



HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2020 February 28.

Published in final edited form as:

Nat Struct Mol Biol. 2019 July ; 26(7): 619–627. doi:10.1038/s41594-019-0248-4.

Cryo-EM Structures of Four Polymorphic TDP-43 Amyloid Cores

Qin Cao^{1,†}, David R. Boyer^{1,†}, Michael R. Sawaya¹, Peng Ge², David S. Eisenberg^{1,*}

¹Department of Chemistry and Biochemistry and Biological Chemistry, UCLA-DOE Institute, Molecular Biology Institute, and Howard Hughes Medical Institute, UCLA, Los Angeles CA 90095

²California NanoSystem Institute, UCLA, Los Angeles CA 90095

Abstract

The DNA/RNA processing protein TDP-43 undergoes both functional and pathogenic aggregation. Functional TDP-43 aggregates form reversible, transient species such as nuclear bodies, stress granules, and myo-granules. Pathogenic, irreversible TDP-43 aggregates form in amyotrophic lateral sclerosis (ALS) and other neurodegenerative conditions. Here we find the features of TDP-43 fibrils that confer both reversibility and irreversibility by determining structures of two segments reported to be the pathogenic cores of human TDP-43 aggregation: SegA (residues 311–360), which forms three polymorphs, all with dagger-shaped folds; and SegB A315E (residues 286–331 containing the ALS hereditary mutation A315E), which forms R-shaped folds. Energetic analysis suggests that the dagger-shaped polymorphs represent irreversible fibril structures, whereas the SegB polymorph may participate in both reversible and irreversible fibrils. Our structures reveal the polymorphic nature of TDP-43 and suggest how the A315E mutation converts the R-shaped polymorph to an irreversible form which enhances pathology.

INTRODUCTION

Amyloid-forming proteins seem to violate the central tenant of protein science—that amino acid sequence determines structure and function¹. In contrast to globular and membrane

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence to: David S. Eisenberg, david@mbi.ucla.edu.

†These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

Q.C. and D.R.B. designed experiments and performed data analysis. Q.C. expressed and purified constructs, and performed biochemical experiments. Q.C. and D.R.B. prepared cryo-EM samples, and performed cryo-EM data collection and processing. P.G. assisted in cryo-EM data collection and processing. M.R.S. performed solvation energy calculation. All authors analyzed the results and wrote the manuscript. D.S.E. supervised and guided the project.

COMPETING INTERESTS

D.S.E. is an advisor and equity shareholder in ADRx, Inc.

CODE AVAILABILITY

Custom software used for solvation energy calculation and for generating multi-layer models is available from the authors upon request.

DATA AVAILABILITY

All structural data have been deposited into the Worldwide Protein Data Bank (wwPDB) and the Electron Microscopy Data Bank (EMDB) with the following accession codes: SegA-sym (PDB 6N37, EMD-9339), SegA-asym (PDB 6N3B, EMD-9350), SegA-slow (PDB 6N3A, EMD-9349), SegB A315E (PDB 6N3C, EMD-0334). Coordinates for structural model of SegB A315E G304W are in Supplementary Data Set 1. All other data are available from the authors upon reasonable request.

proteins each of which folds into a single functional structure, a given amyloid-forming sequence can fold into several distinctly different polymorphic structures^{2,3}.

Irreversible, hyperphosphorylated, amyloid-like aggregates of C-terminal segments of TDP-43 are the primary pathology of amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP)⁴⁻⁶. These aggregates have also been found in Alzheimer's, Parkinson's, CTE, Huntington's disease, Limbic-predominant age-related TDP-43 encephalopathy (LATE) and inclusion body myopathies (IBMs), among others⁷⁻¹¹. Because TDP-43 functions in several essential steps of RNA metabolism^{12,13}, it is widely considered that TDP-43 aggregation is toxic through a loss-of-function mechanism¹⁴⁻¹⁶. Structural studies of amyloid fibrils of β -amyloid^{17,18}, tau^{19,20}, α -synuclein², and β 2-microglobulin²¹ have revealed polymorphs and insights into pathogenesis. Here, we use cryo-EM to determine the overall folds of TDP-43 amyloid cores, expanding structural information about both reversible and irreversible aggregation beyond the local interactions previously revealed by crystallography^{22,23}.

RESULTS

Generating TDP-43 fibrils

We screened for TDP-43 fibrils by incubating various SUMO-tagged TDP-43 segments in the test tube after the cleavage of SUMO tag (Supplementary Figure 1a and Supplementary Table 1). We first aimed to produce fibrils formed by full-length TDP-43, a pathogenic C-terminal fragment (CTF, 208–414) – truncation product that is enriched in disease brain²⁴, or by the Low Complexity Domain (LCD, 274–414) which is considered to be necessary for TDP-43 aggregation^{22,25,26}. However, despite our efforts at optimization, we could observe only highly clumped fibril-like structures and disordered aggregates that are not suitable for cryo-EM structure determination (Supplementary Figure 1b top panel). We suspect that this may be due to the ability of longer constructs of TDP-43 to participate in multi-valent interactions, possibly through LARKS^{22,27} or other adhesive segments outside the LCD²³. These multi-valent interactions have been shown to assemble networks of protein chains and could conceivably explain why longer segments of TDP-43 form amorphous aggregates or fibril clumps. This observation is in line with TDP-43's role in phase separation and stress granule formation, which requires the presence of multi-valent interactions²⁷.

To overcome the hurdle of the disordered assembly of longer segments of TDP-43, we implemented a “divide and conquer” approach whereby we selected known aggregation cores for structure determination. We chose SegA (residues 311–360) and SegB (residues 286–331), guided by the following evidence. SegA was previously identified as an aggregation core of TDP-43 due to the observation that its deletion decreases TDP-43 aggregation in vitro and in cells, whereas addition of SegA to the aggregation-resistant *C. elegans* TDP-43 homolog induces aggregation²⁸. Fibrils of SegB are toxic to primary neurons, and an ALS hereditary mutation A315T together with phosphorylation of the threonine, which is speculated to occur in hyper-phosphorylated aggregates of TDP-43 in disease, increases SegB's cytotoxicity²⁹. With this in mind, we selected SegB A315E – another hereditary mutation with similar effects as A315T²⁹⁻³² and a mimic of A315T with phosphorylation – in order to visualize the structure of a second possible TDP-43

aggregation core and to gain insight into the molecular mechanism of mutation-enhanced TDP-43 pathology. The importance of SegA and SegB in full-length TDP-43 aggregation is also supported by other studies which found that amyloid fibrils containing either a core region (residues 314–353) of SegA or region (residues 274–313) similar to SegB can template aggregation of full-length TDP-43 in SH-SY5Y human neuronal cells³³. Likewise, in the same cell line, deletion of these two regions (residues 314–353 or 274–313) from full-length TDP-43 inhibits aggregation.

Architecture of SegA and SegB fibrils

As we expected, fibrils formed by SegA and SegB A315E were much more homogenous and less bundled than longer segments of TDP-43, including SegAB (286–360) that contains both aggregation cores SegA and SegB (Supplementary Figure 1b). This observation supports the idea that eliminating competing multi-valent interactions helps to produce homogenous fibrils of isolated amyloid cores. Using these homogenous preparations, we determined three polymorphic fibril structures of SegA (termed SegA-sym, SegA-asym, and SegA-slow) and one of SegB A315E (Figure 1, Supplementary Figure 2–4 and Table 1; see Online methods for technical details).

All four fibrils are formed from gradually twisting β -sheets that run the entire length of the fibrils (Figure 1b–e). A thin slab or “layer” perpendicular to each fibril axis (Figure 1, middle) shows that individual SegA and SegB chains are each confined within an essentially two-dimensional layer, in contrast to globular and membrane proteins whose folds occupy three-dimensions. Identical layers stack on one another creating β -sheets that are parallel and in-register. Each layer contains two or more protein chains, giving rise to a corresponding number of protofilaments in the fibril.

The dagger-shaped fold in SegA polymorphs

SegA polymorphs all share a dagger-shaped fold comprised of residues 312–346 with the hydrophobic residues from Phe313 to Ala341 forming tight hydrophobic interactions, and a sharp 160° kink at Gln327 forming the dagger tip (Figure 2a–d and Supplementary Figure 5, detailed comparison in following two paragraphs and Supplementary Figures 5,6). The three SegA polymorphs mainly differ in the number of protofilaments and symmetry. SegA-sym fibrils contain two protofilaments related by a pseudo-2₁ axis, whereas SegA-asym contains two protofilaments with somewhat different conformations. SegA-slow fibrils contain four protofilaments; two protofilaments are related by a central 2-fold axis and contain 50 ordered residues. These are flanked by two other protofilaments containing only 10 ordered residues.

