence of 8 M urea. Under these conditions, the protein is highly soluble (concentration ~1 mM). The increased experimental time caused by maximizing the resolution in the indirect dimensions (^{15}N and ^{13}C) can be compensated by using non-uniform sampling[2,3]. For TDP-43, 100% of the backbone chemical shifts were assigned in this way (Fig. S2a). We titrated the buffer and mapped the chemical shifts through ^{15}N-^{1}H HSQC spectra up to the physiological pH (Fig. S2b). We also collected Cα, Cβ and C' chemical shifts in the presence of 0.8 and 0.08 M urea to confirm the chemical shift mapping and significantly reduce the ambiguity of the sequential chemical shift assignments. As shown in Fig. S2c and S2d, connections that are difficult to identify in the data from the 0.8 M urea sample because of low signal-to-noise (S/N, e.g. S333) or low resolution in the indirect dimension (e.g. S410) are readily identified in the equivalent spectra acquired with 8 M urea. We finally assigned the HSQC spectrum under physiological conditions (Fig. S2e).

Additional Methods

Protein expression and purification for the construct prepared without a hexahistidine tag. A TEV protease-cutting site was inserted between the hexahistidine tag and TDP-43$^{266\text{-}414}$. To obtain this construct, we cloned the target sequence in two steps. In the first, we used two designed primers (forward: 5'-GAGAACCTGTACTTCCAGAGCAATAGACAGTTAGAAAGAAGT-3'; reverse: '5-ACCTCGAGTCATTACATTCCCCAGC-3') to obtain TDP-43$^{266\text{-}414}$ with TEV protease recognition sequence at the N-terminus. In the second, we used other primers (forward: 5'-ATGGCTAGCCATCATCATCATCATCATGAGAACCTGTACTTC CAG-3'; reverse: '5-ACCTCGAGTCATTACATTCCCCAGC-3') to insert the

hexahistidine tag at the N-terminus of the cDNA created in the first step. With this primer deign, only the native sequence is left after TEV-protease digestion.

The protein expression and purification steps are similar to those described for the tagged sample. The pure protein was eluted with 500 mM imidazole in a 10 mM phosphate buffer with 8 M urea at pH 6.5 from a nickel-charged immobilized metal ion affinity chromatography (IMAC) column. The imidazole and urea were removed using a desalting column (PD-10, GE Health Care, Inc.). The hexahistidine-tagged TEV protease was added to this protein solution in a 1:10 ratio and left at room temperature for 4 h. The solution was loaded onto a nickel-charged IMAC column once more; the flow-through was then collected and acidified with trifluoroacetic acid in preparation for HPLC as described for the tagged constructs.

There are no significant differences between the chemical shifts measured for the tagged and untagged construct. (Fig. S3a–c) or between those of the tagged construct in the presence (80 mM) or absence of urea (Fig. S3d–f). In other words, the helical propensity revealed by these data is not affected or caused by urea or the His tag. The chemical shift assignment of the mutants was therefore performed with 80 mM urea samples to maintain the proteins' solubility[4] and allow for longer experiments while increasing the sensitivity in the carbon dimension (as highlighted for S333 in Fig. 2d for example). Other two-dimensional NMR experiments were performed on hexahistidine-tagged constructs (with greater yields) in the absence of urea.

Analytical ultracentrifugation. Measurements were performed on a Beckman Coulter XL-A analytical ultracentrifuge with an An-50 Ti rotor and a standard 12-mm double-sector epon charcoal-filled centerpieces. After checking for sample leakage at 3,000 rpm, the rotation speed was increased to 42,000 rpm and the absorbance at 280 nm was detected for 12 h. All data were analysed using the program SEDFIT[5].

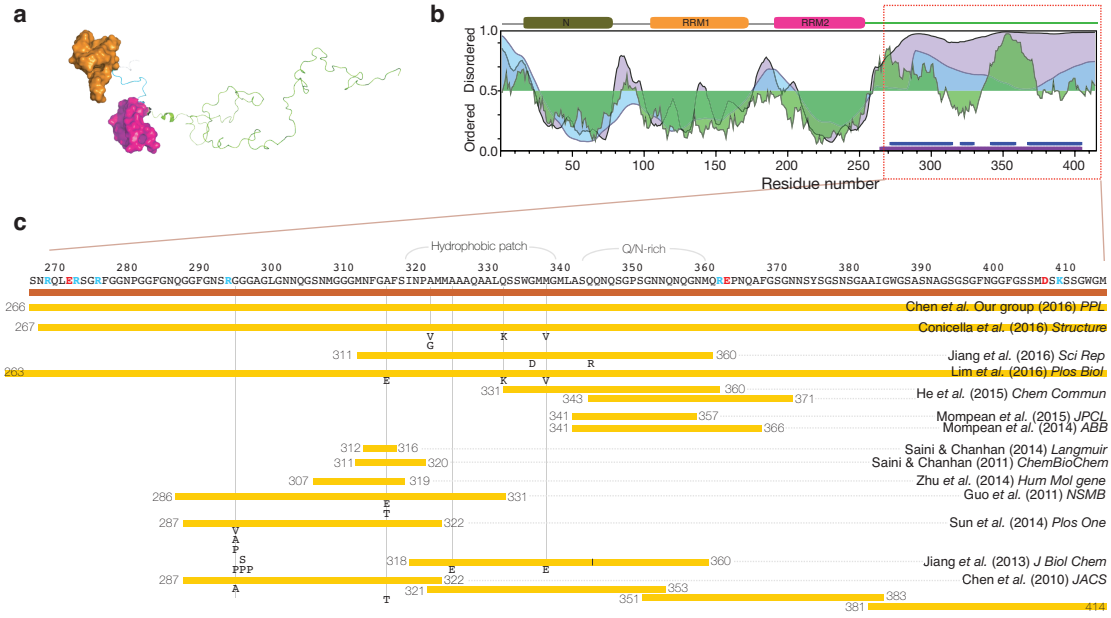

Fig. S1. TDP-43 domains and a review of structural studies of the C-terminal domain. (a) Surface model of full-length TDP-43: the two structured RNA recognition motifs (PDB IDs, 2CQG and 3D2W) are shown in orange and magenta respectively, the linker between the two, in blue, and the intrinsically disordered C-terminus in green. (b) Disorder predictions from the programs PONDR VSL2 (purple) and VL3 (blue)[6,7], and IUPRED (green)[8]. Sequence complexity was calculated using the SEG algorithm[9], revealing the low complexity regions LC1 and LC2 indicated with blue and purple bars respectively. The C-terminal domain has low sequence complexity and is predicted to be intrinsically disordered. (c) Fragments of the C-terminal domain of TDP-43 whose structure has been studied (orange bars) alongside the corresponding reference. Letters indicate the mutations studied. The blue and red letters in the amino acid sequence of TDP-43$^{266–414}$ (upper part of the panel) respectively highlight positively and negatively charged residues.

