

The N-terminal domain of ALS-linked TDP-43 assembles without misfolding

Phoebe S. Tsoi¹, Kyoungjae J. Choi¹, Paul G. Leonard², Antons Sizovs¹, Mahdi Muhammad Moosa¹, Kevin R. MacKenzie^{1,3}, Josephine C. Ferreon^{1*}, Allan Chris M. Ferreon^{1*}

*Corresponding authors

¹Department of Pharmacology and Chemical Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, USA

²Department of Genomic Medicine and Core for Biomolecular Structure and Function, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

³Department of Pathology and Immunology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, USA

*E-mail: allan.ferreon@bcm.edu; josephine.ferreon@bcm.edu

Contributions:

J.C.F. and A.C.F. designed the experiments; P.S.T. and A.C.F. performed and analyzed the ensemble CD and fluorescence spectroscopy, and single-molecule FRET experiments. P.S.T. and J.C.F. performed and analyzed the NMR experiments. K.J.C. prepared the TDP-43 DNA constructs and performed the imaging experiments. K.J.C., J.C.F. and A.C.F. purified and prepared the TDP-43 protein samples. P.G.L. carried out the analytical SEC and sedimentation velocity experiments. A.S. and J.C.F. performed the SEC-static light scattering experiments. K.R.M., M.M.M., J.C.F. and A.C.F. wrote the manuscript, with inputs from all other authors.

Abstract

TDP-43 forms inclusions in several neurodegenerative diseases, and both its N- and C-terminal domains are implicated in this process. We show that the folded TDP-43 N-terminal domain oligomerizes under physiological conditions and propose that, in full-length TDP-43, association between folded N-terminal domains enhances the propensity of the intrinsically unfolded C-terminal domains to drive pathological aggregation.

Amyotrophic lateral sclerosis is a disorder of motor neurons that often results in a progressive fatal muscle atrophy¹, and cytoplasmic TAR DNA-binding protein-43 (TDP-43) inclusion bodies are the salient features of both familial and sporadic disease pathologies². Studies on the mechanism of TDP-43 action have revealed a complex interplay between the *loss of cellular function* and *gain of toxic dysfunction* effects of the *TARDBP* gene product^{1, 3, 4}. TDP-43 comprises an N-terminal domain (TDP-43^{NTD}), two RNA recognition motif single-stranded nucleic acid binding domains, and a disordered C-terminal domain (TDP-43^{CTD})⁵. The glycine rich C-terminal domain, which harbors the vast majority of disease-linked mutations⁴, self-interacts and undergoes both liquid-liquid phase separation and aggregation⁶⁻⁸. While none of the hitherto identified pathogenic mutations reside within the N-terminal domain, recent *in vivo* studies clearly indicate a role for the N-terminal domain in pathogenic aggregation^{3, 9, 10}.

The TDP-43 N-terminal domain is highly conserved across the metazoan lineage¹¹; although its physiological role remains elusive, it has been reported to self-associate *in vitro*¹² and in cells⁹. Conflicting observations have been made about TDP-43^{NTD} physical properties^{5, 13, 14}. Qin *et al*¹³ report that a tagged construct of TDP-43 residues 1-102 adopts an unstable ubiquitin-like fold (~30% unfolded at 25°C, pH 4.0 in unbuffered water) and hypothesize that this unfolded state plays a role in initiating full-length TDP-43 aggregation. Mompeán *et al*⁵ find that a tagged construct of residues 1-77 is completely folded and monomeric under similar solution conditions (25°C, pH 3.8, 3 mM sodium acetate). The contradictory findings may arise from sample conditions, construct lengths, expression tags¹⁴, or expression/purification methods. Here, we integrate observations from ensemble biophysical techniques and single-molecule methods to assess TDP-43^{NTD} stability and self-association, focusing on near-physiological conditions but also examining a wide pH range to attempt to reconcile divergent literature reports.