The core region of the dagger-shaped fold (residue 320–334) is almost identical in all four dagger-shaped folds of three SegA polymorphs (Figure 2a and Supplementary Figure 5). Besides that, detailed single chain superposition shows that SegA-sym and the long chain of SegA-asym have nearly identical dagger-shaped folds, with the “stem” of the long chain SegA-asym (Ala341 to Ser347) bent approximately 40° (Figure 2c). This bending of the long chain SegA-asym stem forces a 40° bend in the “blade” of the SegA-asym short chain relative to the reference SegA-sym fold, corresponding to the shift in registration of the stem

hydrophobic interdigitation (Supplementary Figure 5, bottom left). In addition, the “stem” of SegA-slow (Ala341 to Ser347) is bent approximately 50° compared to SegA-sym (Supplementary Figure 5, bottom right), causing a significant conformational change at Met336. Further, the interaction of the termini in SegA-slow causes a conformational change at Phe313, Phe316 and Met339; and M311 is now visualized and participating in the hydrophobic core (Supplementary Figure 5, bottom right).

The protofilament interface in SegA-sym and SegA-asm are similar (Supplementary Figure 6a,b) with slight differences in the region that has a 40° bend (Supplementary Figure 6b, right). We name this the continuous interface since it is formed by two continuous chains, as opposed to the broken interface in SegA-slow formed by the interaction of the termini (Supplementary Figure 5). The continuous interface has a hydrophobic core formed by Met336 and Leu340 from both monomers and a hydrogen bond network towards the edges (Figure 2d & Supplementary Figure 6a). The broken interface is formed largely by the tight packing of Met359 and the hydrogen bonding network of the peripheral asparagine residues (Supplementary Figure 6a). We observed only part of the continuous interface in SegA-slow, probably due to the conformational change of Met336 (Supplementary Figure 6c,d, see detailed model building in Online methods).

We note that SegA-sym and SegA-asm structures are compatible with full-length TDP-43 due to their free N- and C-termini, whereas the SegA-slow structure is an artifact of the truncation since the N- and C-termini are sequestered in the center of the fibrils (Figure 1d and Supplementary Figure 6a). Therefore, in the following analysis, we mostly focus on SegA-sym and SegA-asm. However, we note that SegA-slow provides valuable information, as discussed below. First, the SegA-slow structure contains the dagger-shaped fold, demonstrating its conservation in the three different polymorphs we observe for SegA (Supplementary Figure 5). We note that the incompatibility of SegA-slow to full-length TDP-43 comes from the tail region (residues 347–360) which is external to the dagger-shaped fold (Supplementary Figure 7a, upper panel). Thus, taken alone, the dagger-shaped fold in SegA-slow (residues 312–346) is still compatible with full-length TDP-43, as in SegA-sym and SegA-asm. Second, the tail region hairpin structure (residues 342–360) in SegA-slow is also compatible with full-length TDP-43 when separated from the dagger-shaped fold, with free N- and C- terminals (Supplementary Figure 7a bottom panel). This hairpin structure is formed by tightly packed hydrophilic zippers stabilized with extensive hydrogen bonds (Supplementary Figure 7b). Based on the analysis above, we speculate that this hairpin structure may also be involved in TDP-43 aggregation as another segmental polymorph. Further study is needed to test this hypothesis as well as identify whether this hairpin contributes to reversible or irreversible aggregation of TDP-43. Third, our SegA-slow structure provides, to our knowledge, the first atomic evidence in support of secondary seeding of amyloid fibrils. Secondary seeding is considered to occur when the side of pre-formed fibrils serves as a template for new monomers to nucleate into an amyloid fibril. This is in contrast to primary seeding where molecules add to the top or bottom of pre-formed fibrils. In this case, the extra density near Q331 and M323 of the long chain of SegA-slow indicates that secondary seeding can occur on these regions (Figure 1d). Because the long chain of SegA-slow contains all 50 residues in SegA, the extra density must come from other SegA molecules, therefore supporting a secondary seeding mechanism. In comparison,

similar extra density has been observed in other amyloid fibril structures, such as the tau Paired Helical Filament (PHF)¹⁹; but in these structures the ordered fibril core does not contain all residues of the molecule. Therefore it is not certain whether these densities come from binding of a different molecule or from sequences outside of the fibril core of the same molecule.

The R-shaped fold in the SegB fibril

SegB A315E forms fibrils of a single morphology, in contrast to polymorphic SegA. These fibrils are characterized by an R-shaped fold spanning residues 288–319 (Figure 2e,f). Overall, the R-shaped fold is more highly kinked than the dagger-shaped fold (Figure 1b–e), probably due to the abundance of glycine residues. Each fibril is wound from four protofilaments (Figure 2e). The two inner protofilaments are related through a tight, pseudo- 2_1 symmetric interface (Figure 2e,g). These protofilaments are flanked by two outer protofilaments through an asymmetric interface (Figure 2e,g). The symmetric interfaces are formed by small side chain residues (Gly298 and Gly300) in the middle (reminiscent of the homodimer interface in the tau PHF structure from Alzheimer's disease¹⁹) and hydrogen bonds among Gln303, Gly295, Ala297 and Gln301 (Figure 2g, left and Supplementary Figure 8a). The asymmetric interfaces are formed by hydrophobic interaction of Ile318, Phe316 with Phe289, the insertion of Leu299 into the pocket formed by the main chain of Asn306-Gly309, hydrophobic interaction of Phe289 (from another inner chain) and the main chain Asn303-Gly304, and hydrogen bonds among Met311, Asn312, and Gly290-Ser292 (Figure 2g, right and Supplementary Figure 8a).

The overall R-shaped folds of the inner and outer chain of the SegB A315E fibril are largely similar with some salient differences (Figure 2f), which recall the positional polymorphism observed in a different, shorter TDP-43 segment²³. The head region of the inner chain is formed by the hydrophobic interaction of Leu299, Met307, Met311 and Phe313, as well as by Asn302 and Ser305 hydrogen bonding; whereas in the outer chain, Leu299 is flipped outside the head to mediate the heterodimer interaction between the inner and outer chain (Figure 2f). Consequently, the head region of the outer chain is formed by the hydrophobic interaction of Met307, Met311 and Phe313, together with hydrogen bonds formed by Asn301, Asn302 and Ser305 (Figure 2f, Supplementary Figure 8b). The neck region of both chains is formed by hydrophobic interaction between Ala297 and Phe313, the Arg293-Glu315 salt-bridge. The 70° bend of the Phe316-Asn319 leg of the outer chain relative to the inner chain is probably caused by the heterodimer interface formed by the legs of the inner and outer chain that restricts the conformation of the inner chain.

In all four R-folds we observe a salt-bridge between Arg293 and Glu315 enabled by the pathogenic A315E mutation. The Arg293-Glu315 salt-bridge is not formed by the residues from the same layer along the fibril axis. Rather, each Arg293 interacts with the Glu315 from one or two layers above (Figure 2h), which may affect the kinetics of fibril growth and nucleation. This salt-bridge suggests a mechanistic explanation for the inclination of A315E toward TDP-43 pathology (see Discussion). Model building suggests that wild-type SegB can form the same R-shaped fold as A315E, with Ala315 participating in a hydrophobic interaction with Ala297 and Phe313. This hypothesis is supported by the similar stability

and morphology of wild-type SegB fibrils in our thermostability assays (Supplementary Figure 1c) compared to SegB A315E fibrils and is also supported by the cross-seeding ability of SegB and SegB A315E (Supplementary Figure 1e).

It is noteworthy that although TDP-43 can form either the dagger-shaped polymorph of SegA or the R-shaped polymorph of SegB, it is unlikely that a given molecule of TDP-43 can form both simultaneously. This exclusivity is indicated by a superposition of the two folds in the overlapping segment Asn312-Asn319 (Supplementary Figure 8c). The superposition reveals that steric overlap between the main chains of SegA and SegB would occur if both folds were formed by a single TDP-43 molecule containing both SegA and SegB sequences.

Stability of SegA and SegB fibrils

The extensive hydrophobic interactions and hydrogen bonds observed in both the dagger- and R-shaped folds suggest that both SegA and SegB A315E fibrils are irreversible. To examine this hypothesis, we calculated modified atomic solvation energies (see Online methods) to quantify the stabilities of the fibrils. The calculated stabilities of both the dagger- and R-shaped fold fibrils (represented by energy per layer and per residue) are comparable to other pathogenic amyloid fibrils, such as the Alzheimer's and Pick's disease fibrils (Figure 3 and Table 2). In contrast, the stability of FUS fibrils, thought to be reversible³⁴, is distinctly lower.

We performed thermostability assays to validate our energetic calculations. When heated to 75 °C for 30 minutes, fibrils formed by SegA, SegB, and SegB A315E were all stable, whereas fibrils formed by mCherry-FUS-LCD (composed of residues 1–214 – identical to the sequence determined in the FUS ssNMR structure except with a His-tag replacing mCherry³⁴) disappeared upon heating at 60 °C (Supplementary Figure 1c). These results are consistent with our energetic calculations and support the idea that the dagger- and R-shaped fold fibrils are irreversible.