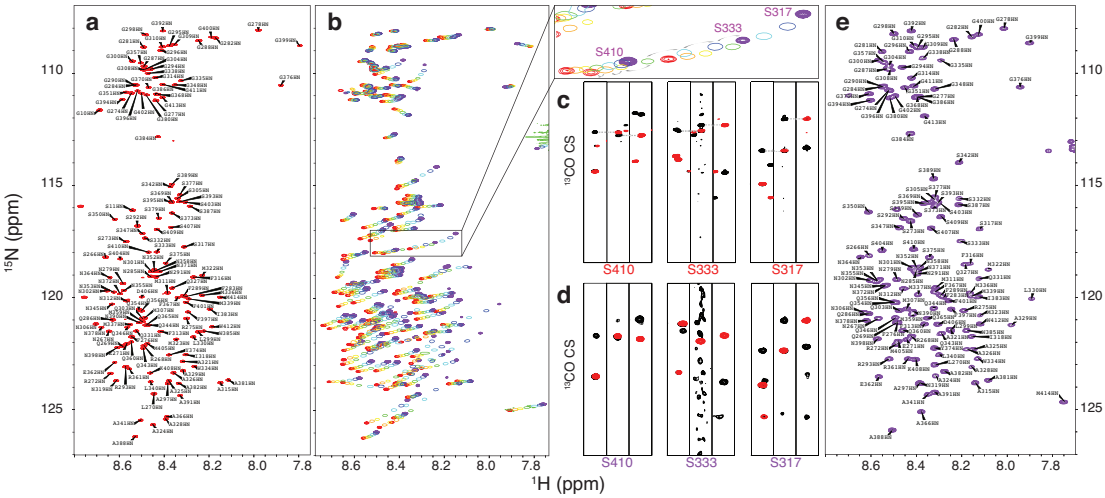

Fig. S2. Chemical shift assignments strategy. (a) Assigned ^{15}N-^{1}H HSQC NMR spectrum of TDP-43$^{266–414}$ at pH 2.5 in the presence of 8 M urea. (b) Tracking the chemical shift assignments of TDP-43$^{266–414}$ from an acidic urea environment to physiological conditions via ^{15}N-^{1}H HSQC spectra. The expanded view (upper right) shows the cross-peaks from residues S410, S333, and S317. (c,d) Sequential connections (dashed lines) across HNCO (red) and HN(CA)CO (black) spectra between the residues highlighted in Panel b. (c) in the presence of 8 M urea at pH 2.5 and (d) under physiological conditions. (e) Assigned ^{15}N-^{1}H HSQC spectrum of TDP-43$^{266–414}$ at pH 6.5 in the absence of urea.

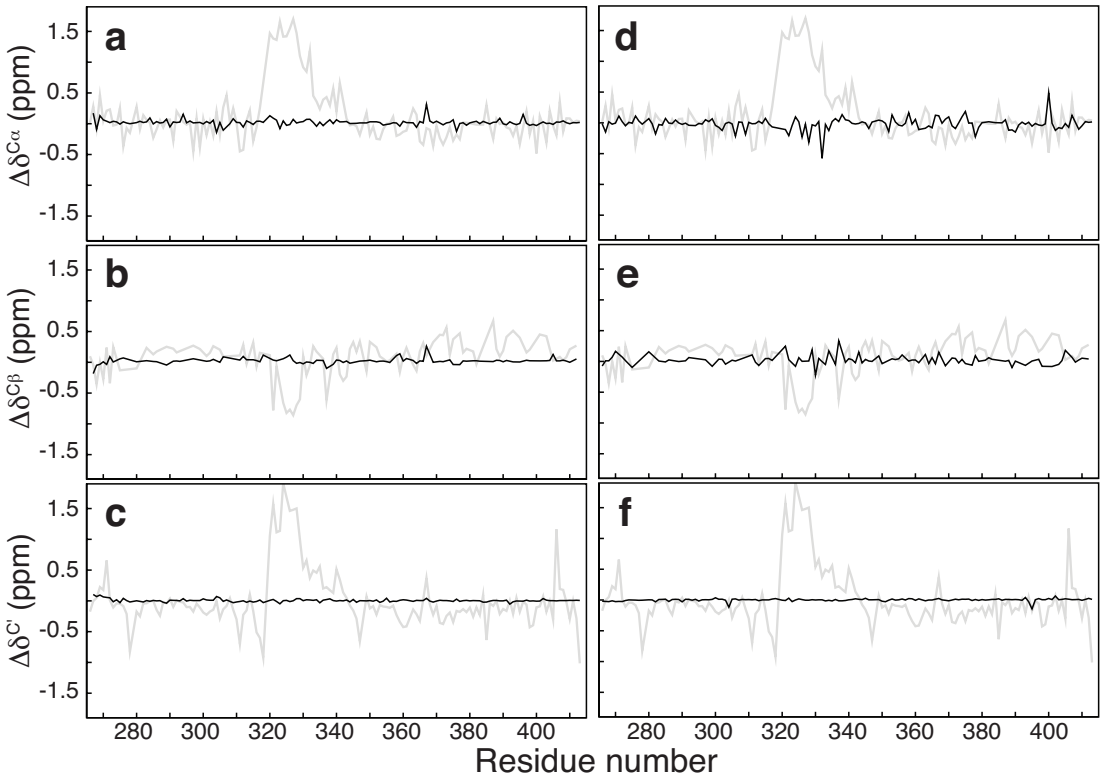

Fig. S3. The chemical shifts (secondary structure propensity) of TDP-43$^{266–414}$ are not affected by the presence of a hexahistidine tag or a small amount of urea. Difference (black lines) between the chemical shifts measured for TDP-43$^{266–414}$ (a–c) in constructs with and without a hexahistidine tag and (d–f) in the presence and absence of 0.08 M urea, compared with the secondary chemical shifts as shown in Fig. 1 b and c (grey lines).

