HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2018 November 21.

Published in final edited form as:

Nat Struct Mol Biol. 2018 June ; 25(6): 463–471. doi:10.1038/s41594-018-0064-2.

Atomic Structures of Segments from TDP-43 LCD and insight into Reversible and Pathogenic Aggregation

Elizabeth L. Guenther[†], **Qin Cao**[†], **Hamilton Trinh**[#], **Jiahui Lu**, **Michael R. Sawaya**, **Duilio Cascio**, **David R. Boyer**, **Jose A. Rodriguez**, **Michael P. Hughes**, and **David S. Eisenberg**[*]

Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, USA

UCLA-DOE Institute, University of California, Los Angeles, Los Angeles, CA, USA

Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA

Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA

Chemistry and Biochemistry Department, University of California, Los Angeles, Los Angeles, CA, USA

Abstract

The normally soluble TAR DNA Binding Protein 43 (TDP-43) is found aggregated both in reversible stress granules and irreversible pathogenic amyloid. In TDP-43, the low complexity domain (LCD) is believed to be involved in both types of aggregation. To discover the structural origins of these two modes of β-sheet rich aggregation, we have determined ten structures of segments of the LCD of human TDP-43. Six of these segments form steric zippers characteristic of the spines of pathogenic amyloid fibrils; four others form LARKS, the labile amyloid-like interactions characteristic of protein hydrogels and proteins found in membrane-less organelles, including stress granules. Supporting a hypothetical pathway from reversible to irreversible amyloid aggregation, we found that familial ALS variants of TDP-43 convert LARKS to irreversible aggregates. Our structures suggest how TDP-43 adopts both reversible and irreversible β-sheet aggregates, and the role of mutation in the possible transition of reversible to irreversible pathogenic aggregation.

[*]Corresponding Author: David S. Eisenberg, david@mbi.ucla.edu.
[#]Present address: Wayne State University School of Medicine, Detroit, MI, USA.
[†]These authors contributed equally to this work.

Author Contributions

E.L.G, Q.C and D.S.E designed the project and wrote the manuscript with input from all other authors especially M.R.S. E.L.G and H.T conducted fibril growth assays and prepared peptide crystals. E.L.G, Q.C and J.L cloned and purified TDP-43 constructs and performed protein aggregation assay and fibril diffraction. M.P.H predicted putative LARKS. E.L.G, H.T and M.R.S processed and solved ^{300}GNNQGSN^{306}. E.L.G and M.R.S processed and solved ^{312}NFGAFS^{317}, ^{312}NFGTFS^{317}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402}. E.L.G, Q.C, J.L and M.R.S processed data and solved the ^{321}AMMAAA^{326} and ^{328}AALQSS^{333}. J.A.R collected microED data on ^{312}NFGEFS^{317} and ^{333}SWGMMGMLASQ^{343}, M.R.S, D.C and E.L.G processed data and solved the structure of ^{312}NFGEFS^{317} and ^{333}SWGMMGMLASQ^{343}. D.R.B collected microED data for ^{312}NFGpTFS^{317}. Q.C, M.R.S, D.C and D.R.B processed data and solved the structure of ^{312}NFGpTFS^{317}. E.L.G, Q.C, M.R.S, M.P.H and D.S.E analyzed structures and designed model of LCD in SG formation and pathological aggregation.

Competing Financial Interests

DSE is an advisor and equity shareholder in ADDRx, Inc.

Introduction

The role of low complexity domains (LCDs) in proteins linked to neurodegenerative diseases is a topic of intense current investigation[1–3]. LCDs are regions of proteins that exhibit biased compositions of amino acids and frequently no secondary structure[4–6]. In some cases the LCDs are the result of nucleotide expansions that cause duplications of short sequences, as in Huntington's Disease[7]. In other cases the LCDs are intrinsic to the protein and believed to have a functional role as in the RNA binding proteins TAR DNA binding protein 43 (TDP-43), Fused in sarcoma (FUS) and hnRNPA1[8], whose LCDs participate in formation of stress granules[1,9,10].

Stress granules (SGs) are aggregates of RNA and RNA binding proteins that form in the cytoplasm when the cell experiences physical, mechanical or chemical stress[11,12]. This process is often initiated by a halt in translation and puts the cell in a hibernating state until the stress subsides[13]. SGs are reversible, meaning that SG-proteins disaggregate when the stress is relieved, the disaggregated proteins and RNA are then free to return to their normal functions in the cell[1,2,14,15]. Phosphorylation has been shown to play a role in these SG-proteins dynamics[16,17]. Over the last decade, researchers have studied how this process occurs and what biophysical properties these aggregates exhibit. Recent studies on G3BP-1, a SG protein, have demonstrated that the SGs exist in a biphasic state with the mRNA packed at the center and the RNA-binding proteins on the periphery[18]. Also, foci composed entirely of RNA have been observed as a product of phase transition in human cells[19]. The mechanism of SGs assembly and the role of RNA are still being investigated[1,3,15,20]. Studies on FUS and hnRNPA1 have demonstrated that the LCDs are sufficient for liquid-liquid phase separation (LLPS)[1,15]. Despite progress, questions remain about the structures of the aggregates and how these may differ between SG assemblies and pathogenic states[4].

Our recent study determined atomic structures of five segments of low-complexity domains of FUS, hnRNPA1, and nup98[21]. All of these displayed kinked β-sheets with weak interactions between sheets. We gave these structures the acronym LARKS (Low-complexity Aromatic-Rich Kinked Segments), and proposed that these LARKS act as Velcro to provide adhesion between LCDs to form SGs or other membrane-less organelle assemblies. These structures contrast with steric-zippers which act as molecular glue to fasten together amyloidogenic segments in amyloid fibrils(reviewed in[22]). Our study of LARKS offered an atomic-resolution hypothesis about the molecular mechanism of SG formation.

In addition to studies aimed at determining the structure of SG assemblies, researchers have focused on the role of SG-associated proteins in neurodegenerative diseases. One of the major hallmarks of neurodegenerative diseases is the deposition of proteinaceous amyloid fibrils in cells[23]. Different from SG assembly, aggregation of proteins into the amyloid state in disease seems a virtually irreversible process[22,24]. Variants found in TDP-43, FUS and hnRNPA-1 have all been linked to Amyotrophic Lateral Sclerosis (ALS) among other diseases[25].

TAR DNA Binding Protein (TDP-43) is a 43kD protein composed of four domains: an N-terminal domain, two RNA recognition domains and a C-terminal low complexity domain (LCD)[26–28]. This protein is involved in a number of cellular functions including mRNA shuttling, transcription, miRNA processing and alternative splicing among others[29,30]. Over the past decade, research into TDP-43 has found aggregates of the protein in patients with ALS, Alzheimer's Disease, Parkinson's Disease, Frontotemporal Lobar Degeneration and Huntington's Disease[31–34]. These aggregates are irreversible, ubiquitinated, hyper-phosphorylated, C-terminal cleavage products ranging in size from 25–35kD[35]. Thus it is hypothesized that the LCD of TDP-43 plays a role in both this irreversible disease aggregation and reversible SG aggregation.

Of proteins that participate in SGs and also form pathogenic amyloid fibrils, TDP-43 stands out for four reasons. First, TDP-43 has over 45 variants that are linked to ALS and other neurodegenerative diseases[25,36,37]. Second, the majority of the variants that cause disease are localized to the LCD[38,39]. Third, the majority of the aggregates found in the cytoplasm of cells of patients are cleavage products inclusive of the LCD[40,41]. Finally, TDP-43 is found to aggregate in 97% of sporadic cases and 45% of familial case of ALS[42,43]. This makes TDP-43 an intriguing candidate for understanding how its LCD participates in both reversible SG assembly and irreversible, pathogenic assembly.

Through structure determination, we can learn which segments of TDP-43 may drive aggregation and how familial variants affect this process. Here we use x-ray diffraction and micro-electron diffraction (MicroED) to determine the atomic resolution structures of ten peptide segments, seven in the reference sequence and three variants, from the LCD of TDP-43. We show that the LCD is capable of forming typical steric-zipper β-sheet structures characteristic of pathogenic aggregation. Also we show that segment 312–317 forms a kinked, labile structure—a LARKS, predicted as a protein interaction motif consistent with properties of SG assemblies[21]. We also show how this structure is perturbed by the familial mutants A315T and A315E, which provides an opportunity for understanding the interplay of formation of membrane-less organelles with pathogenic amyloid.

Results

Six segments from the LCD of TDP-43 form steric zippers

For structural studies we targeted segments throughout the LCD because there is no consensus of which region is the amyloidogenic core. The LCD of TDP-43 is an unstructured domain composed of approximately 140 amino acid residues. Researchers have demonstrated that this region plays a role in aberrant mislocalization and aggregation of TDP-43 to the cytoplasm in mice and neuroblastoma cell lines[44,45]. It was initially suggested that the entire LCD, residues 277–414, served as a prion like domain and was the minimum fragment required for aggregation[46]. However, additional studies suggested smaller regions within the LCD are sufficient for aggregation and neurodegeneration. In the first study, segment 286–331 was identified and shown to form amyloid fibrils and confer neurotoxicity on primary cortical neurons[47]. A second study identified residues 311–360, illustrating through NMR and CD studies that this segment undergoes a helix to β-sheet transition that initiates aggregation and cytoplasmic inclusion[48,49]. A third study suggested

that segment 342–366 is the minimum toxic core that transitions from a random coil to a β-sheet[50]. In short, these studies illustrated that there are several regions of the LCD that can be responsible for amyloid-like aggregation.

Our previous studies demonstrated that adhesive cores of amyloid fibrils are composed of 6–11-residue segments that stack together as steric-zippers[22,51]. To identify which segments from the LCD of TDP-43 might be responsible for TDP-43 aggregation, we consulted ZipperDB[52], (URL: https://services.mbi.ucla.edu/zipperdb). We selected the top 10 hits scored by their adhesive energies (supplementary table 1). Additionally, we also selected segments rich in glutamine and asparagine because the C-terminus of TDP-43 is a prion-like sequence and these two residues are important for prion aggregation[53,54]. Lastly, we selected two segments based on a double mutation reported to reduce TDP-43 aggregation[48] (see methods and materials section for detailed segment selection), In total, we targeted fifteen segments in the LCD of TDP-43 (supplementary table 1).