SegA and SegB folds can be adopted by longer TDP-43 segments

Several lines of evidence suggest that the dagger and R-shaped folds can be adopted by longer TDP-43 segments such as the pathogenic TDP-CTF fragment. (1) Structural conservation: the core region (residue 320–334) of the dagger-shaped fold is structurally conserved in all four dagger-shaped folds despite differences in local environment (Figure 2a). This high degree of structural conservation suggests this core region of dagger-shaped fold is thermodynamically stable, kinetically accessible and is likely to be present in other polymorphs (perhaps including disease relevant TDP-43 aggregates). (2) Model building: the outward facing disposition of the termini in these folds allows longer TDP-43 structures to be modeled from these cores without steric interference. (3) Cross seeding: fibrils formed by TDP-CTF can seed SegA and SegB A315E monomer (Supplementary Figure 1f), and SegA and SegB A315E fibrils can seed TDP-LCD monomer (Figure 4b) (due to its rapid aggregation, TDP-CTF monomer could not be used to measure seeding and we instead used TDP-LCD monomer as the LCD is required for TDP-43 aggregation^{22,25,26}). (4)

Mutagenesis experiments with TDP-CTF reported previously²² and here, as described below.

We found that five individual point mutations to tryptophan delay aggregation of TDP-CTF; these sterically conflict with the tightly packed dagger-shaped (A324W, L330W, Q331W, M337W) and R-shaped folds (Q303W) and delay aggregation of TDP-CTF²². In SegA-sym, three of these mutation sites (Ala324, Leu330 and Met337) point toward the center of the dagger-shaped fold and the other one (Gln331) points toward the “continuous dimer interface” (Figure 2b & Supplementary Figure 9a,b), and in SegB A315E, Gln303 of the inner chain is buried within the inter-chain interface (Supplementary Figure 9a,b and Supplementary Figure 10). All four sites are buried in a tightly packed interface that cannot tolerate a tryptophan replacement. Similarly, the mutations A324E and M337E that are disruptive to the dagger-shaped fold are found to inhibit aggregation of full-length TDP-43 in cells²⁸. As negative controls, we found that A326W and G304W – located in a solvent exposed environment of the dagger-shaped fold and a loose cavity of the R-shaped fold, respectively – can tolerate the bulky tryptophan and did not delay aggregation of TDP-CTF (Supplementary Figure 1d, Supplementary Figure 9a&b)²². Furthermore, a key stabilizing feature of SegB A315E, the Arg293-Glu315 salt bridge, is similarly important for TDP-CTF aggregation. We designed two double mutants, R293E A315E and R293E A315R, based on our structure of SegB A315E. R293E A315E would disrupt the R-shaped fold by electrostatic repulsion whereas R293E A315R would favor the R-shaped fold by restoring the salt-bridge. In accordance with our model, R293E A315E reduces TDP-CTF aggregation whereas R293E A315R enhances TDP-CTF aggregation (Supplementary Figure 1d).

These results suggest that both the dagger- and the R-shaped folds are accessible to TDP-CTF molecules, so that disrupting either one blocks one pathway to fibril formation and hence delays aggregation of TDP-CTF. The observation that none of these mutations targeting the dagger- or R-shaped fold fully eliminates aggregation supports our hypothesis that TDP-CTF can form fibrils using one of multiple aggregation cores, including the dagger- or R-shaped fold, or possibly others not discovered here. In short, conservation of the dagger fold, modeling, seeding experiments, and mutational analysis all suggest that TDP-CTF can form the folds we observe.

Relevance of the dagger- and R- folds for human disease

We investigated whether the dagger- and R-shaped folds are relevant to TDP-43 fibrils in human disease, motivated by the irreversibility of fibrils formed by these folds and the likely accessibility to both folds of TDP-CTF. The scarcity of TDP-43 fibrils in autopsied brains of ALS and FTLN patients daunts structure determination of patient-derived TDP-43 fibrils, so we used seeding experiments to examine the possibility that our structures represent folds adopted by TDP-43 in disease.

We seeded both SegA and SegB A315E monomer with sarkosyl-insoluble material from brain extracts of two patients with FTLN-TDP, which was previously reported to induce FTLN-TDP pathology in a mouse model³⁵. The sarkosyl-insoluble material was able to seed SegA monomer (Supplementary Figure 1g), indicating that in these FTLN-TDP brain extracts, TDP-43 contains a well-folded SegA region that can serve as a template for seeded

aggregation, thereby providing plausibility for the existence of the dagger-shaped fold in these extracts. In contrast we found that the same brain extracts cannot seed SegB A315E monomer (Supplementary Figure 1g), suggesting that the R-shaped fold may not exist in these extracts. This lack of seeding may mean that the R-shaped fold is specific for patients bearing the TDP-43 A315E mutation. Alternatively, the R-shaped fold may not exist in this particular FTLD patient but may exist in other TDP-43 diseases or disease subtypes such as ALS, especially given that the A315E mutation was discovered in the context of ALS³⁶.

We thus propose a one-to-one correspondence between polymorph and disease subtype. More specifically, we speculate that the difference between SegA-sym and SegA-asym are so subtle that they may be associated with the same diseases whereas the differences between the dagger- and R-shaped fold are substantial enough to perhaps be associated with different diseases. This hypothesis is supported by : (1) insoluble TDP-43 extracts from different sub-types of FTLD have distinct morphologies and biochemical properties³⁷, and (2) fibrils of two peptides whose sequences closely match the dagger-shaped fold and the R-shaped fold (Supplementary Figure 9c) seed cell-expressed TDP-43 into distinct aggregates³³.

DISCUSSION

Our finding of polymorphism of TDP-43 fibrils adds TDP-43 to the cohort of other polymorph-forming amyloid proteins^{2,18–20}, and raises the question of why polymorphism is common in amyloid fibrils but not in globular proteins. We suggest that the polymorphism observed in irreversible amyloid fibrils arises from a lack of evolutionary pressure to fold into a particular structure that performs an adaptive function. In contrast, each globular protein and functional aggregate has evolved to fold into a single structure with lower free energy than any other structure that its sequence can form. That is, pathogenic amyloid fibrils lack survival advantage and so can adopt multiple conformations that represent different local energy minima in the protein folding landscape. However, a particular polymorph may give rise to a particular disease³⁸, perhaps nucleated by a mutation such as TDP-43 A315E or a particular cellular environment as suggested for alpha-synuclein³⁹. Polymorph-specific diseases are consistent with results on tau where multiple patients with the same tauopathy all are found to have the same fibril polymorphs and patients with different tauopathies have different fibril polymorphs^{19,20,40}.

Our structure of SegB A315E, to our knowledge the first reported cryo-EM fibril structure containing a hereditary mutation, offers a molecular mechanism in terms of seeding for cell-to-cell spreading of pathology through the brain⁴⁰. As explained above, our SegB A315E structure indicates that A315E facilitates TDP-43 aggregation through electrostatic attraction with Arg293, and suggests A315T (if the threonine is phosphorylated as speculated) can function in a similar way. As summarized in Figure 4a, wild-type TDP-43 monomer can form either the dagger-shaped or R-shaped fold (or possibly other folds), and both folds can act as seeds to recruit additional monomers into the fibril. In A315E, the inter-layer interaction motif of Arg293-Glu315 provides free positive and negative charges on both ends of any R-shaped fibril that consists of 2 layers or more (Figure 2h), and these free charges can attract additional monomers through long-range electrostatic interactions. Thus the

seeding potency of the R-shaped fold with the A315E mutation may exceed that of other folds. Our seeding experiments support this model by showing that SegB A315E fibrils are more effective in seeding TDP-LCD A315E monomer than SegA fibrils as discussed below (Figure 4b).

The observation that both SegA and SegB A315E fibrils can seed LCD monomer (Figure 4b, far-left and middle-left panel) indicates that the LCD can adopt both the dagger- and R-shaped folds when seeded and also suggests roughly equal seeding ability of both dagger- and R-shaped folds. In addition, we found SegA fibrils barely seed LCD A315E monomer (Figure 4b, middle-right panel) whereas SegB A315E fibrils can seed LCD A315E monomer (Figure 4b, far-right panel). We doubt these results arise from the incompatibility of the dagger-shaped fold with the A315E mutation because Ala315 in the dagger-shaped fold faces the solvent so that the A315E mutation should be allowed (Supplementary Figure 9d). Instead, the results shown in Figure 4b middle-right and far-right panels indicate that the seeding ability of the R-shaped fold is much higher than the dagger-shaped fold when seeding A315E monomer. In this case, the observation that SegA fibrils barely seed LCD A315E monomer can be explained by the enhanced primary nucleation and self-seeding of LCD A315E monomer via the Arg293-Glu315 salt-bridge, so that the self-seeded aggregation overwhelms the seeding of SegA fibrils. Taken together, the seeding experiments shown in Figure 4b support our hypothesis that the salt-bridge introduced by A315E mutation accelerates the aggregation of the R-shaped fibrils.