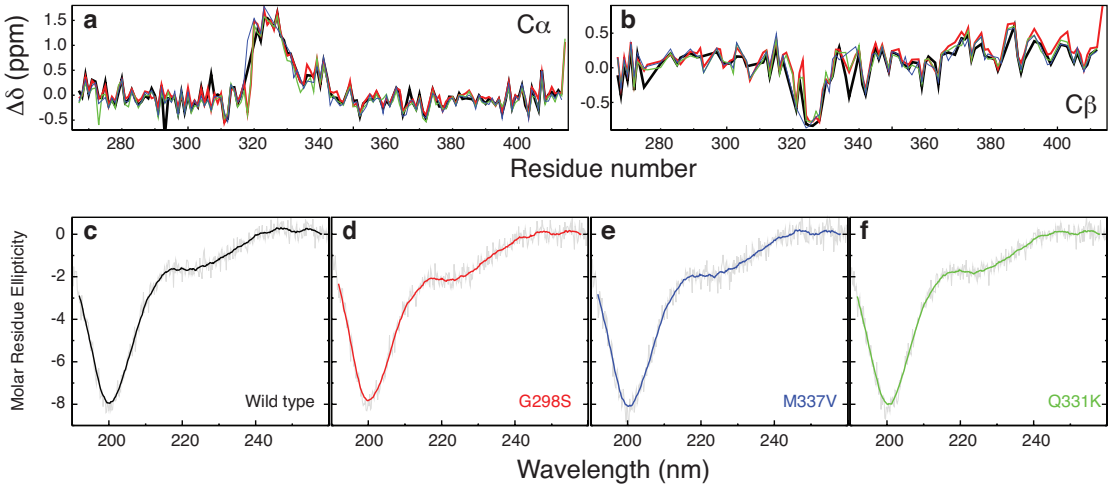

Fig. S4. (a,b) The secondary chemical shifts and (c–f) circular dichroism spectra of wild type TDP-43$^{266–414}$ and three mutants. The chemical shifts in Panels a and b were recorded in samples containing 80 mM urea. The resolution in the carbon dimension and the signal-to-noise is worse than under 0.8 M urea (Fig. 1i and j in the main text) but the conclusion is the same: the three mutations studied do not alter the secondary structure propensity of TDP-43$^{266–414}$. In the circular dichroism spectra (Panels c–f), the grey lines are the raw data, and the coloured lines represent the sliding average over 20 data points. The secondary structure propensities calculated using the program CDNN for the four variants are also very similar (see Fig. S9e and f).

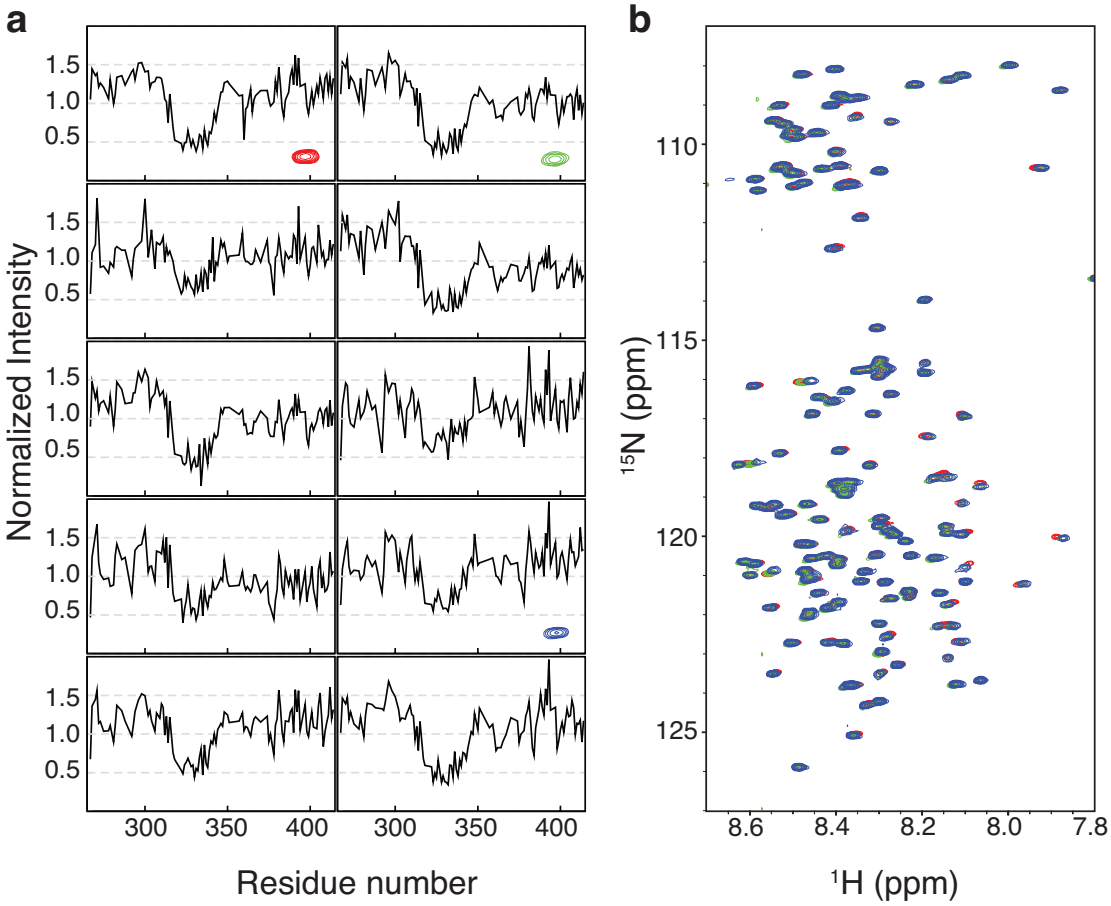

Fig. S5. NMR peak intensity profiles and spectra of TDP-43$^{266–414}$ in identical buffers depend nevertheless on the sample preparation and storage conditions. (The chemical shifts do not vary much.) (a) ^{15}N-HSQC peak intensities versus residue number for different TDP-43$^{266–414}$ samples. The intensities are scaled to the averaged intensity of the N-terminal region (residues 266–315) to facilitate comparison. The relative intensities of peaks from the middle and the C-terminal segments vary. (b) Three overlaid ^{15}N-HSQC spectra from samples with different intensity profiles, shown respectively in red, green (weaker peaks in the C-terminal region) or blue (stronger peaks in the middle segment), as highlighted in Panel a.