Despite reports of poor TDP-43^{NTD} *in vitro* solution properties¹³, we find refolded TDP-43^{NTD} (residues 1-80, no tags) to be readily soluble and suitable for ensemble experiments at near-physiological conditions (pH 7.5, 200 mM NaCl). Thermal denaturation of TDP-43^{NTD} ($\alpha\beta\gamma$ buffer, pH 7.5 \pm 0.05) show clear pre-transition baselines (**Fig. 1a**), indicating that the native folded state is the major ensemble species. To probe for minor species that might escape detection in ensemble-averaged experiments, and to eliminate any self-association effects, we used single-molecule Förster resonance energy transfer (smFRET) experiments that can directly resolve minor species in a mixture¹⁵. We labeled TDP-43^{NTD} with donor and acceptor dyes (Alexa Fluor 488 & Alexa Fluor 594; see **Supplementary Methods**) at residues 2 and 81, which are close in space in the folded domain (PDB 2n4p⁵). As a single doubly-labeled protein in the dilute sample (100 pM) traverses through the femtoliter-observation volume, we record fluorescence signals from the donor and acceptor dyes. Signal intensities from individual channels are analyzed to yield FRET efficiency (E_{FRET}) histograms. Folded, doubly labeled TDP-43^{NTD} gives high E_{FRET} values (**Fig. 1c, top panel**); decreased E_{FRET} implies increased distance between the dyes, from which we infer TDP-43^{NTD} unfolding and expansion. Near-zero E_{FRET} values (gray shaded region at left in **Fig. 1c**) arising from TDP-43^{NTD} molecules with absent, photobleached, or otherwise nonfluorescent acceptor dyes are excluded from our analysis.

Binned smFRET data in the absence of denaturant are consistent with a single-distribution E_{FRET} value of ~0.85, and we infer the existence of a single, folded population at pH 7.5 (**Fig. 1c**,

top). Chemical denaturation with guanidine hydrochloride (GdnHCl) lowers E_{FRET} (with ~ 1 M GdnHCl, $E_{\text{FRET}} \approx 0.65$). At intermediate denaturant concentrations, histograms reveal coexisting E_{FRET} distributions for both folded and unfolded protein states (**Fig. 1c, panels 2-4 from top**), indicating slow interconversion between the states on the experimental time scale (~ 500 μs). The native state mean E_{FRET} remains relatively unchanged from 0 to 2 M denaturant; we infer that the native state does not expand significantly with denaturant but is depopulated in favor of the unfolded state. In contrast, the denatured ensemble mean E_{FRET} decreases at elevated GdnHCl (**Fig. 1c** and **Supplementary Fig.13**), similar to denaturant-induced expansion of unfolded ensembles of globular proteins^{16, 17}.

The TDP-43^{NTD} thermal transition midpoint for unfolding (T_m) is higher than 50°C from pH 6 to 8 but drops dramatically with extremes of pH (**Fig. 1b** and **Supplementary Fig. 7**). The steep dependence of T_m on pH under acidic conditions may explain how experiments using different constructs and buffer conditions access the unfolded state to different extents. At near-physiological salt and pH conditions, however, our ensemble and single-molecule data establish that TDP-43^{NTD} is folded from low pM to μM concentrations. Successive rounds of thermal unfolding and refolding without detectable precipitation and near identical T_m 's show that TDP-43^{NTD} folding is robustly reversible near physiological pH and salt.