Addition mechanisms proposed for hereditary mutations to impart pathology include the shift of TDP-43 from a reversible to an irreversible assembly^{41,42}. Although we have shown that both dagger- and R-shaped folds form irreversible fibrils, and others have shown that the C-terminal domain can form helical assemblies in liquid droplets⁴², under certain conditions the dagger- and R-shaped folds may also participate in reversible aggregation. Energetic analysis suggests this may be possible for the R-shaped fold. The outer two chains in the SegB A315E structure differs noticeably from the inner two chains, having a lower stabilization energy (0.79 kcal/mol per residue), close to the stabilization energy of the FUS structure (0.66 kcal/mol per residue), which undergoes reversible aggregation, versus the inner chain (0.94 kcal/mol per residue, Table 2). We speculate that the structural and energetic differences between the outer and inner chains are perhaps due to the flexibility of the R-shaped fold owing to its high glycine content and relatively small hydrophobic core. In humans lacking the A315E mutation, the absence of the Arg293-Glu315 salt bridge would eliminate the electrostatic attraction that helps to fix the R-shape, perhaps allowing the molecule to partially unfold to a more open U-shape. This conformational change may further compromise the stability of the R-shaped fold by promoting the disassembly of the four-protofilament bundle given the observation that the symmetric interface of the R-shaped fold is relatively weak (Figure 2g) and formation of the asymmetric interface requires the closed conformation of the inner R-shaped folds. Therefore, the transition from reversible to irreversible assemblies may occur through a U-R switch where a more open and less bundled U-shape fold converts to a more compact and more bundled R-shape. This U-R switch is possible for both wild-type and A315E, although perhaps the wild-type sequence would have a higher transition state energy owing to the loss of the Arg293-Glu315

interaction. That is, we speculate that the A315E mutation could impose a switch from a reversible to an irreversible assembly.

We note that our energy analysis relies on the structure of one reversible fibril and that we have no experimental evidence regarding the role of the R-shaped polymorph in reversible TDP-43 aggregation; therefore, further research identifying structures of reversible fibrils and interrogating the role of the R-shape in TDP-43 reversible aggregation is needed.

Our hypotheses proposed in the preceding four paragraphs are only two of numerous possibilities for mutation-enhanced toxicity in TDP-43 proteinopathies and may not apply to hereditary mutations other than A315E. We observe that many hereditary mutations would have a neutral, or even deleterious, effect on the dagger- and R-shaped folds (Supplementary Figure 9d). This is consistent with the idea that there are other observed mechanisms for mutation-enhanced pathology, such as toxic gain-of-function RNA splicing activity⁴³. In addition, these mutations may influence the formation of other amyloid cores not discovered here.

Our four near-atomic resolution structures of TDP-43 amyloid fibrils establish that: (1) TDP-43 is capable of forming multiple fibrillar structural polymorphs; (2) That two sequence segments of TDP-43 can form distinct stable amyloid cores, one with dagger-shaped folds and the other with R-shaped folds; (3) The R293-E315 salt-bridge in R-shaped fold provides a plausible explanation of enhanced TDP-43 pathology by the ALS related hereditary mutation A315E (and possibly A315Tp); and (4) Energetic analysis highlights the structural features of amyloid fibrils that may lead to both reversible and irreversible aggregation.

ONLINE METHODS

Construct design

Construct design of TDP-43 segments follows the same strategy reported previously²². In addition to TDP-LCD (274–414) reported in our previous study²², constructs used in this study include TDP-43 full length (1–414), SegA (311–360), SegB (286–331, wild type or with A315E mutation) and SegAB (286–360). All TDP-43 fragments were conjugated by inserting the cDNA into a pET28a vector that contains an amino terminus conjugation of SUMO protein. The SUMO protein was used to increase solubility and prevent aggregation during expression and purification. The (His)₆-tag on the amino terminus of the SUMO protein was used for Ni-column purification. The expression sequence of each TDP-43 fragments is shown in Supplementary Table 1.

Protein purification and SUMO-tag cleavage

Protein purification follows the same protocol reported previously²². Briefly, all SUMO-tagged proteins were expressed in *Escherichia coli* BL21 (DE3) strain. LB media with 50 µg/ml kanamycin was used for cell culture. After bacterial cells were cultured at 37 °C to an OD₆₀₀ of 0.6–0.8, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for protein expression and the cells were further cultured at 25 °C for 3 h. Cells were harvested and resuspended 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), and were sonicated (3s on/3s of cycle, 10 min) and centrifuged (24,000 g for 20 min) to get cell lysate. The cell lysate was mixed with homemade NucA nuclease (5000 U per liter of cell culture) and filtered before loading onto a HisTrap HP column (GE healthcare) for purification. The HisTrap column was pre-equilibrated with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 20 mM imidazole, and sample was washed with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 200 mM imidazole and eluted with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 500 mM imidazole. Eluted protein was concentrated using Amicon Ultra-15 centrifugal filters (Millipore) and stored at -80°C for future use.

In order to form fibrils of the TDP-43 fragments, the SUMO-tag of each fragment was removed before aggregation. SUMO-tagged protein was mixed with 100:1 (weight basis) homemade ULP1 protease, and incubated for 1 hour. When analyzed by SDS-PAGE, samples mixed with ULP1 at 0 hour show bands of intact SUMO-tagged protein; after 1 hour of cleavage, bands of free SUMO and TDP-43 fragments were shown on the gel, as exemplified by SegA and SegB A315E cleavage shown in Supplementary Figure 1a.

Fibril preparation and optimization

SUMO-tag removed proteins were used for fibril preparation. For all TDP-43 fragments, the starting fibril preparation conditions are the same. Protein was diluted to 50 μM concentration with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μM DTT and filtered using 0.1 μm Ultrafree-MC-VV centrifugal filters (Millipore), and incubated at 37°C with shaking for 3 days. The morphology of each sample was checked by negative stain TEM.

In order to achieve the necessary fibril concentration and distribution for cryo-EM structure determination, the fibril growth condition was optimized for each sample. The optimization includes determining the appropriate concentration of proteins (final values ranging from 3 μM to 100 μM monomer concentration), incubating the solution without shaking, incubation temperature (4°C or 25°C or 37°C), incubation time (overnight to more than 2 weeks), buffer (Tris or phosphate), salt concentration (50 mM to 500 mM NaCl) and seeding with pre-formed fibrils. After optimization of all TDP-43 fragments, SegA and SegB A315E were found to have the best fibril morphology, distribution, and concentration and were therefore selected for cryo-EM structure determination. The optimized fibril morphology of each fragment is shown in Supplementary Figure 1b.

The optimized fibril growth condition for SegA was 50 μM protein concentration, incubation in 20 mM Tris-HCl, pH 7.4, 500 mM NaCl, 10 μM DTT at 37°C with 3 days of shaking. Two percent (molar ratio, monomer equivalent) pre-formed SegA fibrils were added as seeds. After shaking, SegA fibrils were collected by centrifugation at 9,000 g for 5 min and washed twice and resuspended to the starting volume with the identical buffer as above, except using 150 mM NaCl instead of 500 mM. Fibrils in resuspension buffer were further incubated without shaking for 3 days at room temperature. Before cryo-EM sample preparation, SegA fibrils were again collected by centrifugation, and re-suspended with 150 mM NaCl buffer at 5% volume of previous incubation ($20\times$ concentrating), and the concentrated solution was heated to 75°C for 30 min with a PCR machine. The final heating step was necessary to de-clump SegA fibrils and optimize fibril distribution.

The optimized fibril growth condition for SegB A315E was 50 μM initial protein concentration, and incubation in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μM DTT, at 37 $^{\circ}\text{C}$ without shaking. Two percent (molar ratio, monomer equivalent) pre-formed SegB A315E fibrils were added as seeds. After incubation, SegB A315E fibrils were collected by centrifugation at 9,000 g for 5 min and washed twice by the same incubation buffer at 5% volume of previous incubation ($20 \times$ concentrating), and the concentrated solution was used for cryo-EM sample preparation.

Negative stain transmission electron microscopy (TEM)

Negative stain TEM samples were prepared by applying 5 μl of fibril solution to glow-discharged 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.) and incubating on the grid for 2 min. The samples were then blotted off and the grids were stained with 3 μl of 2% uranyl acetate for 1 min. The grids were washed with an additional 3 μl of 2% uranyl acetate and allowed to dry for 1 min. Each grid was imaged using a T12 (FEI) electron microscope.

Cryo-EM data collection and processing

2 μl of fibril solution was applied to a glow-discharged Quantifoil 1.2/1.3 electron microscope grid and plunge-frozen into liquid ethane using a Vitrobot Mark IV (FEI). Data were collected on a Titan Krios (FEI) microscope equipped with a Gatan Quantum LS/K2 Summit direct electron detection camera (operated with 300 kV acceleration voltage and slit width of 20 eV). Super-resolution movies were collected on a Gatan K2 Summit direct electron detector with a nominal physical pixel size of 1.07 \AA /pixel (0.535 \AA /pixel in super-resolution movie frames) with a dose per frame of $\sim 1.1 \text{ e}^{-}/\text{\AA}^2$. A total of 40 frames with a frame rate of 5 Hz were taken for each movie resulting in a final dose of $\sim 44 \text{ e}^{-}/\text{\AA}^2$ per image. Automated data collection was driven by the Legimon automation software package⁴⁴.