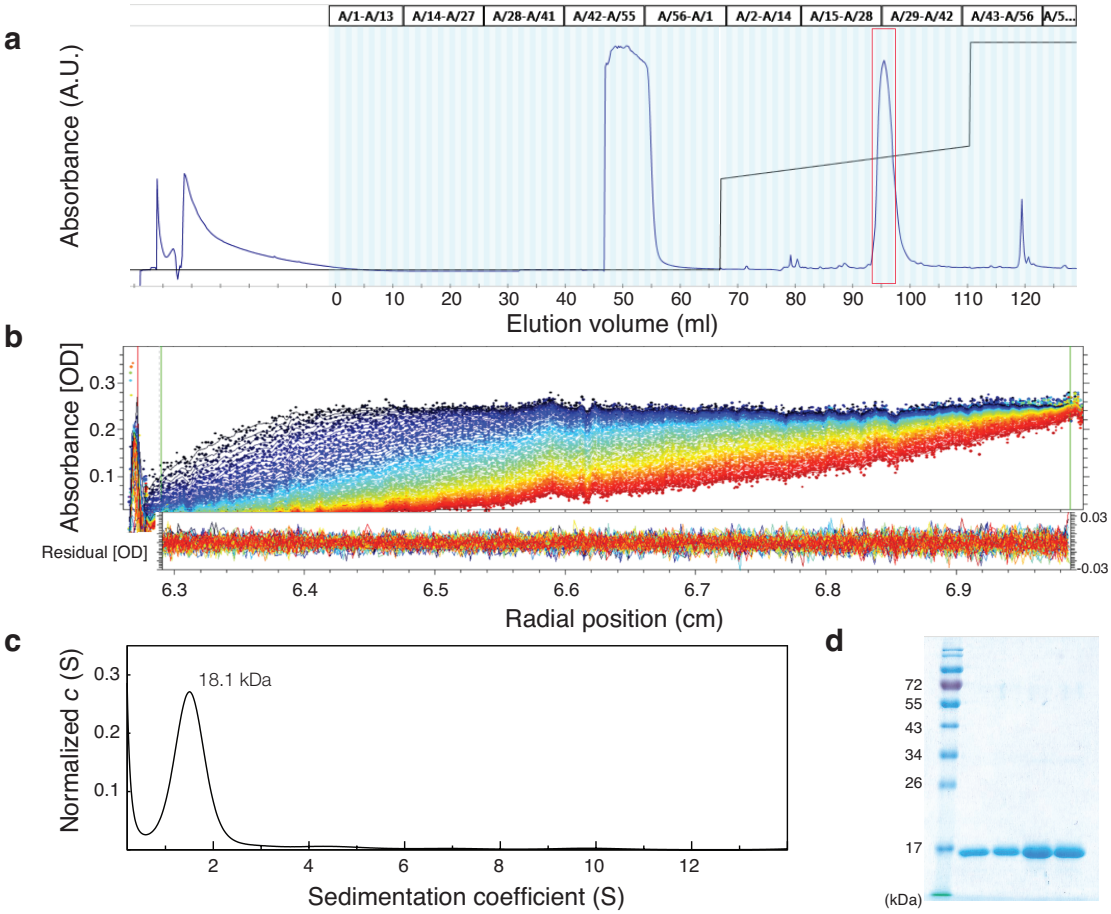

Fig. S6. Sample quality control. (a) A typical HPLC elution profile. The blue line represents the absorbance at 215 nm and the black line shows the gradient of acetonitrile in the elution water. The red box highlights the fractions collected for lyophilization. (b) Raw analytical ultracentrifugation (AUC) data above the fitting residuals (data fitted using SEDFIT[5] with an RMSD of 0.0074). (c) The AUC analysis confirms that after dissolution in phosphate buffer the lyophilized sample is still monomeric. Because the protein is mainly disordered (rather than globular), the derived protein size is larger than the value predicted from the sequence (18.1 versus 15.9 kDa). (d) Photograph of a typical SDS-PAGE gel used to confirm the integrity of the samples after each measurement. Each lane corresponds to one of the four variants of TDP-43$^{266–414}$ studied here (wild type, G298S, M337V, Q331K).