Crystal structures revealed that six of the segments formed steric-zippers: ^{300}GNNQGSN^{306}, ^{321}AMMAAA^{326}, ^{328}AALQSS^{333}, ^{333}SWGMMGMLASQ^{343}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402} (Fig. 1A and B, Table 1). These structures demonstrated tight side chain interdigitation, like the previously reported zippers in other amyloidogenic proteins, such as β-amyloid[51] and Tau[55]. All three six structures are composed of in-register sheets[22]. Segments ^{321}AMMAAA^{326} and ^{333}SWGMMGMLASQ^{343} form anti-parallel sheets (Supplementary Fig. 1) while the other four segments form parallel sheets.

Segments ^{300}GNNQGSN^{306}, ^{328}AALQSS^{333} and ^{370}GNNSYS^{375} all form Class 1 steric zippers in which the β-sheets are mated face-to-face with the same edges of the sheet oriented up the fibril (Fig. 1B)[22,51]. These Class 1 structures have two distinct interfaces (labeled as IF1 and IF2 in Fig.1B). Segments ^{321}AMMAAA^{326} and ^{333}SWGMMGMLASQ^{343} are Class 7 steric zippers with face-to-back and up-up orientation. Segment ^{396}GFNGGFG^{402} forms a Class 4 steric zipper with face-to-back β-sheets and up-down orientation (Fig. 1B)[51]. These Class 7 and Class 4 structures have one unique interface. The area buried (Ab) and shape complementary (Sc) of all interfaces are labeled in Fig. 1B[51].

Six zipper segments are important for pathogenic TDP-43 aggregation

To investigate the connection between our steric-zipper structures and the pathogenic aggregates of TDP-43, we designed mutations to disrupt the zipper structures and tested their influence on aggregation of the TDP-43 protein. The segment we chose to represent pathogenic TDP-43 aggregation contains residues D-414 because it is a pathogenic segment of TDP-43 that was identified from FTLD brains[40]. We term this segment TDP-CTF, where CTF stands for C-terminal fragment. We disrupted steric-zipper formation by substituting tryptophan at the central residue of the sheet-sheet interface; its bulk imposes steric hindrance. We designed 8 mutations, Q303W and G304W targeting interface 2 and 1 of ^{300}GNNQGSN^{306}, respectively; A324W targeting the interface of ^{321}AMMAAA^{326}; L330W and Q331W targeting interfaces 1 and 2 of ^{328}AALQSS^{333}, respectively; M337W targeting the interface of ^{333}SWGMMGMLASQ^{343}; S373W targeting interface 2 of ^{370}GNNSYS^{375}; and G399W targeting the interface of ^{396}GFNGGFG^{402}.We did not design

a mutation in interface 1 of ^{370}GNNSYS^{375}, since the interaction is relatively weak as judged by its small area buried.

Our mutagenesis results indicate that all six short segments are important for TDP-43-CTF aggregation. We achieved an aggregate-free starting point by conjugating TDP-CTF wild type and its mutants with a SUMO tag (Fig.2a, top panel). Aggregation was initiated by ULP1 protease removal of the SUMO tag (Supplementary Fig.2). After 30 min of incubation on ice approximately half of the SUMO-free TDP-CTF aggregated, as judged by the pellet fraction after centrifugation (Fig.2a, middle panel). In contrast, most of the SUMO-free mutants remained soluble, except G304W. After overnight incubation, cleavage of the SUMO-tag was nearly complete and all SUMO-free TDP-CTF wild type was in the pellet (Fig.2a, bottom panel). However, aggregation was incomplete for several mutants, displaying soluble TDP-CTF mutant material in the supernatant, including Q330W, A324W, M337W, L330W, Q331W and G399W. These results indicate that interface 2 of ^{300}GNNQGSN^{306}, the interface of ^{321}AMMAAA^{326}, interfaces 1 and 2 of ^{328}AALQSS^{333}, the interface of ^{333}SWGMMGMLASQ^{343}, interface 2 of ^{370}GNNSYS^{375}, and the interface of ^{396}GFNGGFG^{402} are important for pathogenic aggregation of TDP-43. Potentially, these segments could facilitate pathogenic aggregation of wild type TDP-43 by forming the same steric-zipper structures we report here. Three of these zippers are located within the region 318–343, previously identified as the amyloidogenic core essential for TDP-43 aggregation[48]. This correlation is significant given that it covers a relatively small region compared to the whole LCD (26 residues out of 140). Moreover, a previous discovery that the double mutation A324E-M337E reduces TDP-43 aggregation[48] is consistent with our expectation that these substitutions create steric hindrance and electrostatic repulsion, thereby disrupting steric-zipper interfaces of ^{321}AMMAAA^{326} and ^{333}SWGMMGMLASQ^{343}. These results suggest that these segments are important for pathogenic aggregation of TDP-43, and the steric-zipper structures we report here may represent the structure of the aggregation core.

To further validate the role of these segments in TDP-43 aggregation, we compared the fibril diffraction of the low complexity domain of TDP-43 with the simulated diffraction of the zipper structures. We carried out fibril diffraction on TDP-LCD that contains residues 274–414 and was expressed using the same SUMO-tag strategy employed previously (Supplementary Fig.2). TDP-LCD formed long extended amyloid-like fibrils around 15 nm in diameter (Fig.2B), similar to the fibrils formed by TDP-43 in ALS patients[41]. The fibril diffraction pattern shows the rings at 10 Å and 4.8 Å, the signature feature of amyloid fibrils[22]. In comparison, electron micrographs of the six zipper-forming segments reveal microcrystals (Fig. 2B). We have shown these crystals are composed of cross-β assemblies. That is, the architecture revealed by the crystal structures of the segments is consistent with the cross-β signature of the TDP-LCD. Simulated fibril diffraction patterns calculated from their crystal structures are comparable with experimental diffraction of the TDP-LCD fibril (Fig. 2B); the differences correspond to variation in sheet-to-sheet distance among segments.

Segment 312–317 of the LCD forms a kinked structure that is disrupted by the A315T and A315E familial ALS mutants

In addition to pathogenic aggregation, TDP-43 is also involved in functional aggregation, such as SG formation. In SG assembly, the LCDs of SG-proteins are reported to be the key regions responsible for this reversible aggregation[1–3]. To further explore the role of the LCD in the functional aggregation of TDP-43, we focused on segments that have the potential to contribute to SG formation. The McKnight laboratory previously discovered that glycine residues flanked by aromatic side chains are features of hydrogel formation of the LCD of the protein FUS[2]. The hydrogel they studied is a reversible aggregate formed *in vitro* composed of reversible amyloid fibrils that trap a large amount of water, and is potentially relevant to SG formation *in vivo*[56]. They showed that replacement of aromatic residues in the LCD disrupted this hydrogel formation[2], which led to the hypothesis that aromatic residues play a role in SG assembly. For this reason, our recent study focused on the segments of the LCD of SG proteins that contain an aromatic residue and two proximal small or polar residues[21]. Structural studies of these segments in FUS and hnRNPA1 revealed pairs of kinked sheets, defined as LARKS (Low complexity, Aromatic-Rich, Kinked Segments), that display weak van der Waals interactions between mating sheets. Further research suggests the connection of LARKS with protein adhesion of LCDs in reversible aggregation formation[21].

In the present study, we find ^{312}NFGAFS^{317}, in the LCD of TDP-43, has similar amino acid composition and also forms labile aggregates and LARKS (Fig. 3). This suggests that TDP-43 can adopt the same mechanism of SG-formation previously hypothesized for FUS and hnRNPA1. Under EM, ^{312}NFGAFS^{317} forms microcrystals (Fig. 3A). These microcrystals are labile, meaning they fully dissolve when treated with ≥1% SDS and heat to 70°C for 15 minutes. Their lability was observed by diminished turbidity (Fig. 3B). This behavior is similar to hydrogels formed by LCDs of SG-proteins, which re-solubilize by treatment with SDS and heat[2], unlike pathogenic amyloid fibrils such as β-amyloid fibrils (Supplementary Fig. 3A). The structure of ^{312}NFGAFS^{317} reveals a kinked backbone (Fig. 3C), similar to other LARKS[21]. The kink occurs at Phe316 and has a Ramachandran angles of a left-handed α-helix. The kinked backbone makes two phenylalanine residues at positions 2 and 5 face the same side of the sheet, whereas a canonical β-sheet places the odd numbered sidechains on one face and the even numbered sidechains on the opposite face[51]. The subsequent packing of mating strands creates two different interfaces. Here we see that all the aromatic residues pack near to each other creating a hydrophobic, aromatic interface. The other interface is held together by the hydrogen-bonded interactions of the asparagine side chains. Both interfaces are weaker than most zipper structures as indicated by the relative low surface area buried, and consistent with the observation that these assemblies are labile under treatment with SDS and heat.

We observe that ^{312}NFGAFS^{317} forms a LARKS structure and its microcrystalline aggregates are labile, thus providing another example of LARKS with a potential role in SG formation, and an opportunity to investigate the influence of pathogenic variants of the putative key segment of functional aggregation. Within the NFGAFS segment, there are two ALS-linked familial variants: A315T and A315E[47,57]. Both of these variants have been

shown to increase redistribution of TDP-43 to the cytoplasm and increase neuronal death as compared to the reference sequence[58–61]. Additionally, previous studies have shown that disease-causing mutations in SG-proteins are associated with accumulation of persistent SGs[62–64]. However the molecular mechanism of transformation of SGs to irreversible aggregates has remained elusive. Here we studied the structure and biochemical behavior of ^{312}NFGTFS^{31}, ^{312}NFGEFS^{317}, and ^{312}NFGpTFS^{317} where pT stands for phosphorylated threonine. We studied ^{312}NFGpTFS^{317} because TDP-43 is found to be hyperphosphorylated in ALS patients[35], and the A315T mutation may create a new phosphorylation site *in vivo*.

All three of these mutant segments formed microcrystals under EM (Fig. 3A). ^{312}NFGEFS^{317} and ^{312}NFGpTFS^{317} aggregates were irreversible under the same SDS and heat denaturing assay as ^{312}NFGAFS^{317}. ^{312}NFGTFS^{317} aggregates, however, were still labile (Fig. 3B). These results indicated that the stabilities of the mutant segments A315E and A315pT were significantly enhanced as compared to that of ^{312}NFGAFS^{317}.

To better understand the biochemical findings, we determined the structures of all three mutant segments. All three structures are composed of kinked, in-register, parallel sheets; however the kink positions differ. In ^{312}NFGAFS^{317} the kink occurs at Phe316. In the case of ^{312}NFGTFS^{317} and ^{312}NFGEFS^{317}, the kink occurs at Gly314; and the ^{312}NFGpTFS^{317} structure is kinked at two residues, Phe313 and Gly314. Despite the different positions of the kinks, all of the kinks result in the two phenylalanine residues facing the same side of the sheet (Fig. 3C).