Micrographs containing crystalline ice were used to estimate the anisotropic magnification distortion using `mag_distortion_estimate`⁴⁵. CTF estimation was performed using CTFFIND 4.1.8 on movie stacks with a grouping of 3 frames and correction for anisotropic magnification distortion⁴⁶. Unblur⁴⁷ was used to correct beam-induced motion with dose weighting and anisotropic magnification correction, resulting in a physical pixel size of 1.06 \AA /pixel and 1.064 \AA /pixel for the SegB and SegA data sets, respectively.

All particle picking was performed manually with SegB A315E data set fibrils picked as a group and SegA data set fibrils were picked separately in groups of slow twister and fast twisters using EMAN2 `e2helixboxer.py`⁴⁸ (we could not discriminate between SegA-sym and SegA-asym by eye due to their similar crossover distances and morphology; therefore, SegA-sym and SegA-asym particles were later separated by 2D classification and 3D classification). Particles were extracted in RELION using the 90% overlap scheme into 288 pixel boxes for SegA and 320 pixel boxes for SegB A315E. Classification, helical reconstruction, and 3D refinement were used in RELION as described⁴⁹⁻⁵¹. In general, we performed 2D Classifications with higher `tau_fudge` (ranging from values of 4 to 12) to identify, and select for future 3D classification, particles contributing to 2D classes showing

β -strand separation along the helical axis. We performed 3D classification with the estimated helical parameters from each structure and an elongated Gaussian blob as an initial model to generate starting reconstructions. We ran additional 3D classifications using the preliminary reconstructions from the previous step to select for particles contributing to homogenous classes (stable helicity and separation of β -strands in the X-Y plane). Typically, we performed Class3D jobs with $K=3$ and manual control of the tau_fudge factor and healpix to reach a resolution of $\sim 5\text{--}6$ Å to select for particles that contributed to the highest resolution class for each structure. For the SegB and SegA-slow reconstructions, we performed high-resolution gold-standard refinement as described⁵¹. For SegA-sym and SegA-asym, perhaps due to low particle numbers, Refine3D seemed unable to adequately regulate working resolution to refine with high enough resolution factors to allow SegA-sym and SegA-asym reconstructions to refine past 4.8 Å, and gold-standard refinement did not result in near-atomic resolution maps. Therefore, we employed a “sub-optimal” manual refinement protocol using all particles against a single reference map as described^{2,49}. Final overall resolution estimates were calculated from the corrected Fourier shell correlations of two half maps, using high-resolution phase randomization as implemented in RELION PostProcess to account for effects of a soft-edged solvent mask and to monitor possible overfitting for the SegA-sym and SegA-asym manual refinements⁵². For the SegB and SegA-slow gold-standard refinements, we report the 0.143 FSC resolution cutoff; however, for the SegA-sym and SegA-asym “sub-optimal” refinements we report the 0.5 FSC resolution cutoff (the particles contributing to the final map were split *a posteriori* into even and odd micrographs to prevent particles from the same fibril contributing to different reconstructions and two half-maps were reconstructed using the fully refined center and orientation parameters)⁵². In all cases except the SegA-slow, the default value of 30% was used for the *helical_z_percentage* during 3D classification and refinement. Since the crossover distance of the SegA-slow varied from 1200–2000 Å, we used 10% for the *helical_z_percentage* to minimize the blurring due to the variable twist as described¹⁹. Reconstructions were compared against the reference-free 2D class averages to confirm their validity. In addition, *de novo* model-building also confirms the validity of the 3D reconstructions.

Specific determination of helical symmetry for all structures was performed as follows. For SegA-sym, SegA-asym, and SegB A315E, we estimated the crossover distances using a combination of manual inspection of the micrographs and inspection of 2D class averages using larger particle box sizes to capture the entire crossover distance (box sizes from 432 to 1024 pixels were tested). In the case of the SegA-sym and the SegB A315E, the mirror symmetry axis parallel to the helical axis in the 2D class averages indicated either a two-start helix symmetry with a rise of 2.4 Å or a C_n rotational symmetry. Computed diffraction patterns from the 2D class averages indicated there was no meridional reflection at 4.8 Å, indicating the lack of C_n rotational symmetry. Therefore, for both of these species, a helical rise of 2.4 Å was assumed and the helical twist was calculated from the previously estimated crossover distance. Convergence to a stable helicity and high-resolution refinement to near-atomic resolution allowing *de novo* model building confirmed the helical symmetry. For SegA-slow, the entire crossover distance could not be captured by the 1024 box size, so the crossover distance was estimated from a combination of manual inspection of micrographs

and a “stitched” together fibril using the 2D classes from the 288 box size (Supplementary Figure 3f). Class averages, as well as the computed diffraction pattern from the “stitched” fibril, confirm the presence of a meridional reflection at 4.8 Å, indicating C_n rotational symmetry. Therefore, we assumed a helical rise of 4.8 Å and we calculated the helical twist from the previously estimated crossover distance. High-resolution refinement to near-atomic resolution allowed *de novo* model building, confirming the helical symmetry. In addition, reconstructions with a helical rise of 2.4 Å did not show clear side chain density, and projections of the reconstruction did not match the 2D class averages (Supplementary Figure 3e). For SegA-*asym*, the crossover distance was estimated from a combination of inspection of the micrographs, and 2D class averages with larger box sizes. Due to the lack of apparent two-fold symmetry in the 2D class averages, we assumed a helical rise of 4.8 Å and calculated the helical twist from the measured cross-over distances. Refinement to near-atomic resolution demonstrate that the two SegA-*asym* monomers within each asymmetric unit are not identical, confirming the lack of two-fold symmetry.

Atomic model building

We sharpened the refined maps using phenix.auto_sharpen⁵³ at the resolution cutoff indicated by half-map FSC, and subsequently built atomic models into the refined maps with COOT⁵⁴. For the three SegA polymorphs, we built the model of SegA-*slow* first owing to its high resolution (3.3 Å). The density map of SegA-*slow* clearly shows a cyclic mainchain with the amino and carboxy termini at the center of the fibril. The initial model was built with both N-C orientations, and only the orientation shown in the current model fit all the side chain densities. The conservation of the dagger-fold in the maps of SegA-*sym* and SegA-*asym*, especially the signature ³³⁴WGMMGML³⁴⁰ motif, allowed us to rigid-body fit the model of SegA-*slow* into SegA-*sym* and SegA-*asym* maps. Following rigid-body fitting, we manually inspected and adjusted each residue.

After building atomic models for the main chain of SegA polymorphs, we observed extra density peripheral to the dagger-shaped fold around Gln331 of SegA-*slow* and Met323 of both SegA-*asym* and SegA-*slow* (Figure 1c–d). We believe this density arises from unstable interactions (either due to weak binding or binding of a mixture of different residues) of other protomers with the core protofilaments. We could not build a definitive model due to ambiguity of the side-chain densities; however, we built a hypothetical model into the density near Gln331 of SegA-*slow* based on the speculation that this density represents part of the continuous dimer interface found in SegA-*sym* and SegA-*asym*.

During model building of SegA-*slow*, we found that the model that fully recapitulates the registration of SegA-*sym* was not favorable due to the steric clash at L340 and unexplained side chain density at G338 (Supplementary Figure 6c). Therefore, we built the model with a registration that has a 2 residue offset compared to SegA-*sym*. We believe that this registration shift is allowed by the similarity of nearby residues (Gln331 from dagger-fold flanked by Ser342 and Gln344 from the short chain, compared to Gln331 flanked by Gln344 and Gln346), and is caused by the lack of the hydrophobic anchor formed by Met336 and Leu340 (Supplementary Figure 6d). A similar registration shift was observed in previous reported crystal structures of steric-zippers and was termed registration polymorphism⁵⁵.

The extra densities near Met323 in SegA-asym and SegA-slow could represent an additional dimer interface that protects Ala321, Met323 and Ala325 from the solvent, but without unambiguous side chain density, we could not build in any models for these densities.

The model building of SegB A315E was initiated by the inspection of the density map near the homo-dimer interface on the neck of the inner-chain R-shaped fold. The side chain density and nearby environment indicates the atomic model of this region should feature 4–5 small side chain residues flanked by a big side chain residue at each side, and the only motif in the SegB A315E sequence that matches this criterion is ²⁹³RGGGAGL²⁹⁹, with only one orientation feasible considering the length of density at each side of this motif. With the model of ²⁹³RGGGAGL²⁹⁹ built, the rest of model was built by adding the residues one-by-one. All sidechains fit the density unambiguously. The observation that Glu315 forms a salt-bridge with Arg293 further supports this model. After we built the inner R-shape fold, we speculated that the outer R-shape fold was not a continuation of the same chain as the inner R-shape fold because there were not enough residues in the sequence to accommodate the outer R-shape fold. Therefore, the model of the outer R-shape fold was built *de novo* similarly to the inner-R-shape fold.