Because our refolded TDP-43^{NTD} is stable and soluble over a broad pH range, we used NMR spectroscopy to probe its structure. ^1H - ^{15}N TROSY-HSQC spectra of deuterated TDP-43^{NTD} samples (**Supplementary Figs. 10-12**) at pH 4.0 and zero salt show well-dispersed peaks with shifts similar to those reported by Mompeán et al⁵, indicating that our TDP-43^{NTD} adopts a fold similar to their construct despite lacking a His tag and having four additional C-terminal residues. However, our TDP-43^{NTD} spectra at pH 7.0 and 50 mM NaCl differ strongly from those of Mompeán et al⁵: although the amides are well-dispersed, only half of the expected resonances are visible at 25°C (**Fig. 2a** and **Supplementary Fig. 10a and 11**). Raising the temperature to 40°C and lowering the protein concentration increases the number of peaks but also their heterogeneity (**Supplementary Figs. 11 and 12**). However, we were able to assign the backbone chemical shifts of the domain as a H₆GB1-TDP-43^{NTD} fusion (~ 20 kDa) at pH 6.8 (**Supplementary Table 1**). The backbone shifts are similar⁵ to those reported for the isolated domain at pH 4 (**Supplementary Fig. 10b**), indicating a similar fold. We propose that TDP-43^{NTD} self-association at neutral pH broadens resonances for amides positioned near the protein-protein interface, and that both low pH and N-terminal fusions decrease self-association. Resonances that are assigned in the fusion protein but absent in our spectra of the isolated domain at pH 7.0 map to a contiguous surface in TDP-43^{NTD} corresponding to the last 3 β -strands (**Fig. 2b**), supporting our proposal.

To test directly for self-association, we performed sedimentation velocity analytical ultracentrifugation (AUC) and size-exclusion chromatography (SEC) analyses of TDP-43^{NTD} dilution series. Both AUC and SEC show that TDP-43^{NTD} exists as monomer, dimer, and tetramer species, but not as large aggregates (**Fig. 2c-e** and **Supplementary Figs. 8 and 9**); the smallest SEC species is determined to be a monomer by static light scattering (**Supplementary Fig. 8b**). The higher order species readily dissociate to monomer and dimer, consistent with

equilibrium self-association and in contrast to amyloidogenic proteins that aggregate irreversibly¹⁸.

Ensemble thermal unfolding fluorescence measurements show that the TDP-43^{NTD} thermal unfolding mid-point temperature (T_m) increases with protein concentration (**Supplementary Figs. 2-4**). Because this finding contradicts the concentration-dependent unfolding described by Qin *et al.*¹³, we performed smFRET on doubly-labeled 100 pM TDP-43^{NTD} titrated with up to 120 μ M unlabeled TDP-43^{NTD} (**Supplementary Fig. 14**). The folded state E_{FRET} peak is not significantly perturbed and no minor species are detected across this 10⁶-fold change in protein concentration, so we conclude that self-associating TDP-43^{NTD} maintains its global fold. Together, these data establish that the TDP-43 N-terminal domain self-associates under physiological buffer conditions through a folded, native state. The conservation of the TDP-43^{NTD} sequence¹¹ suggests that oligomerization is important to TDP-43 cellular function, perhaps by modulating the affinity of the TDP-43 RNA recognition motifs for nucleic acids¹².

Although the cellular function of TDP-43^{NTD}-mediated assembly may not be fully elucidated at present, some key implications for protein dysfunction and misfolding are clear. The intrinsically disordered TDP-43 C-terminal domain (TDP-43^{CTD}) is proposed to be the ‘amyloidogenic core’ that drives TDP-43 aggregation^{7, 19}; at 45 μ M, ~8% of TDP-43^{CTD} can initiate assembly, resulting in phase separation⁷. Here, we have shown that TDP-43^{NTD} forms reversible oligomers to a greater extent at lower concentrations (~30% dimer at 2-5 μ M; ~10% tetramer at 20 μ M; see **Supplementary Fig. 9**). We propose that, under physiological conditions, full-length TDP-43 can assemble reversibly to dimers and tetramers via the folded TDP-43^{NTD}, and that these oligomers enhance the rate of TDP-43^{CTD}-mediated liquid-liquid phase separation, irreversible aggregation and/or fibril formation by raising the effective local concentration of the C-terminal domains relative to one another (**Fig. 3**). Our model explains the importance of TDP-43^{NTD} in driving a fusion protein comprising N-terminal and C-terminal TDP-43 domains flanking GFP to form giant liquid droplets with a critical concentration of phase separation of ~5 μ M²⁰. The possibility that TDP-43^{NTD} itself might directly favor phase separation at physiological pH is at odds with the data presented here and is directly rebutted by its failure to do so under conditions where TDP-43^{CTD} readily phase separates and forms fibrils (**Supplementary Fig. 15**). Our model also explains why a minimal TDP-43 construct lacking RRM domains requires the TDP-43 N-terminal domain to induce aggregation of endogenous TDP-43²¹: the N-terminal domain within the artificial construct helps recruit the endogenous protein through its N-terminal domain. TDP-43 may represent a natural example of the concept demonstrated in an artificial construct by Shin *et al.*²², where the light-inducible oligomerization domain of Cry2 dramatically accelerates liquid-liquid phase separation of a fusion protein containing a disordered domain in a light-dependent manner. Furthermore, N-terminal self-assembly may also explain why TDP-43 is found in pathological inclusions with other proteins that contain intrinsically disordered domains²³: if a single TDP-43 were incorporated into an aggregate via its C-terminal domain, it could recruit more TDP-43 through its N-terminal domain.

Other biophysical work has suggested that the unfolded state of TDP-43^{NTD} contributes to aggregation during cellular dysfunction^{13, 14}. Although TDP-43^{NTD} can access the unfolded state readily at extremes of pH, we have shown that under near-physiological conditions there are no

detectable misfolding intermediates by which TDP-43^{NTD} itself might aggregate. Instead, the stable, folded N-terminal domain promotes TDP-43^{CTD}-driven TDP-43 aggregation without itself misfolding, by mediating TDP-43 oligomerization.

Acknowledgements

We thank Jin Wang's laboratory for the use of the SEC-SLS equipment. We thank Debra Townley and Baylor College of Medicine Microscopy Core for the TEM data collection. This work was supported by laboratory startup funds from Baylor College of Medicine (A.C.F. and J.C.F.). The authors declare no competing financial interests.

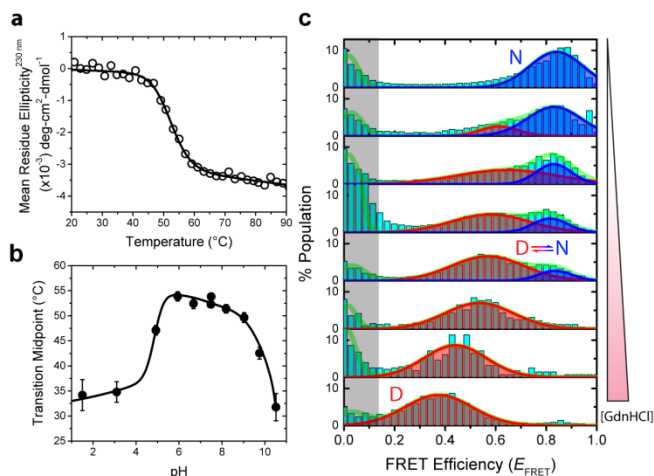


Figure 1. TDP-43^{NTD} is thermodynamically stable at physiological conditions. (a) Thermal unfolding of TDP-43^{NTD} was monitored via far-UV CD spectroscopy. Data is shown as open circles; non-linear least squares (NLS) fit of the data to a two-state protein-denaturation model is shown as a line. Transition midpoint(s) are obtained from the NLS fit. (b) Changes in transition midpoints as a function of pH showing TDP-43^{NTD} is thermodynamically stable at physiological conditions and access acid- and alkaline-unfolded states at respective pH conditions. T_m values are shown as filled circles; the line represents equi-population boundary of folded and unfolded states as determined through global phase diagram analysis. See also **Supplementary Fig. 7**. (c) Single-molecule FRET histograms of two-state unfolding of TDP-43^{NTD} upon chemical denaturation. The protein shows no detectable misfolding intermediate and at intermediate denaturant concentration undergoes slow $N \rightleftharpoons D$ interconversion within the temporal resolution of our experiments (500 μ s). GdmHCl concentrations for individual panels are: 0, 1, 1.25, 1.5, 1.75, 2, 4, and 6 M (top to bottom).

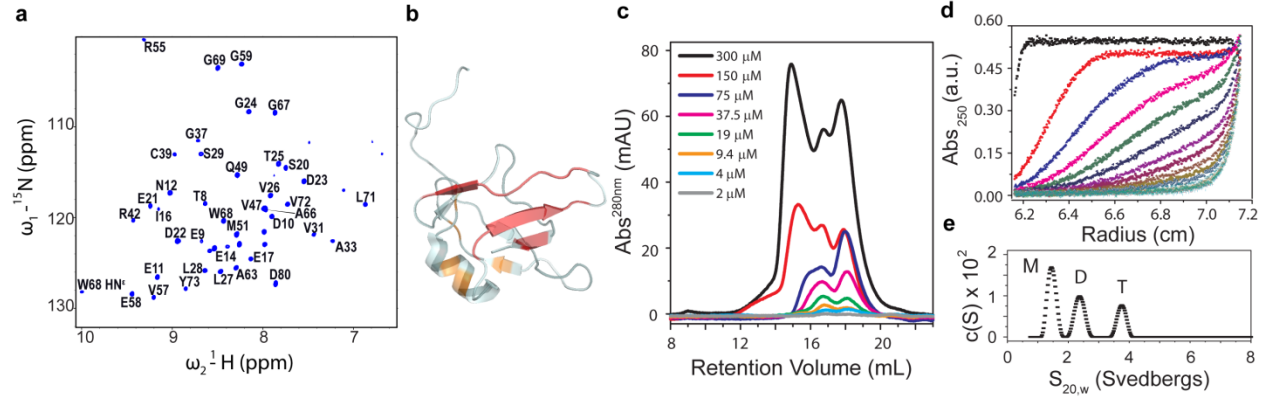


Figure 2. Reversible assembly of TDP-43^{NTD}. (a) ¹H-¹⁵N TROSY-HSQC spectra of TDP-43^{NTD} showing dispersed peaks, albeit with several missing resonances. The missing resonances map to a contiguous surface in TDP-43^{NTD} [(b), red color; PDB: 2n4p⁵]. Residues that sample alternate states are shown in orange (see **Supplementary Figure 11** for details). Size exclusion chromatography (SEC) elution profiles are shown in (c) for samples containing 2-300 μM TDP-43^{NTD} at physiological buffer conditions. Sequential two-fold dilution of 300 μM TDP-43^{NTD} resulted in dissociation of the high MW species (tetramer, T) to lower dimer (D) and monomeric (M) species, indicating a reversibly oligomerizing system that contrasts irreversible fibril formation¹⁸. Sedimentation boundaries (d) and c(S) distributions (e) for 37 μM TDP-43^{NTD} from sedimentation velocity analytical ultracentrifugation experiment shows self-association in a concentration-dependent fashion (see **Supplementary Figure 9** for details).

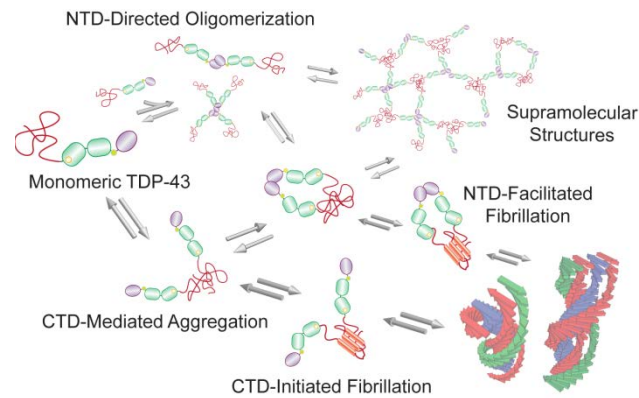


Figure 3. Model for TDP-43 N-terminus- facilitated C-terminal assembly and aggregation.

TDP-43 comprises a fully-folded TDP-43^{NTD} (dark blue) that oligomerizes, two RNA binding domains (light green), and a C-terminal domain (TDP-43^{CTD}, red) that can convert from unstructured to folded intermediate and amyloid fibril states. The N-terminus can facilitate TDP-43^{CTD} phase transitions as well as C-terminus interactions that can lead to TDP-43 fibril formation or pathological aggregation.

References

1. Robberecht, W. & Philips, T. The changing scene of amyotrophic lateral sclerosis. *Nat Rev Neurosci* **14**, 248-264 (2013).
2. Andersen, P.M. & Al-Chalabi, A. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol* **7**, 603-615 (2011).
3. Zhang, Y.-J. et al. Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proc Natl Acad Sci U S A* **106**, 7607-7612 (2009).
4. Lee, E.B., Lee, V.M.Y. & Trojanowski, J.Q. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat Rev Neurosci* **13**, 38-50 (2012).
5. Mompeán, M. et al. The TDP-43 N-terminal domain structure at high resolution. *FEBS J* **283**, 1242-1260 (2016).
6. Wang, I.-F. et al. The self-interaction of native TDP-43 C terminus inhibits its degradation and contributes to early proteinopathies. *Nat Commun* **3**, 766 (2012).
7. Conicella, A.E., Zerbe, G.H., Mittal, J. & Fawzi, N.L. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24**, 1537-1549 (2016).
8. Kitamura, A. et al. Interaction of RNA with a C-terminal fragment of the amyotrophic lateral sclerosis-associated TDP43 reduces cytotoxicity. *Sci Rep* **6**, 19230 (2016).
9. Zhang, Y.-J. et al. The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum Mol Genet* **22**, 3112-3122 (2013).
10. Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F.E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum Mol Genet* **24**, 9-20 (2015).
11. Wang, H.-Y., Wang, I.F., Bose, J. & Shen, C.K.J. Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics* **83**, 130-139 (2004).
12. Chang, C.K. et al. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem Biophys Res Commun* **425**, 219-224 (2012).
13. Qin, H., Lim, L.Z., Wei, Y. & Song, J. 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**, 18619-18624 (2014).
14. Chang, C.K. & Huang, T.H. Untangling the structure of the TDP-43 N-terminal domain. *FEBS J* **283**, 1239-1241 (2016).
15. Schuler, B. & Eaton, W.A. Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* **18**, 16-26 (2008).
16. Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I.V. & Eaton, W.A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci U S A* **104**, 1528-1533 (2007).
17. Ziv, G., Thirumalai, D. & Haran, G. Collapse transition in proteins. *Phys Chem Chem Phys* **11**, 83-93 (2009).
18. Fink, A.L. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**, R9-23 (1998).
19. Jiang, L.L. et al. Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Sci Rep* **6**, 23928 (2016).
20. Schmidt, H.B. & Rohatgi, R. In Vivo Formation of Vacuolated Multi-phase Compartments Lacking Membranes. *Cell Rep* **16**, 1228-1236 (2016).
21. Romano, V., Quadri, Z., Baralle, F.E. & Buratti, E. The structural integrity of TDP-43 N-terminus is required for efficient aggregate entrapment and consequent loss of protein function. *Prion* **9**, 1-9 (2015).
22. Shin, Y. et al. Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171 e114 (2017).

23. Villemagne, V.L., Fodero-Tavoletti, M.T., Masters, C.L. & Rowe, C.C. Tau imaging: early progress and future directions. *Lancet Neurol* **14**, 114-124 (2015).