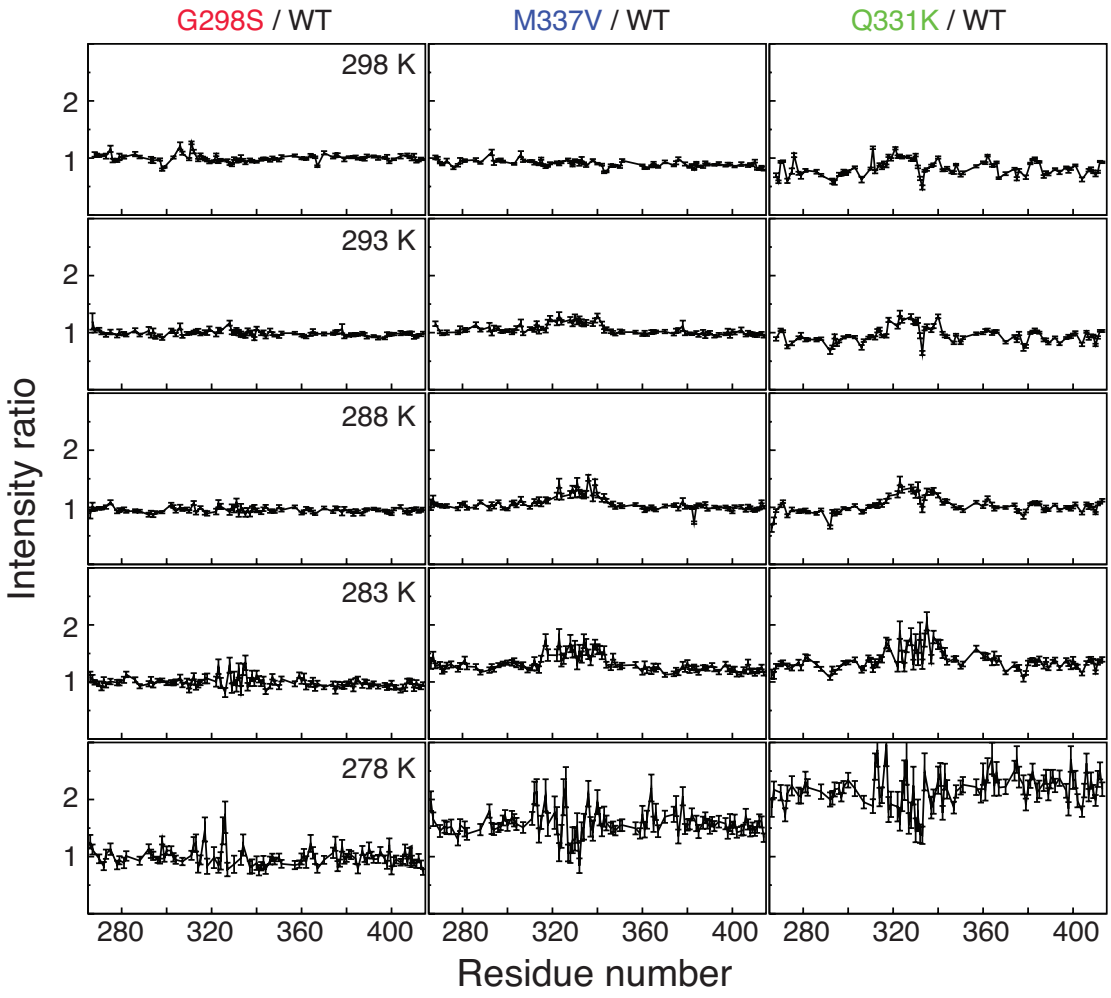

Fig. S7. ^{15}N-HSQC peak intensities of TDP-43$^{266–414}$ mutants relative to those of the wild type. The ratio between mutants and the wild type (WT) at different temperatures. At high temperature (298 K) the intensity ratios are all close to one. The G298S mutant has a similar intensity profile to that of the wild type at all temperatures. At lower temperatures (283 and 278 K), the M337V/WT and Q331K/WT ratios are greater than one (i.e. the signal from the wild type is weaker on average).

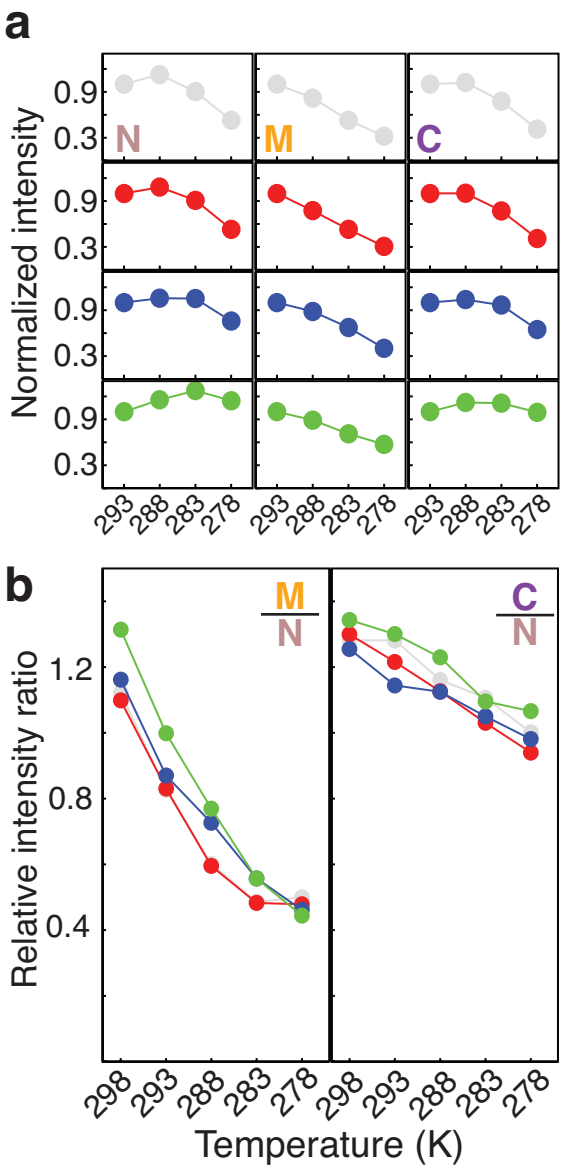

Fig. S8. (a) Average intensity (normalized to the corresponding value at 293 K) of HSQC peaks from the N, M, and C segments of the four variants as a function of temperature. (b) Relative intensities of the M and N parts (M/N) and of the C and N parts (C/N) of the same four variants as a function of temperature.