For all structures, after we built the initial model, we used in-house scripts to generate a 5-layer model using the helical symmetry of the refined map, and used phenix.real_space_refine for model refinement⁵⁶. After several rounds of refinement, we adjusted the orientation of the main chain oxygen and nitrogen to facilitate main chain hydrogen bond within the β -sheet, and we applied these mainchain hydrogen bond restraint for further real-space refinement. As the last step, the rotamer of each serine, glutamine and asparagine residue was manually inspected to ensure the energy favorable hydrogen binding. The final model was validated by MolProbity⁵⁷.

The G304W model of SegB A315E shown in Supplementary Figure 9b was generated by manually changing Gly304 residues of 1-layer SegB A315E model (four glycine residues in four chains) to tryptophan in COOT, and we generated a 5-layer mutated model using the same script as for original SegB A315E model. We refined this 5-layer mutated model by phenix.real_space_refine using the same density map we used to refine the original model. The refined model shows no Ramachandran angles and rotamer outliers, and no significant clashes. The statistics of SegB A315E G304W model are shown in Supplementary Table 3 and the coordinates are listed as Supplementary Data Set 1.

Energetic calculation

The stabilization energy is an adaptation of the solvation free energy described previously⁵⁸, in which the energy is calculated as the sum of products of the area buried of each atom and its corresponding atomic solvation parameter (ASP). ASPs were taken from our previous work⁵⁸. Area buried is calculated as the difference in solvent accessible surface area (SASA) of the reference state (i.e. unfolded state) and the SASA of the folded state. The reference state was measured absent all other atoms in the structure but the residue “i” and main chain atoms of residue i-1 and i+1. The SASA of the folded state was measured for each atom in the context of all amyloid fibril atoms. Fibril coordinates were extended by symmetry by three to five chains on either side of the reported molecule, to ensure the energetic

calculations were representative of the majority of molecules in a fibril, rather than a fibril end. To account for energetic stabilization of main chain hydrogen bonds, the ASP for backbone N/O elements was reassigned from -9 to 0 if they participated in a hydrogen bond. Similarly, if an asparagine or glutamine side chain participated in a polar ladder (two hydrogen bonds per amide), and was shielded from solvent ($SASA_{\text{folded}} < 5 \text{ \AA}^2$), the ASPs of the side chain N and O elements were reassigned from -9 to 0 . Lastly, the ASP of ionizable atoms (e.g. Asp, Glu, Lys, His, Arg, N-terminal amine, or C-terminal carboxylate) were assigned the charged value ($-37/-38$) unless the atoms participated in a buried ion pair, defined as a pair of complementary ionizable atoms within 4.2 \AA distance of each other, each with $SASA_{\text{folded}} < 40 \text{ \AA}^2$). In that case, the ASP of the ion pair was reassigned to -9 . In the energy diagrams, a single color is assigned to each residue, rather than each atom. The color corresponds to the sum of solvation free energy values of each of the atoms in the residue. The energy reported for FUS in Table 2 is the average over 20 NMR models. The standard deviation is 1.8 kcal/mol .

Thermo-stability assays of TDP-43 and FUS fibrils

To validate the energetic calculations and test the stability of TDP-43 segment fibrils experimentally, pre-formed TDP-43 segment fibrils were heated at $75 \text{ }^\circ\text{C}$ for 30 min in a PCR machine, and the existence of the fibrils were checked by negative stain EM. All fibrils were heated in 20 mM Tris pH 7.4, 150 mM NaCl, $10 \text{ }\mu\text{M}$ DTT at $50 \text{ }\mu\text{M}$ monomer-equivalent concentration.

FUS protein was purified and FUS fibrils was formed the same way as previously reported [ref]. FUS protein was expressed using pHis-parallel-mCherry-1FUS214 plasmid that contain the low complexity domain of FUS (1–214) conjugated to mCherry. Purification was done by Ni column and size exclusion column. The hydrogel of FUS was formed by incubating purified mCherry-FUS at $4 \text{ }^\circ\text{C}$ for 2 weeks at a concentration of 1 M . After hydrogel formation, a scrape of hydrogel was mixed with 20 mM Tris pH 7.4, 150 mM NaCl, $10 \text{ }\mu\text{M}$ DTT to approximately 20x of its original volume ($\sim 50 \text{ }\mu\text{M}$ monomer concentration). The mixture was allocated to three PCR tubes, one without heat and two heated at $60 \text{ }^\circ\text{C}$ and $75 \text{ }^\circ\text{C}$ for 30 min, respectively. All three samples were checked by negative stain EM for fibrils. The sample without heat contains abundant fibrils on the EM grid and the sample with heat contains no visible fibrils anywhere on the EM grid.

ThT assays

Protein was diluted into $12 \text{ }\mu\text{M}$ in 20 mM Tris-HCl, pH 7.4, 500 mM NaCl, $10 \text{ }\mu\text{M}$ DTT, $30 \text{ }\mu\text{M}$ ThT, and filtered using $0.1 \text{ }\mu\text{m}$ Ultrafree-MC-VV centrifugal filters (Millipore). Filtered solution was mixed with or without 100:1 (weight basis) ULP1, with or without 5% (molar ratio, monomer equivalent) pre-formed fibril seeds (or 1% v/v for brain extract seeds), and was pipetted into a polybase black 384-well plate with optical bottom (Thermo Scientific) and incubated at $37 \text{ }^\circ\text{C}$ without shaking. ThT fluorescence was measured with excitation and emission wavelength of 440 and 480 nm, respectively, using FLUOstar Omega plate reader (BMG LABTECH). The aggregation curves were averaged from four independent measured replicates and error bars show s.d. of replicate measurements. Fibril seeds were sonicated (1s on/1s off cycle, 3 min) before use to increase the seeding ability.

The brain extracted materials were a generous gift from Prof. Virginia Lee's lab. Brain extract #1 was from a patient diagnosed as FTLN-TDP, whereas brain extract #2 was from a patient diagnosed as FTLN-TDP-GRN. Both brain extracts were extracted from the frontal cortex, and estimated TDP-43 concentration were ~500 ng/ml (measured by ELISA assays) and estimated total protein concentration were 2.5 mg/ml for brain extract #1 and 4 mg/ml for brain extract #2 (measured by BCA assays).

The samples mixed with pre-formed fibril seeds had an additional ThT background reading because of the seeds. To better compare the curves of samples with and without seeding and estimate the ThT signal change that represented the formation of seeded fibrils, we removed the background by subtracting the value of the first reading from each curve. To normalize the different ranges of fluorescence readings observed from each experiments (probably due to the different fluorescence gain settings of the plate reader), we normalized the readings to make the minimum mean value in each panel 0%, and the maximum mean value in each panel 100%.

TDP-CTF aggregation assays

The aggregation assays of pathological fragments of TDP-43 (TDP-CTF) were performed largely the same as previous reported²². The SUMO-TDP-CTF wild type and mutants were diluted into 0.2 mg/ml with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μ M DTT and filtered by 0.1 μ m Ultrafree-MC-VV centrifugal filters (Millipore). To simplify the experiment, we did not remove SUMO-tag by ULP1 protease because we found that wild type TDP-CTF aggregates after overnight incubation even with the protection of SUMO. Filtered samples were incubated at 4 °C overnight, and separated into supernatant and pellet fractions by centrifugation at 18,000 g for 3min at 4 °C. The pellet was resuspended with the same buffer and volume, and both supernatant and pellet were mixed with 3:1 (v/v) of NuPAGE LDS sample buffer (Invitrogen) and heated to 100 °C for 10 min. The samples before incubation were also prepared in the same way. All samples were separated by NuPAGE 4–12% Bis-Tris gels (Invitrogen) and stained by Coomassie blue. TDP-CTF A324W that had been shown to reduce TDP-CTF aggregation previously²² was also tested in this study as a control to make sure that we can get the similar results with or without ULP1 cleavage.

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank H. Zhou for use of Electron Imaging Center for Nanomachines (EICN) resources. We acknowledge the use of instruments at the EICN supported by UCLA and by instrumentation grants from NIH (1S10RR23057 and 1U24GM116792) and NSF (DBI-1338135 and DMR-1548924). The authors acknowledge NIH AG 054022, NIH AG061847, and DOE DE-FC02-02ER63421 for support. D.R.B. was supported by the National Science Foundation Graduate Research Fellowship Program.