Further analysis of the structures illustrates that the packing of sheets in both ^{312}NFGTFS^{317} and ^{312}NFGpTFS^{317} are face-to-face, as in ^{312}NFGAFS^{31}. Both have two distinct interfaces of mating strands, one dry interface formed by hydrophobic residues (mainly phenylalanine residues, labeled as IF1 in Fig. 3C) and one wet interface formed by mostly hydrophilic residues (labeled as IF2 in Fig. 3C). The packing of ^{312}NFGEFS^{317} is face-to-back so its two interfaces are identical, with hydrophobic interactions in the middle and hydrophilic interactions on both sides. For the ^{312}NFGpTFS^{317} structure, the phosphate is clearly observed in the Fo-Fc omit map (Supplementary Fig. 3B). To our knowledge, this is the first time a post-translational modification has been reported for an amyloid mimetic structure. The enhanced stability of the A315E and A315pT of ^{312}NFGAFS^{317} assemblies can be explained by the increased interface strengths, indicated by area buried for the dry interface and hydrogen network of the wet interface. The area buried for the dry interface increased significantly from 63.0 Å^2/chain in ^{312}NFGAFS^{317} to 137.0 Å^2/chain in ^{312}NFGEFS^{317} and 175.6 Å^2/chain in ^{312}NFGpTFS^{317} (Fig. 3C). In the case of ^{312}NFGpTFS^{317}, increased buried area results from the four phenylalanine residues of two strands perfectly interdigitating with each other. Additionally, on the wet interface, the side chains of Glu315 and pThr315 create a hydrogen bonding network, which further strengthens the assembly (Supplementary Fig. 3C). This stacking of negatively charged Glu315 and pThr315 is facilitated by the acidic pH of crystallization conditions, in which the residual TFA from peptide synthesis significantly decreases the pH, and results in partially protonated glutamate and P-threonine side chains (see discussion). For ^{312}NFGTFS^{317}, the area buried for dry interface and hydrogen network for the wet interface is similar that of

^{312}NFGAFS^{317}, which is consistent with the observation that ^{312}NFGTFS^{317} has similar behavior in our denaturing assay as ^{312}NFGAFS^{317}.

In summary, we find that the ^{312}NFGAFS^{317} segment of TDP-43 forms a kinked β-sheet structure—a LARKS—which is labile. We find that two familial variants within this segment, A315T and A315E, together with phosphorylation, strengthen the reference sequence assembly and making it irreversible by creating a more stable structure with stronger interaction between each the sheets. These results further support our hypothesis that LARKS are the potential molecular mechanism of functional, reversible aggregate formation, by showing that LARKS exist in another SG-forming protein, TDP-43. Moreover, these results indicate that the pathogenic mutations and or post-translational modifications can strengthen LARKS and drive SG-proteins towards pathogenic aggregation.

Discussion

Our previous studies suggested connections of cross-β structures to both amyloid aggregation through steric-zippers forming irreversible pathogenic fibrils[22] and LARKS forming reversible hydrogels and possibly membrane-less organelles[21]. In this study, we determined structures of segments from the LCD of TDP-43 and found that that this domain has the capacity to form both steric-zippers and LARKS. This is consistent with the dual nature of the TDP-43 LCD which can drive aggregation of the protein into either functional SG aggregates or into pathogenic amyloid fibrils as in ALS or FTLD. In Figure 4 we present a speculative model that incorporates our findings and explains how SG assembly might lead to pathogenic assembly. In physiological conditions, TDP-43 can execute its normal functions without irreversible aggregation, even though steric-zipper (red bars) and LARKS (kinked blue bars) segments exist in their sequences (Fig. 4, top panel). In pathogenic conditions, the steric-zipper forming segments interact with each other to form irreversible assemblies, driving the protein into pathogenic amyloid fibrils (Fig. 4, left pathway). In stress conditions, the LARKS forming segments interact with each other to form reversible assemblies and drive the protein to functional aggregates such as SGs. Since functional assemblies are reversible[1,2,14], the aggregation of proteins can be reversed when stress is relieved (Fig. 4, right pathway). Numerous factors are believed to influence the fate of the proteins, including intra/inter-cellular signals, chaperones, RNA binding, protein folding, microenvironments such as pH, protein concentration, etc. Here, our study suggests that the mutations and or post-translational modifications can also change the pathway of assembly by directly changing the character of assembly-driving segments (Fig. 4, middle pathway). As exemplified by A315E and A315pT, the mutation and phosphorylation strengthen the ^{312}NFGAFS^{317} assembly, converting a labile LARKS to irreversible aggregates. Our hypothesis is that under conditions that favor SGs, LARKS segments interact with each other, and should one or more of them convert to an irreversible structure, the protein may form a pathological, irreversible assembly.

The mechanism represented in Figure 4 may apply for other regions of TDP-43 and other SG-forming proteins. The ^{312}NFGAFS^{317} segment is not the only region proposed to be important for reversible aggregation of TDP-43; residues 320–340 have been reported to be

involved in granule formation[3,65]. Applying a threading algorithm for predicting LARKS from amino acid sequence, we predicted additional LARKS in the LCD of TDP-43[21]. We compared the LARKS predictions with the familial variants reported on TDP-43 LCD, and found that 28 out of 42 reported familial variants are in the regions of predicted LARKS (supplementary table 2). This observation supports our hypothesis that the mechanism of conversion of A315E and A315pT on ^{312}NFGAFS^{317} may be a general mechanism.

Notice that both ^{312}NFGEFS^{317} and ^{312}NFGpTFS^{317} crystals were grown at acidic pH, which facilitated the stacking of glutamate and phosphorylated threonine residues. A recent study also shows that decrease in pH drives phase separation in yeast and in *in vitro* systems[66]. This observation suggests the importance of pH on SG formation and the role of pH in transition of SG to pathogenic aggregation.

In this study, we also reported six steric-zipper structures from LCD of TDP-43, and showed the connection between these structures with pathogenic aggregates of TDP-43 by mutagenesis experiments. The observation that all six segments are important for TDP-43 aggregation, suggests a segmental polymorphism, meaning the aggregation of TDP-43 can be formed via multiple cores. Although recent structural studies have shown that the amyloid fibril core usually contains tens of residues (e.g. cryo-EM structures of Tau[67]), our study on short peptides facilitates the research of amyloidogenic proteins by identifying the high-resolution structures of core aggregation-prone segments, and more importantly, by providing structural insights for inhibitor designs. By designing peptide inhibitors that bind to the ends of steric-zipper structures and cap their growth, we have successfully developed the inhibitors for aggregation of several amyloidogenic proteins, including p53[68] and Tau[55]. The zipper structures we reported here will provide the opportunity for structure-based design of TDP-43 inhibitors.

In summary, the present study provides atomic-resolution structures of irreversible and reversible aggregates of segments of TDP-43, and the possible roles of familial mutations and or post-translational modifications in the interconversion of these two distinct aggregation pathways.

Online Methods

Selection and Generation of TDP-43 Peptide Targets

We chose our peptides of interest by a combination of computational prediction and literature reports (supplementary table 1). For computational prediction, we selected the segments that contain high rank predictions from ZipperDB (ranked by the predicted Rosetta energy). From literature reports, we have selected the segments containing the features important for LCD aggregation. First, the LCD of TDP-43 is believed to be a prion-like domain, and glutamine and asparagine residues are found to be important for prion aggregation[53,54], so we selected three segments from LCD of TDP-43 that have the highest Gln and Asn density. These three segments are GNNQGSN (4 out of 7), ASQQNQ (4 out of 6) and GNNQNQ (5 out of 6); the latter two were also identified by ZipperDB. Second, aromatic residues are reported to be important in LCD aggregation[2], so we selected two segments from the LCD of TDP-43 that have multiple aromatic residues close to each other;

these are NFGAFS (two residues between two phenylalanine residues) and GFNGGFG (three residues between two phenylalanine residues). Third, the A324E-M337E double mutation is reported to reduce TDP-43 aggregation[48], so AMMAAA and SWGMMGMLASQ were selected to study the molecular mechanism of this double mutation. All peptides were purchased from Genscript at a purity of 95% or higher. Notice that from fifteen peptides we selected, we determined the structures for ten of them. The other segments we selected (supplementary table 1), failed to generate structures because of poor crystal formation or poor quality of crystal diffraction.

Peptide Aggregates Formation

All peptide segments were dissolved in phosphate buffered saline (PBS) at pH 7.5 to a final concentration of 20mM. Samples were then incubated at 37°C in a Torrey Pine Scientific shaker at level 9 for 80 hours. Samples were examined by TEM following fibril formation.

Construct design

TDP-CTF (208–414) was designed to mimic the pathological fragment of TDP-43 found in FTLD brain, and the low complexity domain of TDP-43 (TDP-LCD, 274–414) was also designed. Both TDP-43 fragments was conjugated to the carboxyl-terminus of small ubiquitin-like modifier (SUMO) proteins to increase its solubility and prevent its aggregation during expression and purification. The SUMO protein was also conjugated with $(His)_6$-tag on its amino-terminus for easier purification. The SUMO-TDP constructs were inserted into a pET28a vector with following expressed sequences:

SUMO-TDP-CTF:
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIF
FKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR
EQIGGREFFSQYGDVMDVFIPKPFRAFAFVTFADDQIAQSLCGEDLIIKGISVHISNAEP
KHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAF
SINPAMMAAAQAALQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGS
GNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM

SUMO-TDP-LCD:
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIF
FKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR
EQIGGGRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAM
MAAAQAALQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYS
GSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM

Protein purification and validation

Both SUMO-TDP-CTF and SUMO-TDP-LCD protein was expressed in *Escherichia coli* BL21 (DE3) strain. Bacterial cells were cultured in LB media supplemented with 50 μg/ml kanamycin at 37 °C to an OD600 of 0.6–0.8. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25 °C for 3 hours. Cells were harvested and re-suspended in 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM Imidazole, 10% (v/v) glycerol, supplemented with 1% (v/v) Halt Protease Inhibitor single-use cocktail (Thermo

Scientific). The re-suspension solution was sonicated (3s on/3s off cycle, 10 mins) and the supernatant solution was collected after centrifugation (24,000 × g for 20 mins). The supernatant solution was mixed with homemade NucA nuclease (5000 U per liter of cell culture) and filtered before load to HisTrap HP column (GE health care). The HisTrap column was equalized with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM Imidazole before loading of supernatant solution, and washed with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 200 mM Imidazole after loading. The protein expected was eluted by 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 500 mM Imidazole and concentrated by Amicon Ultra-15 centrifugal filters (Millipore) to at least 2 mg/ml (10 fold higher than the working concentration of aggregation assays). SUMO-TDP-LCD was concentrated to 10 mg/ml (10 fold higher than its fibril formation concentration).