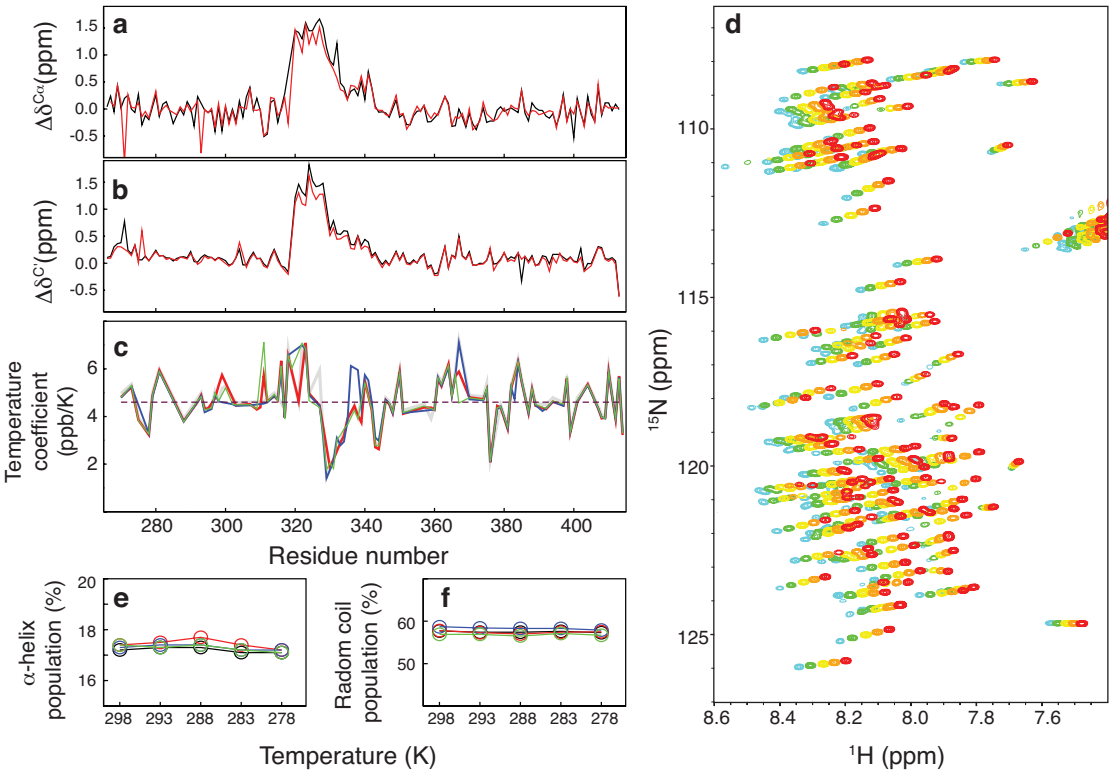

Fig. S9. The α-helical component of TDP-43$^{266–414}$ at different temperatures. (a) Cα and (b) C' secondary chemical shifts of wild type TDP-43$^{266–414}$ at 298 (red lines) and 283 K (black). The similar profiles show that the protein's α-helical propensity is the same at both temperatures. (c) Similar temperature coefficient profiles (average chemical shift divided by the temperature difference; this figure is derived from the 278 and 298 K data, but all other temperature combinations yield the same result) are obtained for wild type TDP-43$^{266–414}$ (black line) and the G298S (red), M337V (blue) and Q331K (green) mutants. Around residues 320–340, the temperature coefficient is less than 4.6 ppb/K, which is characteristic of hydrogen-bonded secondary structure. (d) Temperature titration spectra recorded for the wild type (red: 298 K, orange: 293 K; yellow 288 K; green: 283 K; cyan: 278 K). (e,f) Secondary structure propensities (α-helix and random coil) derived from circular dichroism data at different temperatures for the four variants.

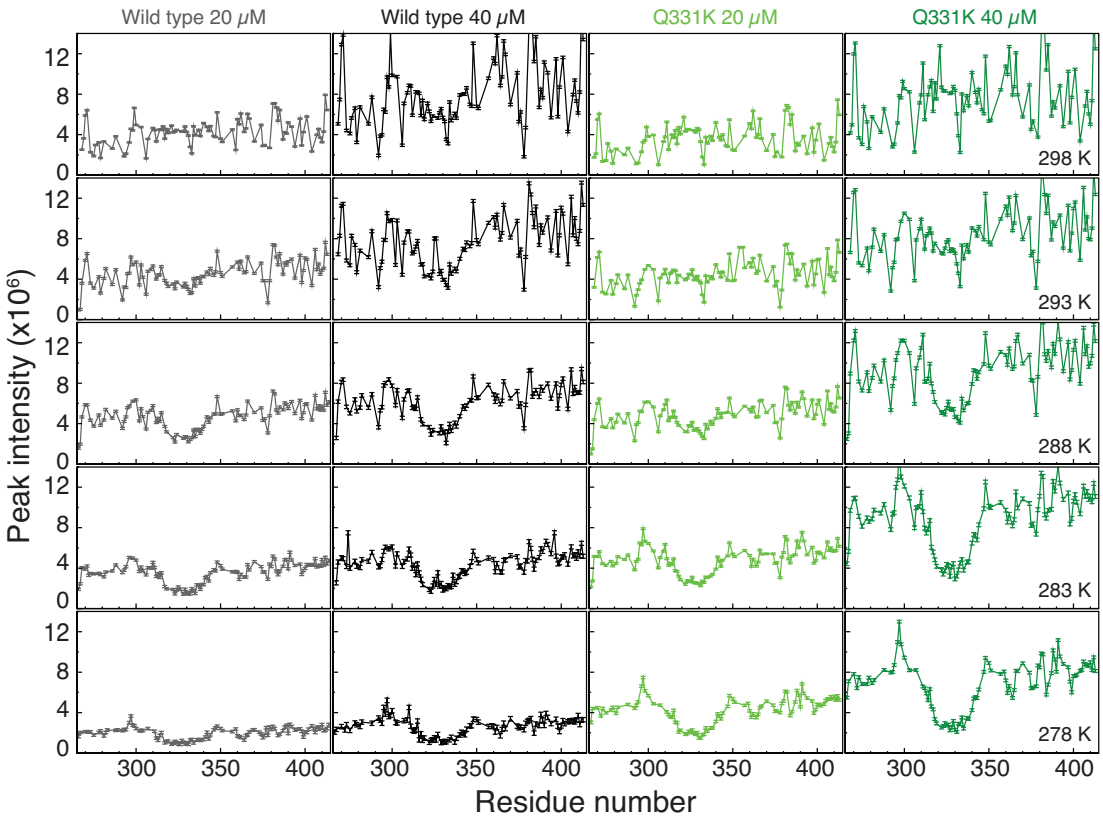

Fig. S10. NMR peak intensity profiles of wild type TDP-43$^{266–414}$ (left two panels, grey) and the Q331K mutant (right two panels, green) at concentrations of 20 (light grey/green) and 40 μM (dark grey/green). These intensities were used to calculate the intensity ratios shown in Fig. 4a and 4d

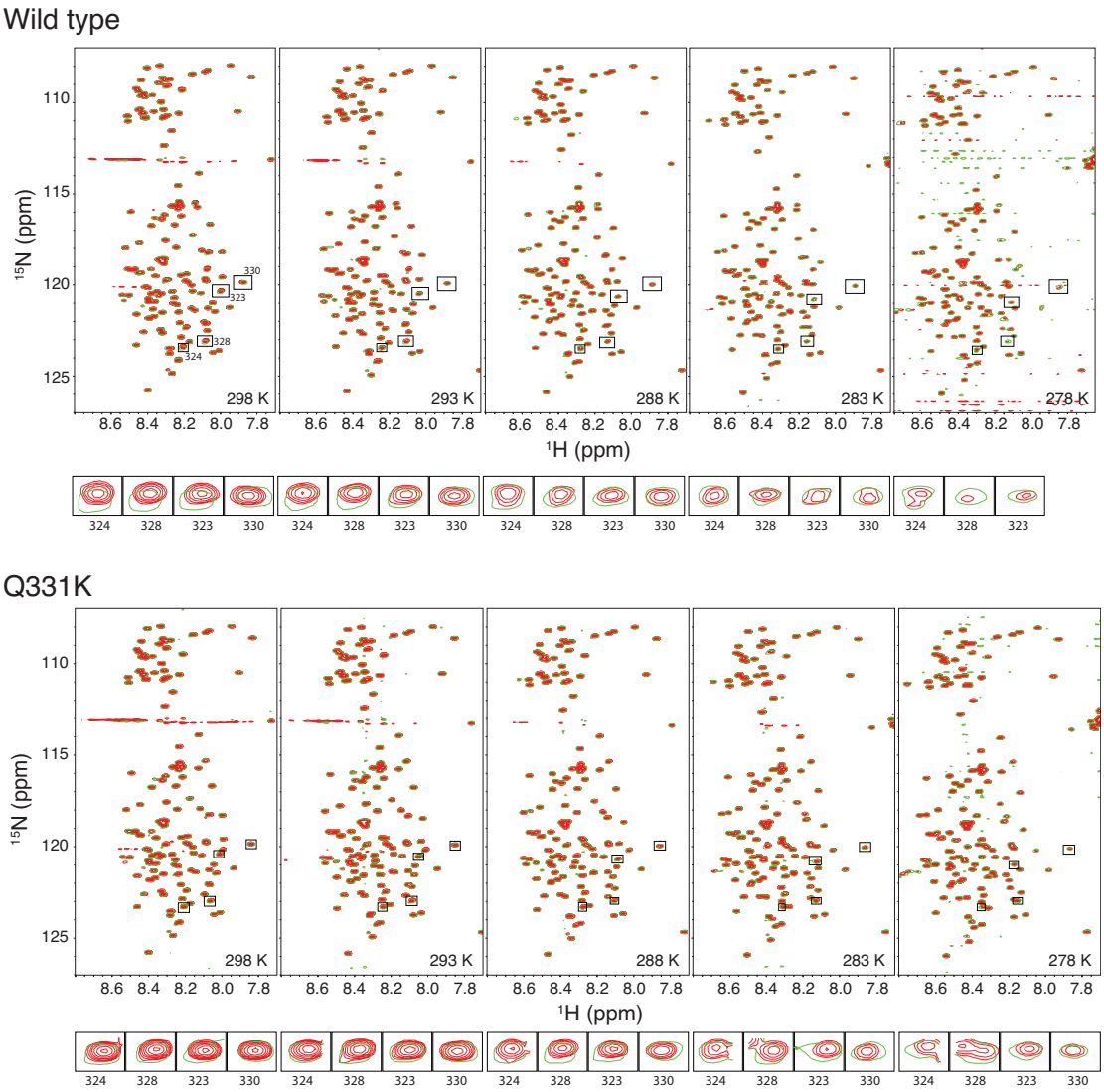

Fig. S11. ^{15}N-HSQC spectra of wild type TDP-43$^{266–414}$ and the Q331K mutant at different temperatures and concentrations. The spectrum recorded for the 40 µM sample is overlaid in red on those from 20 µM samples (green) at the five temperatures indicated. The peaks (A324, A328, M323 and L330) for which the changes in chemical shifts at 298 K are the most noticeable (shown in Fig. 4b and 44) are highlighted below each spectrum combination.

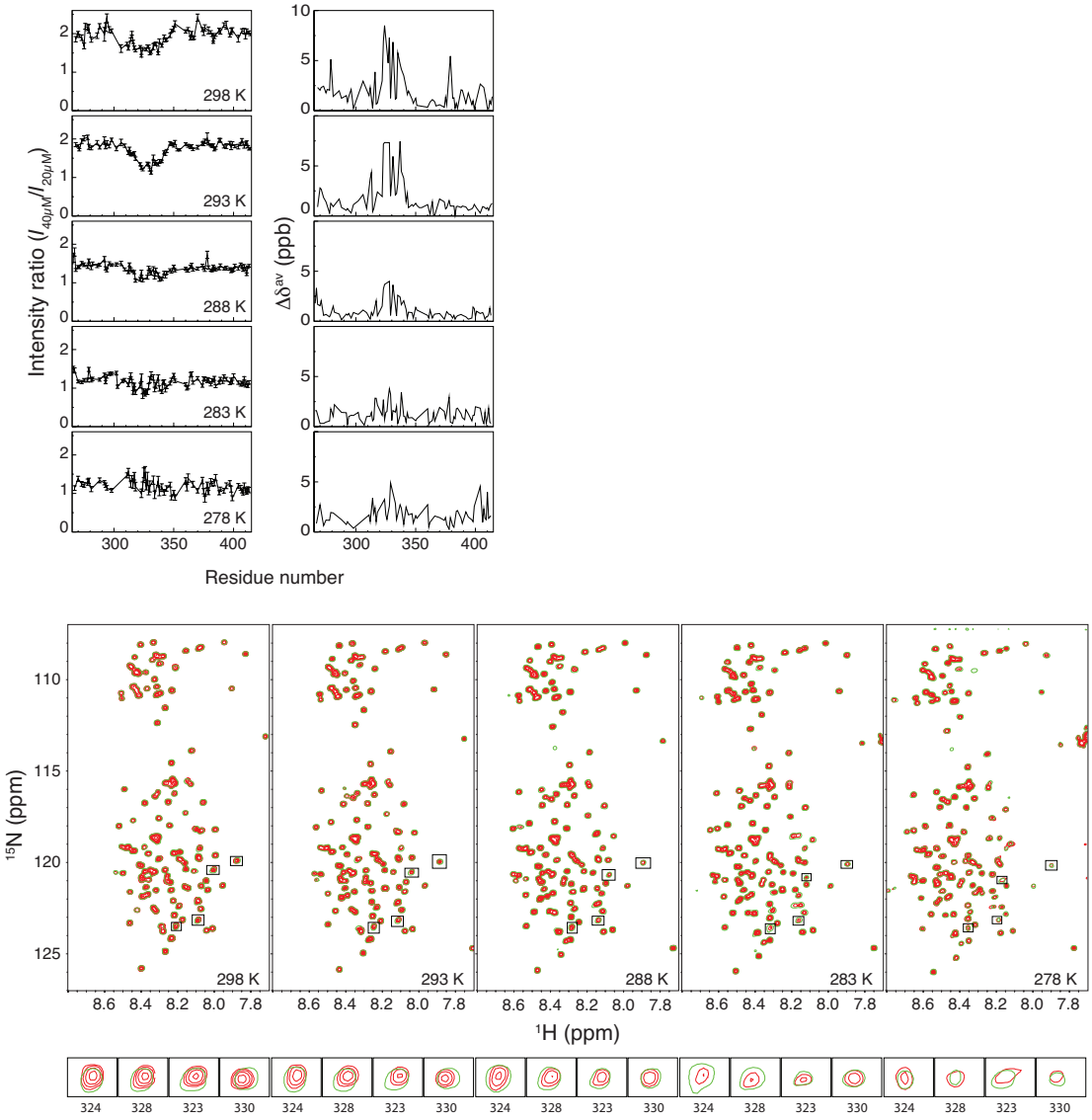

Fig. S12. Intensity ratios and ^{15}N-HSQC spectra of wild type TDP-43$^{266–414}$ at 20 and 40 μM in the presence of 80 mM urea at 278–298 K. The same analysis procedure was used as described for Figs. 4 and S11. The intensity ratios and chemical shift perturbations show that the sample is self-associated at high temperature (293 and 298 K). At low temperature, the 40 μM sample undergoes liquid-liquid phase separation, whereas the 20 μM sample does not, and thus the intensity ratio is close to one with little chemical shift perturbation.

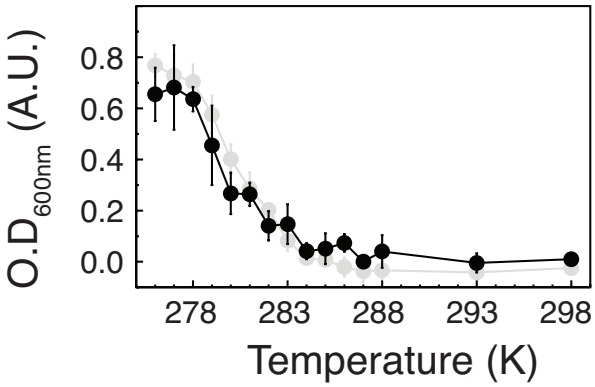

Fig. S13. Solution turbidity assays for TDP-43$^{266–414}$ with (grey) or without (black) a hexahistidine tag.

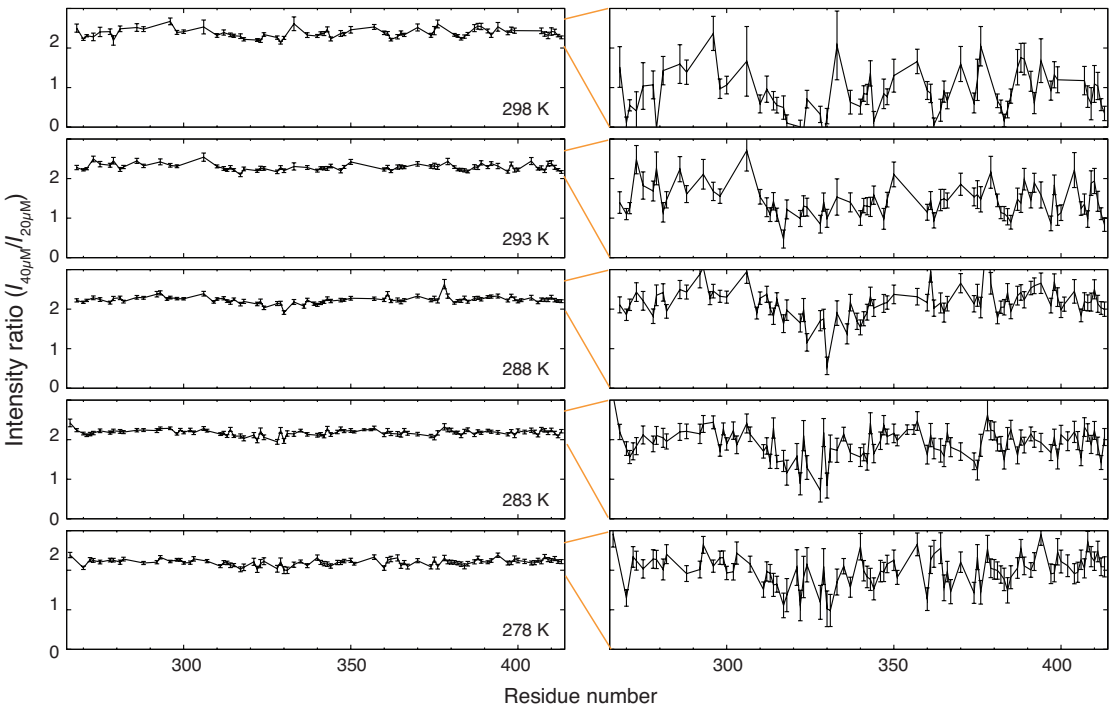

Fig. S14. Intensity ratios of the 20 and 40 μM W334G mutants from 278 to 298 K. The same analysis as for Fig. 4 was applied. The intensity ratios show that the self-association is significantly reduced. However, a very weak level of reduced intensity ratio in the α-helical region is still observed (right Panel; the zoom-in of the left Panel).

Reference:

1 Chen, T. C., Hsiao, C. L., Huang, S. J. & Huang, J. R. The Nearest-Neighbor Effect on Random-Coil NMR Chemical Shifts Demonstrated Using a Low-Complexity Amino-Acid Sequence. *Protein Pept Lett* **23**, 967-975 (2016).

2 Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* **52**, 315-327, doi:10.1007/s10858-012-9611-z (2012).

3 Hyberts, S. G., Frueh, D. P., Arthanari, H. & Wagner, G. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *J Biomol NMR* **45**, 283-294, doi:10.1007/s10858-009-9368-1 (2009).

4 Kim, J. R., Muresan, A., Lee, K. Y. & Murphy, R. M. Urea modulation of beta-amyloid fibril growth: experimental studies and kinetic models. *Protein Sci* **13**, 2888-2898, doi:10.1110/ps.04847404 (2004).

5 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619, doi:10.1016/S0006-3495(00)76713-0 (2000).

6 Radivojac, P., Obradovic, Z., Brown, C. J. & Dunker, A. K. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, 216-227 (2003).

7 Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. & Dunker, A. K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **61 Suppl 7**, 176-182, doi:10.1002/prot.20735 (2005).

8 Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434, doi:10.1093/bioinformatics/bti541 (2005).

9 Wootton, J. C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* **18**, 269-285

(1994).