REFERENCES

1. Anfinsen CB Principles that govern the folding of protein chains. *Science* 181, 223–230 (1973). [PubMed: 4124164]
2. Li B. et al. Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* 9, 3609 (2018). [PubMed: 30190461]
3. Zhang W. et al. Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's disease. (2018). doi:10.1101/468892
4. Neumann M. et al. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science* 314, 130–133 (2006). [PubMed: 17023659]
5. Hasegawa M. et al. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann. Neurol.* 64, 60–70 (2008). [PubMed: 18546284]
6. Irwin DJ et al. Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine. *Acta Neuropathol. (Berl.)* 129, 469–491 (2015). [PubMed: 25549971]
7. Amador-Ortiz C. et al. TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann. Neurol.* 61, 435–445 (2007). [PubMed: 17469117]
8. Nakashima-Yasuda H. et al. Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathol. (Berl.)* 114, 221–229 (2007). [PubMed: 17653732]
9. 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.* 67, 1159–1165 (2008). [PubMed: 19018245]
10. Salajegheh M. et al. Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis. *Muscle Nerve* 40, 19–31 (2009). [PubMed: 19533646]
11. Nelson PT et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain J. Neurol.* (2019). doi:10.1093/brain/awz099
12. Buratti E. & Baralle FE The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biol.* 7, 420–429 (2010). [PubMed: 20639693]
13. Lee EB, Lee VM-Y & Trojanowski JQ Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci.* 13, 38–50 (2011). [PubMed: 22127299]
14. Iguchi Y. et al. TDP-43 depletion induces neuronal cell damage through dysregulation of Rho family GTPases. *J. Biol. Chem.* 284, 22059–22066 (2009).
15. Kraemer BC et al. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol. (Berl.)* 119, 409–419 (2010). [PubMed: 20198480]
16. Wang J. et al. Cell-autonomous requirement of TDP-43, an ALS/FTD signature protein, for oligodendrocyte survival and myelination. *Proc. Natl. Acad. Sci.* 115, E10941–E10950 (2018).
17. Tycko R. Molecular Structure of Aggregated Amyloid- β : Insights from Solid-State Nuclear Magnetic Resonance. *Cold Spring Harb. Perspect. Med.* 6, (2016).
18. Gremer L. et al. Fibril structure of amyloid- β (1–42) by cryo-electron microscopy. *Science* 358, 116–119 (2017). [PubMed: 28882996]
19. Fitzpatrick AWP et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* 547, 185–190 (2017). [PubMed: 28678775]
20. Falcon B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* 561, 137–140 (2018). [PubMed: 30158706]
21. Iadanza MG et al. The structure of a β 2 -microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat. Commun.* 9, 4517 (2018). [PubMed: 30375379]
22. Guenther EL et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* 25, 463–471 (2018). [PubMed: 29786080]
23. Guenther EL et al. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol.* 25, 311–319 (2018). [PubMed: 29531287]
24. Igaz LM et al. Expression of TDP-43 C-terminal Fragments in Vitro Recapitulates Pathological Features of TDP-43 Proteinopathies. *J. Biol. Chem.* 284, 8516–8524 (2009). [PubMed: 19164285]

25. Pesiridis GS, Lee VM-Y & Trojanowski JQ Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 18, R156–162 (2009). [PubMed: 19808791]
26. Kato M. & McKnight SL A Solid-State Conceptualization of Information Transfer from Gene to Message to Protein. *Annu. Rev. Biochem.* 87, 351–390 (2018). [PubMed: 29195049]
27. Hughes MP et al. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* 359, 698–701 (2018). [PubMed: 29439243]
28. Jiang L-L et al. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* 288, 19614–19624 (2013). [PubMed: 23689371]
29. Guo W. et al. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol.* 18, 822–830 (2011). [PubMed: 21666678]
30. 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.* 14, e1002338 (2016).
31. Wegorzewska I, Bell S, Cairns NJ, Miller TM & Baloh RH TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. *Proc. Natl. Acad. Sci.* 106, 18809–18814 (2009).
32. Xu M. et al. Characterization of β -domains in C-terminal fragments of TDP-43 by scanning tunneling microscopy. *J. Struct. Biol.* 181, 11–16 (2013). [PubMed: 23138004]
33. Shimonaka S, Nonaka T, Suzuki G, Hisanaga S. & Hasegawa M. Templated Aggregation of TAR DNA-binding Protein of 43 kDa (TDP-43) by Seeding with TDP-43 Peptide Fibrils. *J. Biol. Chem.* 291, 8896–8907 (2016). [PubMed: 26887947]
34. Murray DT et al. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* 171, 615–627.e16 (2017).
35. Porta S. et al. Patient-derived frontotemporal lobar degeneration brain extracts induce formation and spreading of TDP-43 pathology in vivo. *Nat. Commun.* 9, 4220 (2018). [PubMed: 30310141]
36. Fujita Y, Ikeda M, Yanagisawa T, Senoo Y. & Okamoto K. Different clinical and neuropathologic phenotypes of familial ALS with A315E TARDBP mutation. *Neurology* 77, 1427–1431 (2011). [PubMed: 21956716]
37. Laferrière F. et al. TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci.* 22, 65 (2019). [PubMed: 30559480]
38. Clavaguera F. et al. Brain homogenates from human tauopathies induce tau inclusions in mouse brain. *Proc. Natl. Acad. Sci.* 110, 9535–9540 (2013). [PubMed: 23690619]
39. Peng C. et al. Cellular milieu imparts distinct pathological α -synuclein strains in α -synucleinopathies. *Nature* 557, 558 (2018). [PubMed: 29743672]
40. Falcon B. et al. Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol. (Berl.)* (2018). doi:10.1007/s00401-018-1914-z
41. Ramaswami M, Taylor JP & Parker R. Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell* 154, 727–736 (2013). [PubMed: 23953108]
42. Conicella AE, Zerze GH, Mittal J. & Fawzi NL ALS mutations disrupt phase separation mediated by -helical structure in the TDP-43 low complexity C-terminal domain. *Structure* 24, 1537–1549 (2016). [PubMed: 27545621]
43. Fratta P. et al. Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. *EMBO J.* 37, (2018).

METHODS-ONLY REFERENCES

44. Suloway C. et al. Automated molecular microscopy: the new Legation system. *J. Struct. Biol.* 151, 41–60 (2005). [PubMed: 15890530]
45. Grant T. & Grigorieff N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* 192, 204–208 (2015). [PubMed: 26278979]

46. Rohou A. & Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221 (2015). [PubMed: 26278980]
47. Grant T. & Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* 4, e06980 (2015).
48. Tang G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* 157, 38–46 (2007). [PubMed: 16859925]
49. Guenther EL et al. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol.* 25, 311–319 (2018). [PubMed: 29531287]
50. He S. & Scheres SHW Helical reconstruction in RELION. *J. Struct. Biol.* 198, 163–176 (2017). [PubMed: 28193500]
51. Scheres SHW RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530 (2012). [PubMed: 23000701]
52. Chen S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135, 24–35 (2013). [PubMed: 23872039]
53. Terwilliger TC, Sobolev OV, Afonine PV & Adams PD Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. Struct. Biol.* 74, 545–559 (2018).
54. Emsley P, Lohkamp B, Scott WG & Cowtan K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501 (2010). [PubMed: 20383002]
55. Wiltzius JJW et al. Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* 16, 973–978 (2009). [PubMed: 19684598]
56. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* 74, 531–544 (2018).
57. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21 (2010). [PubMed: 20057044]
58. D E, M, W. & M Y. Interpretation of protein folding and binding with atomic solvation parameters. *Chemica Scripta* (1989). Available at: <https://eurekamag.com/research/032/001/032001963.php>. (Accessed: 20th December 2018)

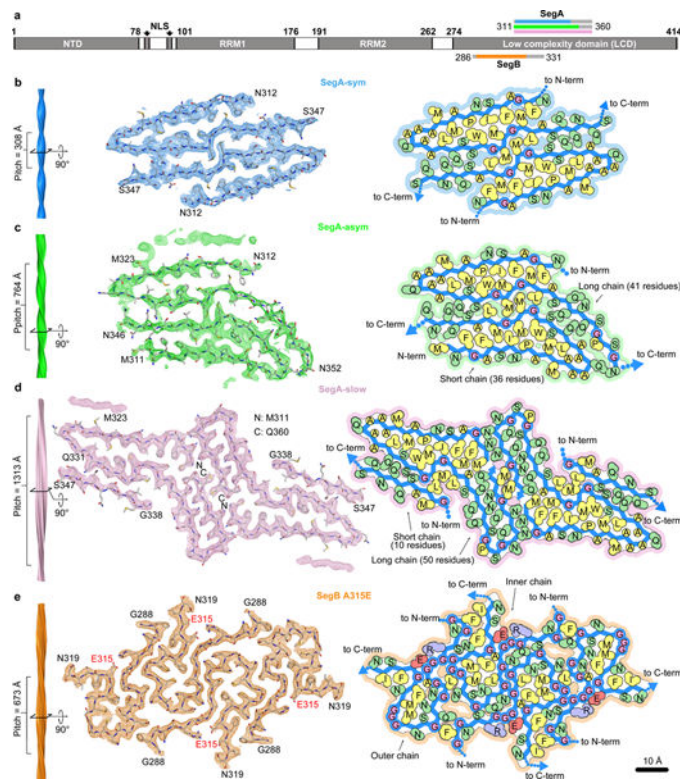


Figure 1. Cryo-EM Structure of TDP-43 polymorphic fibrils.

a, Schematic of full-length TDP-43. SegA (residues 311–360) and SegB (residues 286–331) identified for structural determination are shown as gray bars, respectively above and below the low-complexity domain (LCD). The color bars show the range of residues visualized in the structure of each polymorph. Panels **b–e**: (left) fibril reconstructions showing left-handed twist and pitch; (middle) density and atomic model of one cross-sectional layer of each fibril; (right) schematic model showing protein chain (blue) and residues (hydrophobic in yellow, polar in green, glycine in pink, glutamate in red, and arginine in blue).