To validate the cleavage of SUMO-tag and existence of TDP fragment, purified SUMO-TDP-CTF and SUMO-TDP-LCD was diluted into 0.2 mg/ml with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μM DTT, mixed with 100:1 (weight basis) homemade ULP1 protease, and incubated on ice. Samples were taken before and 1 hour after adding ULP1 protease, and separated by NuPAGE 4–12% Bir-Tris gel (Invitrogen). The pellet of SUMO-TDP-LCD was also collected and sampled from fibril formation assay and loaded to the same gel. The SDS-PAGE was stained by coomassie blue, or probed with TDP-43 polyclonal antibody (Invitrogen, catalog number PA5-27221, 1:1000 dilution) and Anti-Rabbit IgG (whole molecule)-Peroxidase antibody (Sigma, catalog number A0545, 1:5000 dilution) after transferred to a nitrocellulose membrane (Invitrogen). The membrane was visualized with SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Scientific) and BIOMAX Light Film (Carestream).

Fibril formation and diffraction of TDP-LCD

SUMO-TDP-LCD was diluted into 1 mg/ml with the same dilution buffer used in aggregation assays of TDP-CTF, and mixed with 100:1 (weight basis) homemade ULP1 protease. The fibril was formed by shaking the solution at 37 °C for 3 days and sitting it on bench for another 2 days. The fibril was collected by centrifugation at 9,000 × g for 5 mins and washed by water for 2 times. The fibril was sampled for TEM analysis and SDS-PAGE, and then concentrated to 2% of the original volume by centrifugation and removal of supernatant. The concentrated solution was loaded between two siliconized glass capillaries with their heads facing each other. The solution was then dried out and the fibril remained on one of the capillary heads in solid form. The capillary with fibril was mounted on an in-house X-ray machine and the fibril was shot by X-ray for 3 mins. The diffraction pattern was detected and recorded by a CCD detector.

Fibril diffraction simulation of zipper forming segments

The crystal structures of zipper forming segments were used for generation of fibril diffraction simulation. To simulate the fibril diffraction patterns, the space group of each crystal structure was expanded into P1, and structure factors were calculated using the CCP4 program sfall[69]. The diffraction patterns were cylindrically averaged 180 degrees around the fibril axis using custom written software.

Negative Stain Transmission Electron Microscopy (TEM)

Samples for TEM were prepared by applying 5 μL of sample on hydrophilic 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The fibrils were allowed to adhere for 4 min and washed twice with water. The samples were then stained for 2 min with 2% uranyl acetate, leaving a few nanoliters of stain on the grid. Grids were allowed to dry for 1 minute before storage. Each grid was inspected on a T12 (FEI) microscope.

Aggregation assays of TDP-CTF

SUMO-TDP-CTF wild type and its mutants were diluted into 0.2 mg/ml with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μM DTT and filtered by 0.1 μm Ultrafree-MC-VV centrifugal filters (Millipore). The filtered solution was mixed with 100:1 (weight basis) homemade ULP1 protease and incubated on ice to remove the SUMO-tag from TDP-CTF. The samples were taken before and 30 mins after adding the protease, and the 30-minute-samples were divided into supernatant and pellet fraction by centrifugation at 18,000 × g for 3 mins at 4 °C. The pellet fraction was re-suspended with the same buffer used in dilution and with the same volume of the original samples. The samples were mixed with 3:1 (v/v) of NuPAGE LDS sample buffer (Invitrogen) and heated to 100 °C for 10 mins. After the 30-minute-samples were taken, the cleavage solutions were moved to refrigerator (4 °C) and incubated overnight. The overnight samples were then taken and prepared same as 30-minute-samples. All samples were loaded to and separated by NuPAGE 4–12% Bis-Tris gels (Invitrogen) and stained by coomassie blue. To make better comparisons, the samples of wild type TDP-CTF were loaded to both gels when the samples taken from the same time point were split into two gels because of loading well limitation.

Denaturing assays of ^{312}NFGAFS^{317} and its variants

Twenty-percent (m/v) SDS solution (in water) was diluted to 15%, 10% and 5% (m/v) by water, and 20% (v/v) of SDS-water stocks or water were mixed with 80% (v/v) of PBS to make SDS-PBS stocks with 4%, 3%, 2%, 1% and 0% (m/v) of SDS. SDS-PBS stocks with varies SDS content were mixed with 20 mM peptide-PBS solution in 1:1 volume ratio to get final peptide solutions with 2%, 1.5%, 1%, 0.5% and 0% (m/v) SDS. All solutions were divided into 3 PCR tubes (5 μl per tube), and the solutions contain 2%, 1.5%, 1% and 0.5% (m/v) SDS were heated to 70 °C for 15 mins by a thermocycler. The OD340 of solutions in each tube were measured by P330 NanoPhotometer (Implen), and the readings from same peptide were normalized to the mean values of the solutions contain no SDS and without heating.

Aβ fibril denaturing assay

Aβ purification and preparation please see our previous study[70]. Aβ1–42 was diluted into 50 μM with PBS buffer, and incubated at 37 °C overnight to form Aβ fibrils. The fibril solution was mixed with 10% (v/v) of 20% (m/v) SDS stock to make 2% (m/v) final SDS concentration, and heated to 70 °C for 15 mins. The samples were taken before and after adding SDS and heating, and analyzed by TEM.

Crystallization

All crystals, except ^{312}NFGEFS^{317}, ^{312}NFGpTFS^{317} and ^{333}SWGMMGMLASQ^{343} were grown by hanging drop vapor diffusion. ^{300}GNNQGSN^{306} was dissolved in water to a final concentration of 10 mg/ml. The reservoir solution was optimized around the initial hit of 0.2M sodium sulfate, 0.1M Bis tris propane pH 7.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein: reservoir.

^{321}AMMAAA^{326} was dissolved in water to a final concentration of 10mg/ml, supplemented with 29.4 mM of LiOH. Crystals were grown over a reservoir solution containing 1.3 M Ammonium Phosphate and 0.1 M sodium acetate pH 6.27. Hanging drops contained a 1:2 solution of protein: reservoir.

^{328}AALQSS^{333} was dissolved in water to a final concentration of 10mg/ml. Crystals were grown over a reservoir solution containing 0.01 M nickel chloride, 0.1 M Tris pH 8.5 and 20% (w/v) PDG 2000. Hanging drops contained a 2:1 solution of protein: reservoir.

^{370}GNNSYS^{375} was dissolved in water to a final concentration of 10mg/mL. Crystals were grown over a reservoir solution containing 0.2M sodium nitrate, 0.1M Bis tris propane pH 8.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein: reservoir.

^{396}GFNGGFG^{402} was dissolved in water to a final concentration of 12 mg/mL. Crystals formed with the reservoir solution: 800 mM Sodium phosphate monobasic/ 1200 mM Potassium phosphate dibasic, 100 mM Sodium acetate/ Acetic acid pH 4.5. Hanging drops contained a 2:1 solution of protein: reservoir.

^{312}NFGAFS^{317} was dissolved in water to a final concentration of 11mg/mL. Crystals formed with a reservoir solution of 0.1M phosphate/citrate pH 4.2, 40% w/v ethanol and 5% w/v PEG 1,000. The drops contained a 2:1 ratio of protein solution: reservoir.

^{312}NFGTFS^{317} was dissolved in water to a final concentration of 11mg/mL. Crystals formed with a reservoir solution of 0.2M ammonium acetate, 0.1M sodium acetate pH 4.6, and 30% w/v PEG 4,000. The drops contained a 2:1 ratio of protein solution: reservoir.

^{312}NFGEFS^{317}, ^{312}NFGpTFS^{317} and ^{333}SWGMMGMLASQ^{343} were dissolved in PBS (pH 7.5) to a final concentration of 20mM. The peptide solution was shaken at 37°C in a Torrey Pines shaker at level 9 for 80 hours. To better understand the stacking of glutamate and phosphorylated threonine in ^{312}NFGEFS^{317} and ^{312}NFGpTFS^{317} structures, the pH of both peptide solutions was tested by MColorpHast pH-indicator strips (Millipore) and the pH of both solutions was around 4. For ^{312}NFGpTFS^{317}, initial crystals growing from shaking diffracted poorly, and in order to get better diffracting crystals, an additional round of crystal seeding was performed. Fresh made 20 mM ^{312}NFGpTFS^{317} PBS solution was filtered and mixed with 2% (v/v) of pre-formed ^{312}NFGpTFS^{317} crystals, and incubate at 37°C without shaking for 120 hours. Diffraction data was collected from seeded crystals, and the initial crystals were used in denaturing assay and EM observation showing in Fig.3.

X-ray Diffraction Data Collection and Processing

All x-ray diffraction data were collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA, at a wavelength of 0.971 Å and temperature of 100 K. For NFGAFS, data were collected at 5° oscillations and 130 mm detector distance. Detector ADSC Q315 was used. Indexing and integration were done using XDS in $P2_12_12_1$ and scaled with XSCALE to 1.1 Å resolution[71]. The data were imported using the CCP4 suite[69]. Molecular replacement was done using Phaser[72]. Initial models of different idealized polyalanine β-strands did not yield a structure. The final model used was a kinked strand with the sequence NFGAAS, which was based on the truncated structure of NNFGAIL[73]. The model was manually adjusted over 3 cycles to the correct sequence and rotamer orientation in COOT[74]. The atomic model was refined by Phenix[75] with a final R_{work}/R_{free} of 13.3/16.0% and 100% of Ramachandran angles favored.

For NFGTFS, data were collected at 5° oscillations and 125 mm detector distance. Detector ADSC Q315 was used. Indexing and integration was done using XDS in $P2_1$ and scaled with XSCALE to 1.0Å resolution[71]. The data were imported using the CCP4 suite[69]. The structure was solved by direct methods using the program SHELX[76]. The model was manually adjusted over 3 cycles to the correct sequence and rotamer orientation in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 8.7/9.3% and 100% of Ramachandran angles favored.

For GFNGGFG, data were collected at 5° oscillations and 150 mm detector distance. Detector ADSC Q315 was used. Indexing and integration was done using Denzo in P1 and scaled with SCALEPACK to 1.25Å resolution[78]. The data were imported using the CCP4 suite[69]. Molecular replacement was done using Phaser[72]. Initial models of different idealized polyalanine β-strands did not yield a structure. The final model used was two kinked strands with the sequence GFAGG, which was based on the truncated structure of SYSGYS[21]. The model was manually refined using Phenix[75] over 4 cycles to the correct 7 residue sequence and rotamer orientation in COOT[74]. The atomic model has a final R_{work}/R_{free} of 15.2/16.8% and 100% of Ramachandran angles favored.

For AMMAAA, data were collected at 5° oscillations and 130 mm detector distance. Detector EIGER 16M was used. Indexing and integration was done using XDS in $P2_12_12_1$ and scaled with XSCALE to 1.2Å resolution[71]. The data were imported using the CCP4 suite[69]. Molecular replacement was done using Phaser[72] with an idealized polyalanine β-strand (AAAAAA). The model was manually adjusted over 14 cycles to the correct 6 residue sequence and rotamer orientation in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 19.8/21.8% and 100% of Ramachandran angles favored.

For AALQSS, data were collected at 5° oscillations and 130 mm detector distance. Detector EIGER 16M was used. Indexing and integration was done using XDS in $P2_12_12_1$ and scaled with XSCALE to 1.1 Å resolution[71]. The data were imported using the CCP4 suite[69]. The structure was solved by direct methods using the program SHELX[76]. The model was manually adjusted over 13 cycles to the correct 6 residue sequence and rotamer orientation

in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 19.5/24.3% and 100% of Ramachandran angles favored.

For GNNQGSN, data were collected at 5° oscillations and 125 mm detector distance. Detector ADSC Q315 was used. Indexing and integration was done using XDS[71] in C2 and scaled with SCALEPACK[79] to 1.8Å resolution. The data were imported using the CCP4 suite[69]. Molecular replacement was done using Phaser[72] with an idealized β-strand of the sequence GNNAGA. The model was manually adjusted over 38 cycles to the correct 7 residue sequence and rotamer orientation in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 18.3/19.4% and 80% of Ramachandran angles favored as well as 20% allowed.

For GNNSYS, data were collected at 3° oscillations and 125mm detector distance. Detector ADSC Q315 was used. Indexing and integration was done using Denzo in $P2_12_12_1$ and scaled with SCALEPACK to 1.00Å resolution[78]. The data were imported using the CCP4 suite[69]. Molecular replacement was done using Phaser with an idealized 5 residue polyalanine β-strand[72]. The model was manually refined using Phenix over 2 cycles to the correct 6 residue sequence and rotamer orientation in COOT[74,75]. The atomic model has a final R_{work}/R_{free} of 15.5/18.1% and 100% of Ramachandran angles favored.

MicroED Data Collection and Processing

The MicroED data were collected using the same protocol as Rodriguez et al. with minor differences detailed below[80–82]. Data was collected at HHMI Janelia Research Campus using the Tecnai F20 TEM operating at 200 kV recorded using a bottom mount TVIPS F416 CMOS camera.

For NFGEFS, Individual image frames were exposed for 3s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.2° per second. Diffraction data were collected and merged from 4 individual crystals. Indexing and integration were done using XDS[71] in $P2_12_12$ and scaled with SCALEPACK[79] to 1.0Å resolution. The data was imported using the CCP4 suite[69]. The structure was solved by direct methods using the program SHELX[76]. The model was manually adjusted over 4 cycles to the correct sequence and rotamer orientation in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 22.0/27.0% and 75% of Ramachandran angles favored as well as 25% allowed.

For NFGpTFS, Individual image frames were exposed for 3s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.295° per second. Diffraction data was collected on a single crystal. Indexing and integration were done using XDS in $P2_12_12_1$ and scaled with XSCALE to 0.75 Å resolution[71]. The data was imported using the CCP4 suite[69]. The structure was solved by direct methods using the program SHELX[76]. The model was manually adjusted over 16 cycles to the correct sequence and rotamer orientation in COOT[74]. The atomic model was refined by Refmac[77] with a final R_{work}/R_{free} of 23.2/25.1% and 100% of Ramachandran angles favored.

For SWGMMGMLASQ, individual images were exposed for 2s. During each exposure, the crystals were continuously rotated within the beam at a rate of 0.315° per second. Diffraction data were collected on seven crystals. Indexing and integration were performed XDS in space group P1 and scaled with XSCALE to 1.5 Å resolution[71]. The data were imported using the CCP4 suite[69]. The structure was solved by molecular replacement using the program phaser[72]and a search model consisting of an 9-alanine-long ideal beta-strand. The model was manually adjusted in COOT[74]. The atomic model was refined by Refmac[77] and Buster[83] with a final R_{work}/R_{free} of 28.0/31.3% and 94.5% of Ramachandran angles favored as well as 5.5% allowed. These R-factors are higher than expected for a data set of this quality. We think the discrepancy is due to a disorder in the crystal. The P1 unit cell dimensions of SWGMMGLASQ allow pseudo-merohedral twinning (pseudo space group C2). Also, the native Patterson map suggests that the crystal contains a translocation disorder (peak at 0.379, −0.264, −0.471 with height of 36% of the origin peak). Attempts to correct for these disorders did not improve the R-factors significantly.

Structure Analysis

All images of crystallographic structures were generated in Pymol. Area buried and shape complementarity values were calculated based on published methods[84,85].

Data Availability

The density maps and atomic coordinates of all models reported here have been deposited to the Protein Data Bank (PDB). The MicroED model has also been reported to the EM Data Bank. Accession codes are provided here: GNNQGSN (5WKD); AMMAAA (6CEW); AALQSS (6CB9); SWGMMGMLASQ (6CFH, EMD-7467); GFNGGFG (5WIQ); GNNSYS (5WIA); NFGAFS (5WHN); NFGTFS (5WHP); NFGEFS (5WKB, EMD-8857); NFGpTFS (6CF4, EMD-7466). All other data are available upon reasonable requests to the authors.

Accession Codes Availability

The fibril diffraction simulation code is available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Sangwan and P. Seidler for discussion, D. Shi and T. Gonen at Janelia for microscope support, and M. Collazo at UCLA-DOE Macromolecular Crystallization Core Technology Center for crystallization support. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Eiger 16M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This research was supported in part by grants from the National Institutes of Health NIH NIA AG029430, NIH NIA AG054022, Howard Hughes Medical Institute and the Janelia Research Campus visitor program. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by

UCLA and by instrumentation grants from NIH (1S10RR23057 and 1U24GM116792) and NSF (DBI-1338135). D.R.B was supported by the National Science Foundation Graduate Research Fellowship Program.

Abbreviations

Ab	Area buried
FUS	Fused in sarcoma
LARKS	Low complexity aromatic-rich kinked segment
LCD	Low complexity domain
CTF	C-terminal fragment
LLPS	Liquid-liquid phase separation
MicroED	Micro electron diffraction
pT	phosphorylated threonine
SG	stress granule
TDP-43	TAR DNA Binding Protein 43

References

1. Molliex A, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell. 2015; 163:123–133. [PubMed: 26406374]
2. Kato M, et al. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012; 149:753–767. [PubMed: 22579281]
3. Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure. 2016; 24:1537–1549. [PubMed: 27545621]
4. Kato M, Mcknight SL. A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. Annu. Rev. Biochem. 2018; 87
5. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. A census of protein repeats. J. Mol. Biol. 1999; 293:151–160. [PubMed: 10512723]
6. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 2005; 6:197–208. [PubMed: 15738986]
7. Zoghbi HY, Orr HT. Glutamine Repeats and Neurodegeneration. Annu. Rev. Neurosci. 2000; 23:217–247. [PubMed: 10845064]
8. March ZM, King OD, Shorter J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. Brain Res. 2016; 1647:9–18. [PubMed: 26996412]
9. Bentmann E, et al. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). J. Biol. Chem. 2012; 287:23079–23094. [PubMed: 22563080]
10. Colombrita C, et al. TDP-43 is recruited to stress granules in conditions of oxidative insult. J. Neurochem. 2009; 111:1051–1061. [PubMed: 19765185]
11. Dewey CM, et al. TDP-43 Is Directed to Stress Granules by Sorbitol, a Novel Physiological Osmotic and Oxidative Stressor. Mol. Cell. Biol. 2011; 31:1098–1108. [PubMed: 21173160]
12. Gilks N, et al. Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1. Mol. Biol. Cell. 2004; 15:5383–5398. [PubMed: 15371533]

13. Dang Y, et al. Eukaryotic initiation factor 2α-independent pathway of stress granule induction by the natural product pateamine A. J. Biol. Chem. 2006; 281:32870–32878. [PubMed: 16951406]
14. Kedersha N, et al. Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. J. Cell Biol. 2000; 151:1257–1268. [PubMed: 11121440]
15. Patel A, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell. 2015; 162:1066–1077. [PubMed: 26317470]
16. Reineke LC, et al. Casein Kinase 2 Is Linked to Stress Granule Dynamics through Phosphorylation of the Stress Granule Nucleating Protein G3BP1. Mol. Cell. Biol. 2017; 37:e00596–16. [PubMed: 27920254]
17. Murray DT, et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. Cell. 2017; 171:615–627.e16. [PubMed: 28942918]
18. Wheeler JR, Matheny T, Jain S, Abrisch R, Parker R. Distinct stages in stress granule assembly and disassembly. Elife. 2016; 5:1–25.
19. Jain A, Vale RD. RNA phase transitions in repeat expansion disorders. Nature. 2017; 546:243–247. [PubMed: 28562589]
20. Lin Y, Protter DSW, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol. Cell. 2015; 60:208–219. [PubMed: 26412307]
21. Hughes M, et al. Atomic structures of low-complexity protein segments reveal kinked β-sheets that assemble networks. Science. 2018; 359:698–701. [PubMed: 29439243]
22. Eisenberg DS, Sawaya MR. Structural Studies of Amyloid Proteins at the Molecular Level. Annu. Rev. Biochem. 2017; 86:69–95. [PubMed: 28125289]
23. Knowles TPJ, Vendruscolo M, Dobson CM. The amyloid state and its association with protein misfolding diseases. Nat. Rev. Mol. Cell Biol. 2014; 15:384–96. [PubMed: 24854788]
24. Meersman F, Dobson CM. Probing the pressure–temperature stability of amyloid fibrils provides new insights into their molecular properties. Biochim. Biophys. Acta - Proteins Proteomics. 2006; 1764:452–460.
25. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. Biochem. J. 2017; 474:1417–1438. [PubMed: 28389532]
26. Chang, Cke, et al. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. Biochem. Biophys. Res. Commun. 2012; 425:219–224. [PubMed: 22835933]
27. Lukavsky PJ, et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. Nat. Struct. Mol. Biol. 2013; 20:1443–1449. [PubMed: 24240615]
28. Saini A, Chauhan VS. Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. Langmuir. 2014; 30:3845–3856. [PubMed: 24559403]
29. Buratti E, Baralle FE. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. RNA Biol. 2010; 7:420–429. [PubMed: 20639693]
30. Lee EB, Lee VMY, Trojanowski JQ. Gains or losses: Molecular mechanisms of TDP43-mediated neurodegeneration. Nat. Rev. Neurosci. 2012; 13:38–50.
31. Schwab C, Arai T, Hasegawa M, Yu S, Mcgeer PL. Colocalization of Transactivation-Responsive DNA-Binding Protein 43 and Huntingtin in Inclusions of Huntington Disease. J Neuropathol Exp Neurol. 2008; 67:1159–1165. [PubMed: 19018245]
32. Amador-Ortiz C, et al. TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. Ann. Neurol. 2007; 61:435–445. [PubMed: 17469117]
33. Nakashima-Yasuda H, et al. Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. Acta Neuropathol. 2007; 114:221–229. [PubMed: 17653732]
34. Higashi S, et al. Concurrence of TDP-43, tau and α-synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. Brain Res. 2007; 1184:284–294. [PubMed: 17963732]
35. Neumann M, et al. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science. 2006; 314:130–133. [PubMed: 17023659]
36. Kabashi E, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. 2008; 40:572–574. [PubMed: 18372902]

37. Chiang C-H, et al. Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. Sci. Rep. 2016; 6:21581. [PubMed: 26883171]
38. Lagier-Tourenne C, Polymenidou M, Cleveland DW. TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. 2010; 19:46–64.
39. Pesiridis GS, Lee VMY, Trojanowski JQ. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. Hum. Mol. Genet. 2009; 18:156–162.
40. Igaz LM, et al. Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. J. Biol. Chem. 2009; 284:8516–8524. [PubMed: 19164285]
41. Hasegawa M, et al. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Ann. Neurol. 2008; 64:60–70. [PubMed: 18546284]
42. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. Neuron. 2013; 79:416–438. [PubMed: 23931993]
43. Ling JP, Pletnikova O, Troncoso JC, Wong PC. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science (80-.). 2015; 349:650–655.
44. Yang C, et al. The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. PLoS One. 2010; 5:e15878. [PubMed: 21209826]
45. Chen AK, et al. Induction of Amyloid Fibrils by the C-Terminal Fragments of TDP-43 in Amyotrophic Lateral Sclerosis. J. Am. Chem. Soc. 2010; 132:1186–1187. [PubMed: 20055380]
46. Cushman M, Johnson BS, King OD, Gitler AD, Shorter J. Prion-like disorders: blurring the divide between transmissibility and infectivity. J. Cell Sci. 2010; 123:1191–1201. [PubMed: 20356930]
47. Guo W, et al. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. Nat. Struct. Mol. Biol. 2011; 18:822–830. [PubMed: 21666678]
48. Jiang LL, et al. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. J. Biol. Chem. 2013; 288:19614–19624. [PubMed: 23689371]
49. Jiang LL, et al. Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. Sci. Rep. 2016; 6:23928. [PubMed: 27030292]
50. Mompeán M, et al. Structural characterization of the minimal segment of TDP-43 competent for aggregation. Arch. Biochem. Biophys. 2014; 545:53–62. [PubMed: 24440310]
51. Sawaya MR, et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature. 2007; 447:453–7. [PubMed: 17468747]
52. Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc. Natl. Acad. Sci. 2010; 107:3487–92. [PubMed: 20133726]
53. Budini M, et al. Cellular model of TAR DNA-binding Protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. J. Biol. Chem. 2012; 287:7512–7525. [PubMed: 22235134]
54. Fuentealba RA, et al. Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. J. Biol. Chem. 2010; 285:26304–26314. [PubMed: 20554523]
55. Seidler PM, et al. Structure-based inhibitors of tau aggregation. Nat. Chem. 2018; 10:170–6. [PubMed: 29359764]
56. Xiang S, et al. The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. Cell. 2015; 163:829–39. [PubMed: 26544936]
57. Gitcho MA, et al. TDP-43 A315T mutation in familial motor neuron disease. Ann. Neurol. 2008; 63:535–538. [PubMed: 18288693]
58. Xu M, et al. Characterization of b -domains in C-terminal fragments of TDP-43 by scanning tunneling microscopy. J. Struct. Biol. 2013; 181:11–16. [PubMed: 23138004]
59. Lim L, Wei Y, Lu Y, Song J. 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. 2016; 14:1–35.
60. Estes PS, et al. Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a Drosophila model of ALS. Hum. Mol. Genet. 2011; 20:2308–2321. [PubMed: 21441568]

61. Walker AK, et al. ALS-associated TDP-43 induces endoplasmic reticulum stress, which drives cytoplasmic TDP-43 accumulation and stress granule formation. PLoS One. 2013; 8:1–12.
62. Bosco DA, et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Hum. Mol. Genet. 2010; 19:4160–4175. [PubMed: 20699327]
63. Hackman P, et al. Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. Ann. Neurol. 2013; 73:500–509. [PubMed: 23401021]
64. Kim HJ, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013; 495:467–473. [PubMed: 23455423]
65. Schmidt HB, Rohatgi R. In Vivo Formation of Vacuolated Multi-phase Compartments Lacking Membranes. Cell Rep. 2016; 16:1228–1236. [PubMed: 27452472]
66. Franzmann TM, et al. Phase separation of a yeast prion protein promotes cellular fitness. Science. 2018; 359:eaao5654. [PubMed: 29301985]
67. Fitzpatrick AWP, et al. Cryo-EM structures of tau filaments from Alzheimer's disease. Nature. 2017; 547:185–190. [PubMed: 28678775]
68. Soragni A, et al. A Designed Inhibitor of p53 Aggregation Rescues p53 Tumor Suppression in Ovarian Carcinomas. Cancer Cell. 2016; 29:90–103. [PubMed: 26748848]
69. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography. 2011; 67:235–242. [PubMed: 21460441]
70. Cheng P-N, Liu C, Zhao M, Eisenberg D, Nowick JS. Amyloid β-sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. Nat. Chem. 2012; 4:927–933. [PubMed: 23089868]
71. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]
72. McCoy AJ, et al. Phaser crystallographic software. J. Appl. Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
73. Wiltzius JEDJW, et al. Atomic structure of the cross-b spine of islet amyloid polypeptide (amylin). Protein Sci. 2008; 17:1467–1474. [PubMed: 18556473]
74. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
75. Adams PD, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
76. Sheldrick G. A short history of SHELX. Acta Crystallogr A. 2008; 64:112–122. [PubMed: 18156677]
77. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D: Biological Crystallography. 1997; 53:240–255. [PubMed: 15299926]
78. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. 1997; 276:307–326.
79. Kabasch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. 1993; 26:795–800.
80. Rodriguez JA, et al. Structure of the toxic core of α-synuclein from invisible crystals. Nature. 2015; 525:486–490. [PubMed: 26352473]
81. Hattne J, et al. MicroED data collection and processing. Acta Crystallogr. Sect. A Found. Adv. 2015; 71:353–360. [PubMed: 26131894]
82. Shi D, et al. The collection of MicroED data for macromolecular crystallography. Nat. Protoc. 2016; 11:895–904. [PubMed: 27077331]
83. Bricogne, G., et al. BUSTER version 2.9. United Kindom Glob. Phasing Ltd; 2010.
84. Lawrence M, Colman P. Shape complemntarity at protein/protein interfaces. J Mol Biol. 1993; 234:956–950.
85. Anonymous. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

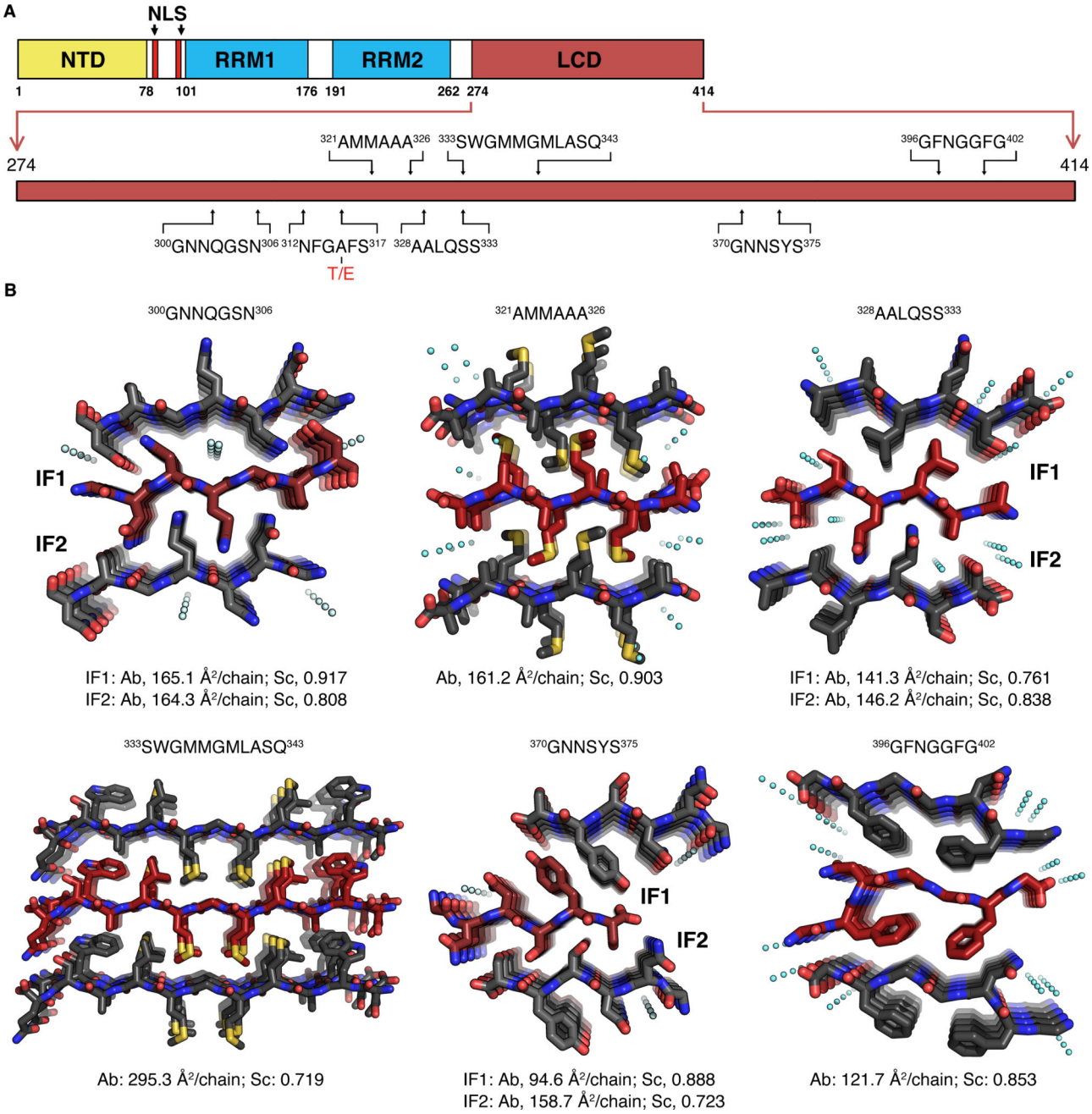

Figure 1. Segments from the Low Complexity Domain (LCD) of TDP-43 form steric zippers

(A) Schematic representation of full length TDP-43, the low complexity domain is enlarged and the sequences and relative positions of segments for which we determined structures are shown. (B) Six steric-zipper structures we determined in this study. For each structure, three β-sheets and two interfaces between these sheets are shown, and each sheet shows 5 of its thousands of strands. For ^{300}GNNQGSN^{306}, ^{328}AALQSS^{333} and ^{370}GNNSYS^{375}, two interfaces (IF) are distinct and are labeled as IF1 and IF2; For ^{321}AMMAAA^{326}, ^{333}SWGMMGMLASQ^{343} and ^{396}GFNGGFG^{402}, all interfaces are identical because of the symmetry of packing. The area buried (Ab) and shape complementary (Sc) values for each

interface are shown below the structure. Water molecules found in the structure are shown as cyan droplets. All the structures are shown in top view, viewed down the fibril axis.

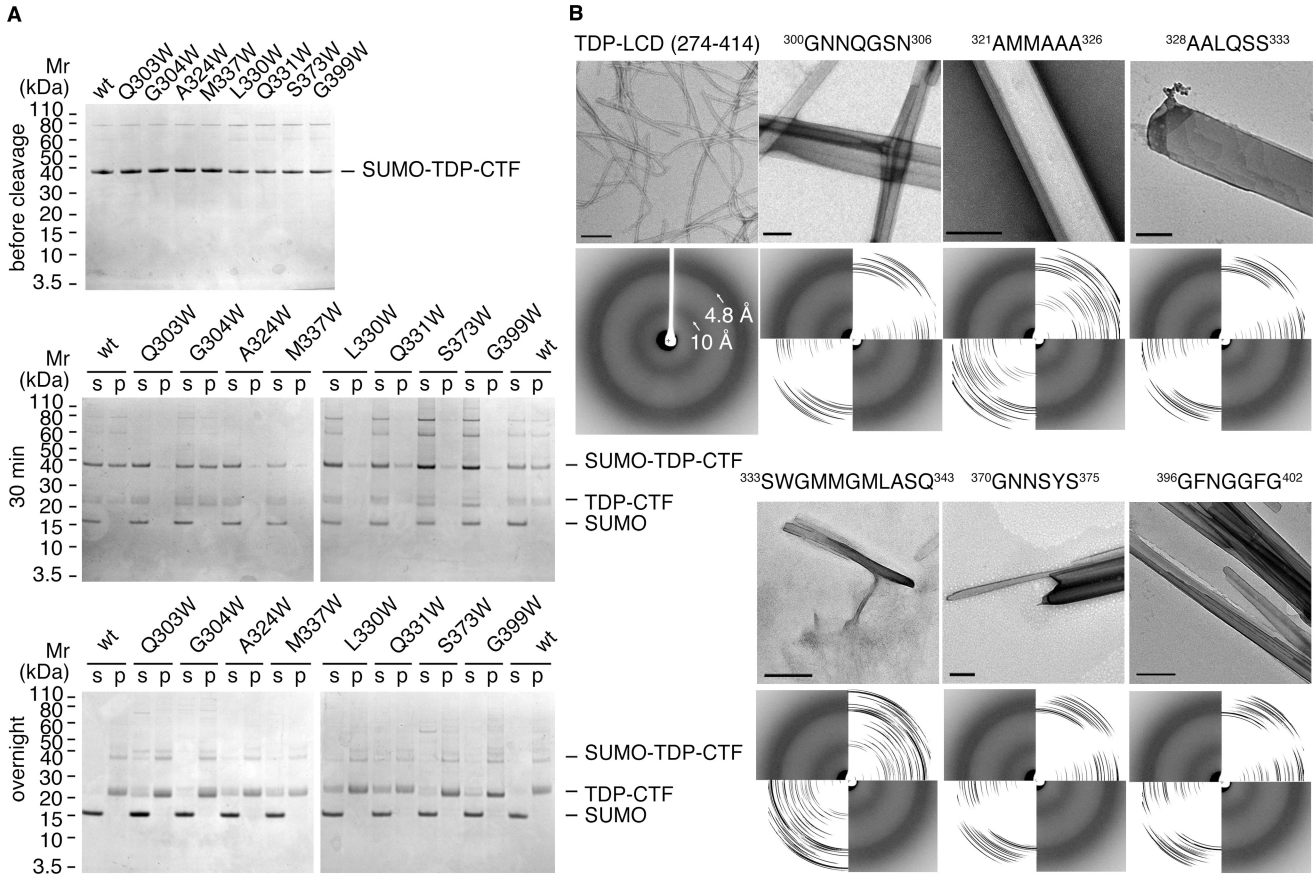

Figure 2. Validation of steric-zipper forming segments

(A) Aggregation assays of the pathological fragment TDP-CTF (208–414), both wild type (wt) and with mutations to block steric-zipper formation. TDP-CTF was conjugated with a SUMO protein tag to prevent aggregation, and ULP1 protease was added to remove the SUMO tag. TDP-CTF was incubated on ice for 30 mins after cleavage and continued to incubate at 4 °C overnight. Samples were taken after incubation and separated into supernatant (s) and pellet (p) by centrifugation, and analyzed by SDS-PAGE. Notice that TDP-CTF G304W has similar behavior to wild type TDP-CTF, whereas other mutants show reduced aggregation compared to wild type. (B) Negative stain EM and fibril diffraction pattern of TDP-LCD and steric-zipper segments. TDP-LCD was conjugated with SUMO tag and mixed with ULP1 protease to remove SUMO tag. The diffraction pattern of TDP-LCD fibrils was acquired by X-ray diffraction and the diffraction patterns of zipper forming segments were computationally simulated from their crystal structures. The diffraction pattern of TDP-LCD fibrils was overlaid with the ones of zipper-forming segments for comparison. Notice that TDP-LCD fibrils generate diffraction rings at 4.8 Å and 10 Å, indicating that they have cross-β spines. Scale bar = 200 nm.

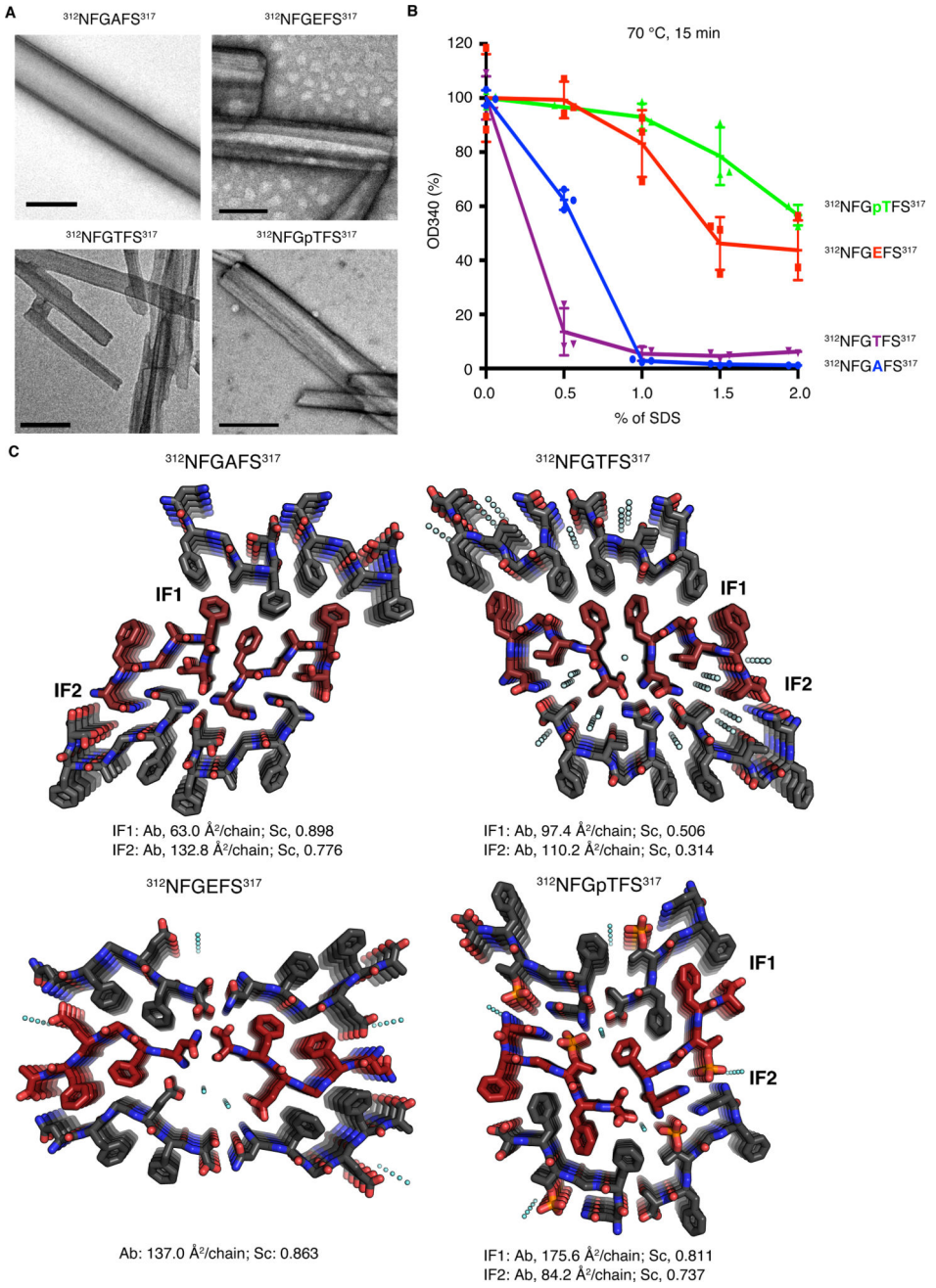

Figure 3. The ^{312}NFGAFS^{317} segments forms a kinked β-sheet that is strengthened by familial variants A315T and A315E

(A) Negative stain EM images of aggregates of ^{312}NFGAFS^{317} and its variants after 4 days of shaking at 37 °C. Scale bar = 200 nm. (B) Denaturing assays of aggregates of ^{312}NFGAFS^{317} and its variants. The aggregates of each segment were treated with the indicated percentage of SDS, and heat to 70 °C for 15 mins. The absorbance at wavelength 340 (OD340) of the solution after treatment was measured to represent turbidity. The readings from same segment were shown as percentage of the mean values of the segment treated with 0% SDS and without heat, and were shown as mean ± SD (n=3 independent experiments) with individual data overlaid. (C) Structures of ^{312}NFGAFS^{317} and its variants.

For each structure, six sheets are shown and two interfaces (IF1 and IF2) are labeled between three parallel packed sheets. For ^{312}NFGEFS^{317}, the two interfaces between the parallel packed sheets are identical because of its face-to-back packing, so the interfaces are not labeled. The area buried (Ab) and shape complementary (Sc) values for each interface are shown below the structure. Water molecules found in the structure are shown in cyan droplets. All the structures are shown viewed down the fibril axis. Notice that the larger interfaces in both ^{312}NFGEFS^{317} and ^{312}NFGpTFS^{317} are increased over those in ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317}, explaining in part their greater stability. See Supplementary Fig. 3 for additional hydrogen bond stabilization.

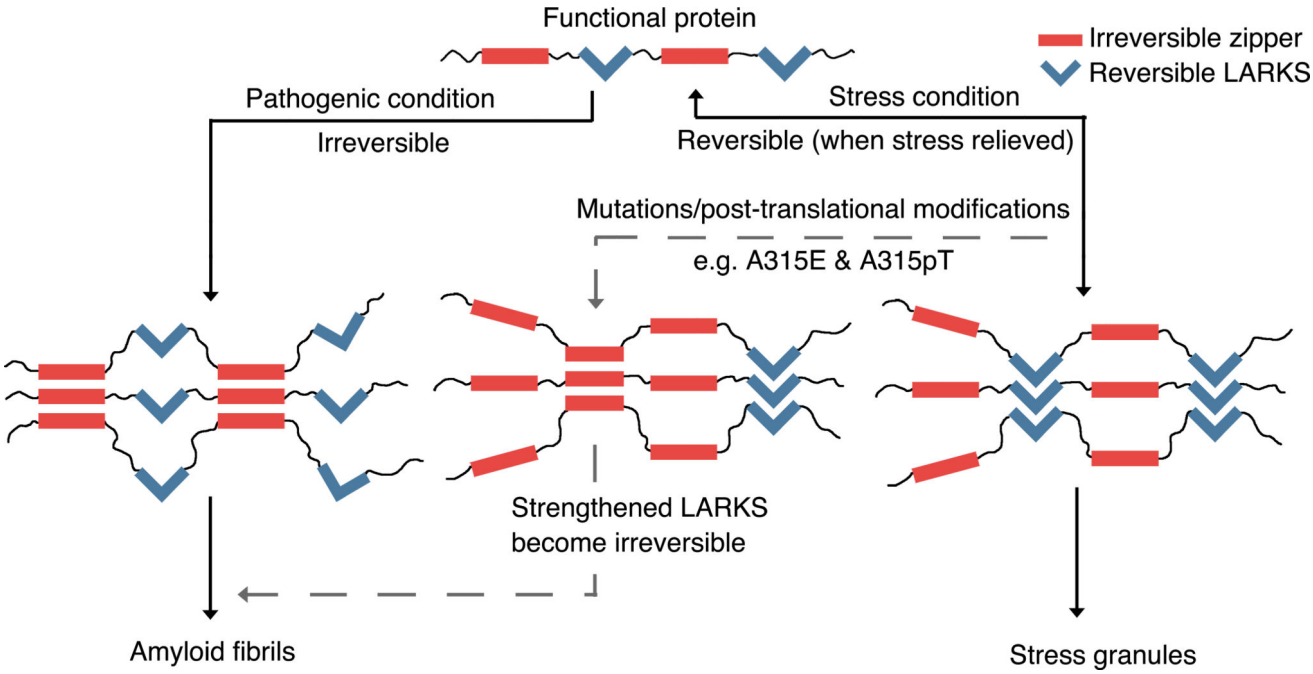

Figure 4. Speculative model showing the alternative pathways of formation of stress granules with pathogenic amyloid

TDP-43 forms several irreversible steric zippers (represented by red bars) and several reversible LARKS (represented by blue kinked bars). When performing its normal functions, the protein is not aggregated (top panel). Under pathogenic conditions (left pathway), the steric zipper segments interact with each other, acting as molecular glue to drive the protein into irreversible aggregation, and forming amyloid fibrils; under stress conditions (right pathway), the LARKS interact with each other, and acting as molecular Velcro, providing adhesion to form a hydrogel as in stress granules; these interactions are reversible. The crystal structures of ^{312}NFGAFS^{317}, ^{312}NFGEFS^{317} and ^{312}NFGpTFS^{317} show that A315E and A315T with phosphorylated Thr can strengthen the LARKS to irreversible aggregates. Either mutation or post-translational modifications can act as the agent of conversion (middle pathway). Since the involvement of ^{312}NFGAFS^{317} in stress granules formation is yet to be validated, the middle pathway is shown by gray dashed lines, emphasizing that this pathway is speculative.

Table 1

Statistics of x-ray diffraction data collection and atomic refinement.

	GNNQGSN (5WKD)	AMMAAA (6CEW)	AALQSS (6CB9)	GFNGGFG (5WIQ)	GNNSYS (5WIA)	NFGAFS (5WHN)	NFGTFS (5WHP)
Data Collection							
Space group	C2	$P2_12_12_1$	$P2_12_12_1$	P1	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Cell dimensions							
a,*b*,*c* (Å)	50.35, 4.77, 14.75	9.52, 15.44, 44.1	4.79, 15.97, 43.43	4.81, 16.44, 22.67	4.79, 15.62,40.68	13.82, 4.85, 46.74	15.33, 4.84, 23.57
α,β,γ(°)	90, 101.73, 90	90, 90, 90	90, 90, 90	90.63, 96.37, 91.26	90, 90, 90	90, 90, 90	90.00, 96.91, 90.00
Resolution(Å)	1.80 (1.86–1.80)[a]	1.20 (1.23–1.20)	1.10 (1.14–1.10)	1.24 (1.29–1.25)	1.00 (1.04–1.00)	1.10 (1.14–1.10)	1.00 (1.03–1.00)
R_{merge} (%)	15.8 (25.0)	18.1 (68.3)	21.1 (56.2)	16.1 (40.6)	15.5 (55.4)	17.2 (72.6)	8.6 (10.4)
$I/\sigma(I)$	6.9 (3.3)	3.2 (1.4)	5.1 (2.8)	7.1 (2.0)	9.2 (2.1)	9.7 (2.5)	15.0 (7.6)
$CC_{1/2}$	97.1 (81.6)	99.5 (69.2)	97.0 (91.3)	98.5 (69.0)	99.2 (66.8)	99.1 (82.3)	99.3 (98.0)
Completeness (%)	91.6 (84.8)	96.0 (99.3)	90.7 (72.1)	94.4 (95.4)	95.2 (62.7)	94.0 (75.9)	87.6 (37.0)
Redundancy	2.3 (2.2)	3.5 (3.5)	6.9 (7.3)	4.1 (1.8)	5.4 (1.9)	8.3 (6.4)	4.9 (2.8)
Refinement							
Resolution (Å)	24.32–1.80	22.05–1.20	21.22–1.10	22.53–1.24	20.34–1.00	13.25–1.10	15.22–1.00
No. reflections	345	2,078	1,288	1,796	1,836	1,300	1,639
R_{work}/R_{free}	18.3/19.4	19.8/21.8	19.5/24.3	15.2/16.8	15.5/18.1	13.3/16.0	8.7/9.3
No. atoms							
Protein	48	79	40	94	45	50	48
Water	2	4	4	8	1	0	3
B factors (Å^2)							
Protein atoms	8.7	9.1	7.3	3.2	3.5	4.6	2.4
Water	12.2	20.7	20.0	7.6	15.3	-	14.8
R.m.s. deviations							
Bond lengths (Å)	0.021	0.012	0.010	0.005	0.005	0.006	0.019
Bond angles (°)	1.804	1.431	1.422	0.692	1.18	0.95	1.79

[a]Values in parentheses are for highest-resolution shell.

Table 2

Statistics of MicroED data collection and atomic refinement.

	SWGMMGMLASQ (EMD-7467, 6CFH)	NFGEFS (EMD-8857, 5WKB)	NFGpTFS (EMD-7466, 6CF4)
Data Collection			
Space group	P1	$P2_12_12$	$P2_12_12_1$
Cell dimensions			
a, b, c (A°)	8.56, 9.60, 39.97	42.77, 17.42, 4.90	23.65, 4.72, 30.06
α, β, γ (°)	97.2, 92.9, 105.9	90, 90, 90	90, 90, 90
Resolution (Å)	1.50 (1.55–1.50)[a]	1.00 (1.03–1.00)	0.75 (0.77–0.75)
R_{merge} (%)	20.8 (85.5)[b]	28.3 (99.3)[c]	17.2 (66.1)
$I/\sigma(I)$	3.3 (0.7)	4.6 (1.1)	3.9 (1.3)
$CC_{1/2}$	98.7 (72.0)	99.2 (36.0)	98.9(59.9)
Completeness (%)	93.5 (89.4)	88.7 (78.0)	86.6(87.2)
Redundancy	4.2 (2.4)	9.5 (4.8)	3.8 (4.0)
Refinement			
Resolution (Å)	13.17–1.50	21.39–1.00	7.65–0.75
No. reflections	1,819	1,801	4,178
R_{work}/R_{free}	28.0/31.3	22.0/27.0	23.2/25.1
No. atoms			
Protein	162	90	93
Water	0	2	1
B factors			
Protein atoms	18.1	5.3	19.6
Water	-	25.3	3.1
R.m.s. deviations			
Bond lengths (Å)	0.007	0.016	0.021
Bond angles (°)	0.91	1.09	2.04

[a] Highest resolution shell shown in parenthesis.

[b] Data are from 7 crystals.

[c] Data are from 4 crystals.