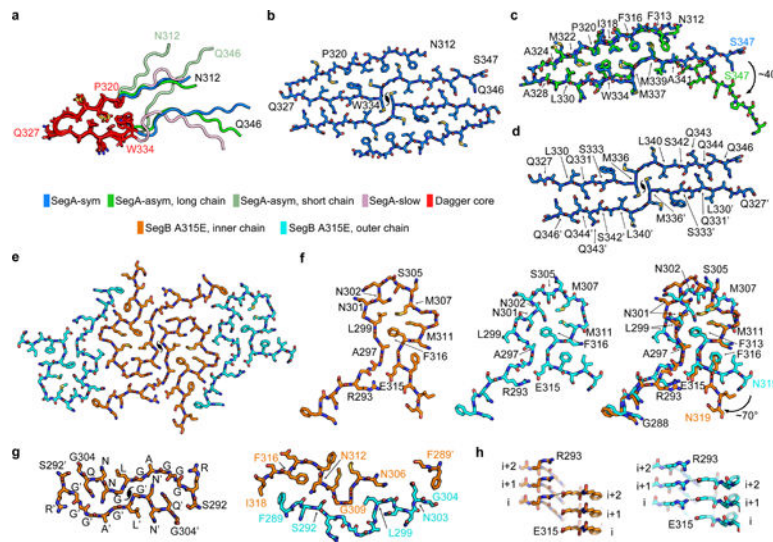


Figure 2. Structure of the dagger-shaped fold and the R-shaped fold.

a, Superposition of the dagger-shaped folds (residue 312–346) in three SegA polymorphic fibrils. The core region of the dagger-shaped folds (residue 320–334) is colored red with side chains shown, and the rest of the dagger-shaped fold is colored according to the key. **b**, Atomic model of SegA-sym. **c**, Pairwise superposition of the dagger-shaped fold of SegA-sym vs. SegA-asym long chain. **d**, Continuous dimer interface of SegA-sym. **e**, Atomic model of SegB A315E. Each of the four chains is R-shaped, with a pseudo- 2_1 axis relating the two pairs. The inner two chains are identical as are the two outer chains. **f**, Structures of the inner (left) and outer (middle) chains and their superposition (right). Notice the proximity of R293 to the pathogenic variant E315 which replaces A315 in the reference sequence. **g**, View parallel to the fibril axis showing the symmetric interface between the two inner chains (left) and the asymmetric interface between the inner and outer chains (right). **h**, View perpendicular to the fibril showing the inter-layer interaction of Arg293 with Glu315 on the inner chain (left) and outer chain (right). (Detailed alignment parameters and RMSD values listed in Supplementary Table 2).

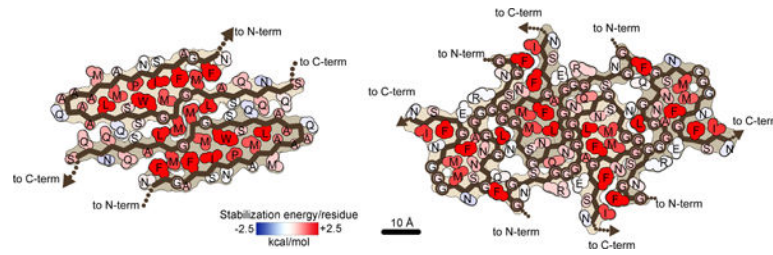


Figure 3. Calculated solvation energy maps for TDP-43 fibrils based on atomic solvation parameters.

Solvation energy maps of SegA-sym (left) and SegB A315E (right). Residues are colored from unfavorable (blue, -2.5 kcal/mol) to favorable stabilization energy (red, 2.5 kcal/mol). Notice that both the dagger-shaped fold (SegA-sym) and the R-shaped fold (SegB A315E) show stable cores.

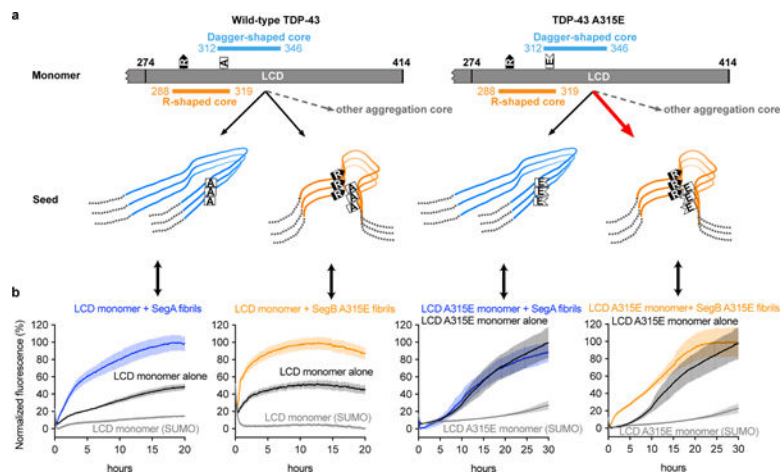


Figure 4. Speculative model of TDP-43 amyloid aggregation based on fibril structures reported here.

a, (Left) Monomeric TDP-43 can form amyloid fibrils with either dagger-shaped or R-shaped cores (or perhaps other so far unidentified cores). The ALS hereditary A315E mutation accelerates the R-shaped core through electrostatic interaction of Glu315 with Arg293. **b**, Experiments seeding TDP-43 fibril formation support the model of Panel A. Notice that both SegA and SegB A315E fibrils seed LCD monomer (far left and middle left). In contrast, only SegB A315E fibrils seed LCD A315E monomer whereas SegA fibrils barely seed LCD A315E monomer (far right and middle right). Data are shown as mean \pm s.d., n=4 independent experiments.

Table 1

Cryo-EM data collection, refinement and validation statistics

	SegA-sym (EMD-9339, PDB 6N37)	SegA-asym (EMD-9350, PDB 6N3B)	SegA-slow (EMD-9349, PDB 6N3A)	SegB A315E (EMD-0334, PDB 6N3C)
Data collection and processing				
Magnification	×130,000	×130,000	×130,000	×130,000
Voltage (kV)	300	300	300	300
Electron exposure (e ⁻ /Å ²)	44	44	44	44
Defocus range (μm)	1.1–3.9	1.1–4.4	1.0–9.9	0.5–4.9
Pixel size (Å)	1.064	1.064	1.064	1.06
Symmetry imposed	C ₁	C ₁	C ₂	C ₁
Helical rise (Å)	2.38	4.81	4.84	2.41
Helical twist (°)	178.61	-2.27	179.34	179.36
Initial particle images (no.)	92,568	92,568	499,065	361,182
Final particle images (no.)	8,033	15,464	167,087	44,085
Map resolution (Å)	3.8	3.8	3.3	3.3
FSC threshold	0.5	0.5	0.143	0.143
Map resolution range (Å)	200–3.8	200–3.8	200–3.3	200–3.3
Refinement				
Initial model used (PDB code)	De novo	De novo	De novo	De novo
Model resolution (Å)	3.8	3.8	3.2	3.2
FSC threshold	0.5	0.5	0.5	0.5
Model resolution range (Å)	200–3.8	200–3.8	200–3.2	200–3.2
Map sharpening <i>B</i> factor (Å ²)	-87	-93	-134	-154
Model composition				
Nonhydrogen atoms	2580	2735	4310	4260
Protein residues	360	385	600	640
Ligands	0	0	0	0
<i>B</i> factors (Å²)				
Protein	75.8	105.4	108.6	72.2
Ligand	-	-	-	-
R.m.s. deviations				
Bond lengths (Å)	0.009	0.010	0.008	0.009
Bond angles (°)	1.548	1.372	1.143	1.306
Validation				
MolProbity score	2.28	2.47	2.25	2.35
Clashscore	12.95	22.11	17.25	18.47
Poor rotamers (%)	0	0	0	0
Ramachandran plot				
Favored (%)	85.3	86.3	91.1	88.5
Allowed (%)	14.7	13.7	8.9	11.5

	SegA-sym (EMD-9339, PDB 6N37)	SegA-asym (EMD-9350, PDB 6N3B)	SegA-slow (EMD-9349, PDB 6N3A)	SegB A315E (EMD-0334, PDB 6N3C)
Disallowed (%)	0	0	0	0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Comparative solvation energy calculations

Fibril Structure	Energy per layer (kcal/mol)	Energy per residue (kcal/mol)
SegA-sym (PDB 6N37)	73	1.0
SegA-asym (PDB 6N3B)	76	0.97 (long chain, 0.92; short chain, 1.04)
SegB A315E (PDB 6N3C)	110	0.86 (inner chain, 0.94; outer chain, 0.79)
Tau PHF (PDB 5O3L)	128	0.88
Tau SF (PDB 5O3T)	144	1.0 (chain A, 0.98; chain B, 1.02)
Tau Pick's (PDB 6GX5)	98	1.0
FUS (PDB 5W3N)	41	0.66

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript