Cryo-EM Structure of the Full-length hnRNPA1 Amyloid Fibril

Kartikay Sharma [1*], **Sambhasan Banerjee** [1], **Dilan Savran** [2], **Cedric Rajes** [1], **Sebastian Wiese** [3], **Amandeep Girdhar** [4], **Nadine Schwierz** [5], **Christopher Lee** [4], **James Shorter** [6], **Matthias Schmidt** [1], **Lin Guo** [4,6] and **Marcus Fändrich** [1]

1 - Institute of Protein Biochemistry, Ulm University, 89081 Ulm, Germany

2 - Dr. Senckenberg Institute of Pathology, University Hospital Frankfurt, Frankfurt am Main, Germany

3 - Core Unit Mass Spectrometry and Proteomics, Ulm University, 89081 Ulm, Germany

4 - Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107, USA

5 - Institute of Physics, University of Augsburg, 86159 Augsburg, Germany

6 - Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Correspondence to Kartikay Sharma: *kartikay.sharma@uni-ulm.de* *(K. Sharma)* *@SAMBHASANBANERJ* 🐦 *(S. Banerjee),* *@shorterlab* 🐦 *(J. Shorter)*

Introduction

The misfolding of heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) protein is a common characteristic of several neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD) and multisystem proteinopathy (MSP).[1–3] This member of the family of multifunctional RNA-binding proteins, is thought to play an important role in transcription, mRNA splicing, stability, transport and translation.[4,5] It contains two RNA recognition motifs (RRMs) and a so-called low-complexity domain (LCD) that is glycine-rich and contains several RNA-binding Arg-Gly-Gly repeats.[6,7] The LCD contains many uncharged polar amino acids and is classified as prion-like due to the similarity of its amino-acid composition to the prion domains of certain yeast proteins.[2,8]

The LCD of recombinant and putatively monomeric hnRNPA1 protein is structurally flexible and lacks stable conformational elements.[2,9] It carries a nuclear-localization signal

sequence (NLS) that governs the nuclear localization of the protein under physiological conditions.[10] The nuclear-cytoplasmic shuttling of hnRNPA1 is mediated by karyopherin-β2 (Kapβ2, also known as transportin-1) protein which binds directly to the NLS.[11,12] Cellular stress leads to a translocation of hnRNPA1 into the cytoplasm where it undergoes a liquid–liquid phase separation and participates in the formation of stress granules (SGs).[13–15] Recently, Kapβ2 protein was shown to act as a molecular chaperone of several RNA-binding proteins and inhibited and reversed the formation of amyloid fibrils from hnRNPA1 protein.[16]

Several changes to the hnRNPA1 amino acid sequence were found to be associated with familial forms of ALS or MSP.[2,3,13,17] These changes include the single-site mutations Asp262Asn, Asp262Val, Asn267Ser, Pro288Ser and Pro288Ala as well as frameshift mutations that affect the C-terminal length of hnRNPA1 beyond residue Gly303.[2,3,13,17] The mutational changes Asp262Val, Pro288Ala and Pro288Ser were reported to accelerate the formation of fibrils *in vitro* and to recruit hnRNPA1 protein into SGs when expressed in cultured cells,[2,3,13] Asp262Val and Pro288Ala, decelerate SG disassembly in cultured cells, but do not notably affect the liquid–liquid phase separation of the mutant compared to wild-type (WT) protein *in vitro*.[3] The frameshift mutations affect SG formation in cultured cells and the ability of hnRNPA1 to undergo the liquid–liquid phase separation as well as to form fibrils *in vitro*.[3]

To gain a deeper understanding of mechanism by which full-length hnRNPA1 protein forms amyloid fibrils, we here determined the molecular structure of the full-length hnRNPA1 amyloid fibril using cryo-electron microscopy (cryo-EM). We find that the fibril contains a stable core that is composed of a 45-residue segment of the LCD, whereas the majority of the fibril protein does not participate in formation of the filament structure and decorates the fibril core. The importance of the fibril core is further supported by the observation that the several of disease-related mutations are clustered in the region found ordered by cryo-EM.

Results and Discussion

Cryo-EM structure of the full-length hnRNPA1 amyloid fibril

Amyloid fibrils from full-length hnRNPA1 protein (residues 1–320) were plunge frozen in liquid ethane and imaged with a cryo electron microscope at 300 kV. The resulting images show the presence of mainly one dominant fibril morphology that is defined by a width of 19.9 ± 1.1 nm and a fuzzy appearance without well-resolved cross-overs (Figure 1(a)). Two-dimensional (2D) class averages obtained with the computationally cropped segments of the fibril revealed the helical structure of the fibril and

showed a well-resolved feature with 4.7–4.8 Å spacing originating from the stacking of the β-strands along the fibril axis (Supplementary Figure 1(a)). The three-dimensional (3D) map, which was reconstructed from the cryo-EM data, achieved a spatial resolution of 3.32 Å, based on the 0.143 Fourier Shell Correlation (FSC) criterion (Figure 1(b), Supplementary Figure 1(b,c)). The reconstruction implemented a left-hand fibril twist, as a left-handed twist of the fibrils was obtained with platinum-side shadowing and Transmission electron microscope (TEM) (Figure 1(b), Supplementary Figure 1(d)). The reconstructed 3D map was used to build an atomic model of the fibril (Figure 1(b)).

The fibril core is formed by 45 residues (Gly251 to Tyr295) from the LCD (Figure 1(c)). Each fibril molecules extends by ∼7 Å along the fibril axis, as measured at the Cα atoms (Figure 1(d)). Residues Met1 to Asp250 and Phe296 to Phe320 are not seen in our 3D map suggesting that these N- and C-terminal segments of the fibril protein are conformationally disordered. The N- and C-terminal residues of the fibril core (Gly251 and Tyr295) are solvent exposed (Figure 1(c)), allowing conformationally heterogenous N- and C-terminal segments to protrude from and to decorate the fibril core. This conclusion is consistent with the diffuse appearance of the fibrils in the cryo-EM micrographs (Figure 1(a)) and the substantially smaller width of the 3D map (∼6.2 nm) compared to the fibril width measured with the cryo-EM micrographs (Figure 2).

Proteolysis of full-length hnRNPA1 amyloid fibril

To further confirm that the N- and C-terminal segments of hnRNPA1 protein surround the fibril core, we subjected the hnRNPA1 fibrils to proteolysis with proteinase K (PK). Proteolysis resulted in three bands on the gel at approximately 20, 10 and 7 kDa (Supplementary Figure 2). To ascertain the molecular identity of these bands, three bands were digested with trypsin and analyzed with mass spectrometry (Supplementary Figure 3–5, Supplementary Tables 3–5). Analysis of the six most intense peaks in the mass spectra of the 20 and 7 kDa bands resulted in the identification of a number of N-terminal protein fragments but no fragment from LCD. In the case of the 10 kDa band, a similar analysis identified one fragment from the fibril core along with several N-terminal fragments. Our results suggest that the fibril core of hnRNPA1 fibrils is very unstable when exposed to PK and becomes digested more rapidly than the N-terminal RRM domains, presumably because they are folded into globular domains. Our Results further strengthen the conclusion that the N- and C-terminal segments of hnRNPA1 protein surround the fibril core.
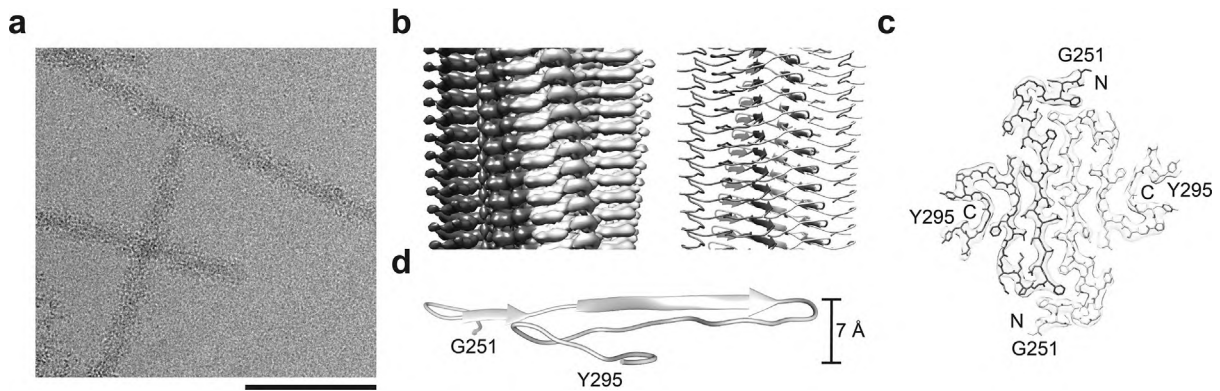
Figure 1. Cryo-EM structure of the full-length hnRNPA1 fibril. (a) Cryo-EM image of the fibrils. Scale bar: 100 nm. (b) Side view of the 3D map (left) and of the molecular model represented as ribbon diagram (right). (c) Cross-sectional view of one molecular layer of the 3D map, superimposed with the molecular model. N-terminal and C-terminal amino acid residues of the fibril core are labelled. In panels (b) and (c), the two fibril protein stacks are colored in different shades of gray to guide the eye. d) Side view of fibril protein showing z-axial height of the protein backbone.
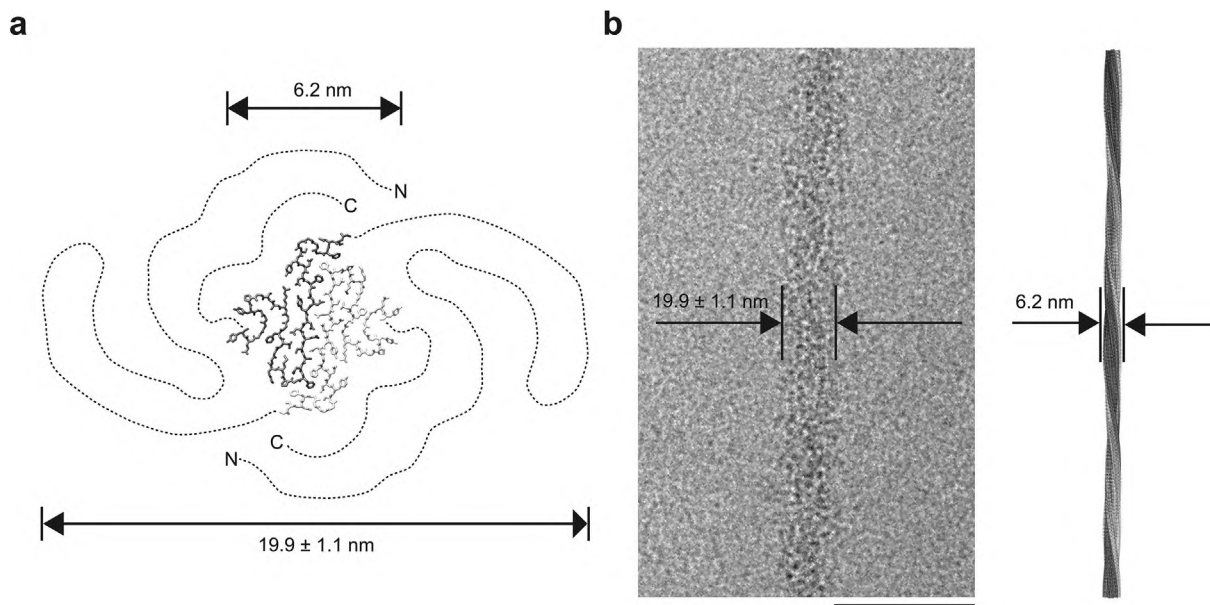

Figure 2. Comparison of fibril width in the micrograph and in the 3D map. (a) Cross-sectional representation of the fibril core with flexible and structurally disordered segments being added as dotted lines in arbitrary conformation. It is possible that the N-terminal segment contains regions with stable conformation at the RRM domains not shown in the figure. (b) Cryo-EM micrograph image (left) and 3D map of the fibril (right) scaled to the same size. Scale bar: 50 nm.

Structural features of the fibril

The fibril is polar and consists of two staggered fibril protein stacks (Figure 1b). The fibril shows a pseudo 2_1-screw symmetry. Together with the axial height change of the protein molecules (see above), the symmetry helps to sterically interlock the proteins in the fibril. There are two β-strands per fibril protein, termed here β1 and β2, which extend from residue Tyr260 to Asp262 and from residue Asn265 to Asn272 (Figure 3). The two

fibril protein stacks pack at a distance of ~7.3 Å perpendicular to the fibril axis. This spacing defines, together with the 4.7–4.8 Å spacing of the molecules in the direction of the fibril axis, the fibril cross-β structure. The fibril structure appears to be stabilized through highly cooperative networks of inter- and intramolecular interactions. These interactions can roughly be divided into interactions that extend within the cross-sectional plane of the fibril, such as intra- and inter-molecular side chain-side chain interactions, and
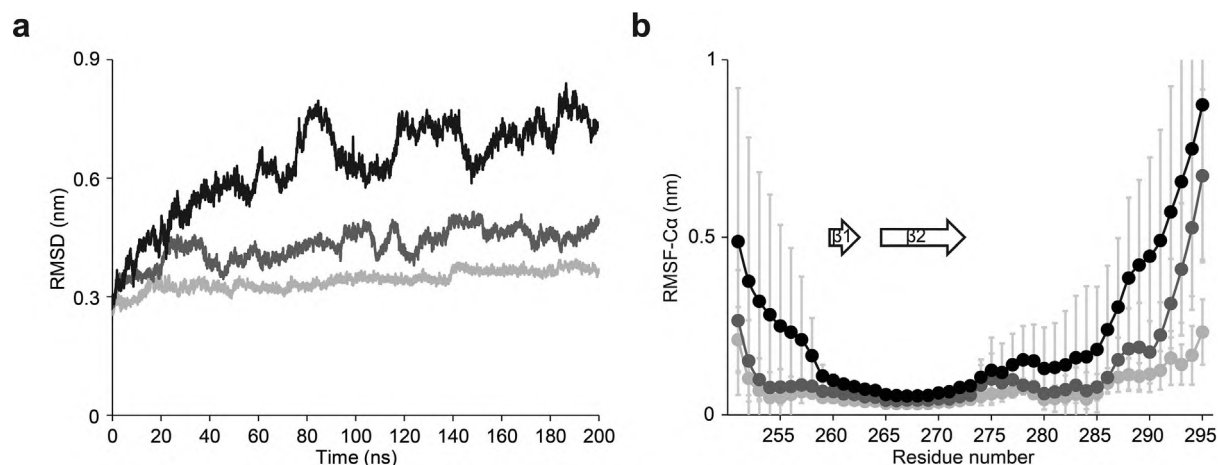
Figure 3. MD simulations of the hnRNPA1 fibril structure. (a) Temporal development of the RMSD at different temperatures (light gray: 300 K, dark gray: 350 K, black: 400 K). (b) Averaged RMSF values of the Cα atoms of residues Gly251-Tyr295 (light gray: 300 K, dark gray: 350 K, black: 400 K). Arrows above the x-axis indicate the position of β-strands in the fibril protein. Error bars show the standard deviation ($n = 12$).

interactions that run along the fibril z-axis. The latter include the backbone-backbone hydrogen bonds, which define the fibril cross-β structure, and intermolecular side chain-side chain interactions that run across different layers of the fibril. These interactions are visualized by the molecular model, although the current resolution of our reconstruction does not allow their clear observation in the 3D map.

To obtain further insights into the molecular features responsible for the fibril stability, we carried out molecular dynamics (MD) simulations of the fibril core at 300 K, 350 K and 400 K. These methods reveal relatively low root mean square fluctuations (RMSFs) at the Cα atoms of residues Tyr260 to Phe273 (Figure 3), indicating a high structural stability of this sequence segment. As these residues form also the fibril cross-β structure, we conclude that the β-sheets constitute the most stable structural elements of the fibril. The core residues, which are closer to the disordered parts of the fibril protein, are more vulnerable to a loosening of their structure, as indicated by their higher structural fluctuations (Figure 3). This effect is particularly strong at residues Gly287-Tyr295, which are also prone to dissociate from the fibril core as demonstrated by measurement of the distance between the residues in the terminal fibril layers ($i + 2$ and $i − 3$) and their next neighbors (layers $i + 1$ and $i − 2$) (Supplementary Figure 6). Interestingly, layer $i − 3$ shows a greater propensity to dissociate from the fibril protein stack than layer $i + 2$, indicating that the molecules at the two opposing ends of the fibril differ in their dissociation susceptibility (Supplementary Figure 6).

The segment, which is in particular affected by the structural loosening in our MD simulation, relates to the NLS of hnRNPA1 and thus to the region of the protein that is bound by Kapβ2.[12] Kapβ2 dissociates preformed hnRNPA1 aggregates *in vitro* and in cultured cells[16] and requires the hnRNPA1 PY (Pro288-Tyr289) motif at the NLS as well as residue Arg284 for binding.[12] Interestingly, all three residues show increased distances between the terminal layers at the end of our MD analysis, indicating that they support the dissociation of molecules from the fibril ends (Supplementary Figure 6).

The fibril cross-β structure does not match well with the regions of high aggregation propensity

Although the involvement of the LCD in fibril formation was reported by several previous studies,[2,14,15,18] its importance for the formation of the fibril cross-β structure is surprising − at least when comparing the fibril cross-β structure with the intrinsic aggregation propensity of hnRNPA1. The intrinsic aggregation propensity of a sequence can be revealed by widely used computational tools, such as ZipperDB,[19] Waltz,[20] Aggrescan,[21] FoldAmyloid,[22] PASTA 2.0,[23] TANGO 2.2[24] and Amylpred 2.[25] Analysis of the hnRNPA1 sequence with these methods consistently identifies segments within the two RRMs, but not within the cross-β sheet forming sequence elements, as particularly aggregation-prone (Figure 4). Five out of seven tested methods (Aggrescan, FoldAmyloid, Pasta 2.0, TANGO 2.2 and Amylpred 2) lacked any hit within the fibril cross-β structure (Figure 4). Only two (ZipperDB and WALTZ) showed hits within these secondary structural elements. In case of ZipperDB, however, the observation of true positives is compromised by a very high number (85) of false positives in the disordered segments of the fibril protein (Figure 4), indicating that this method
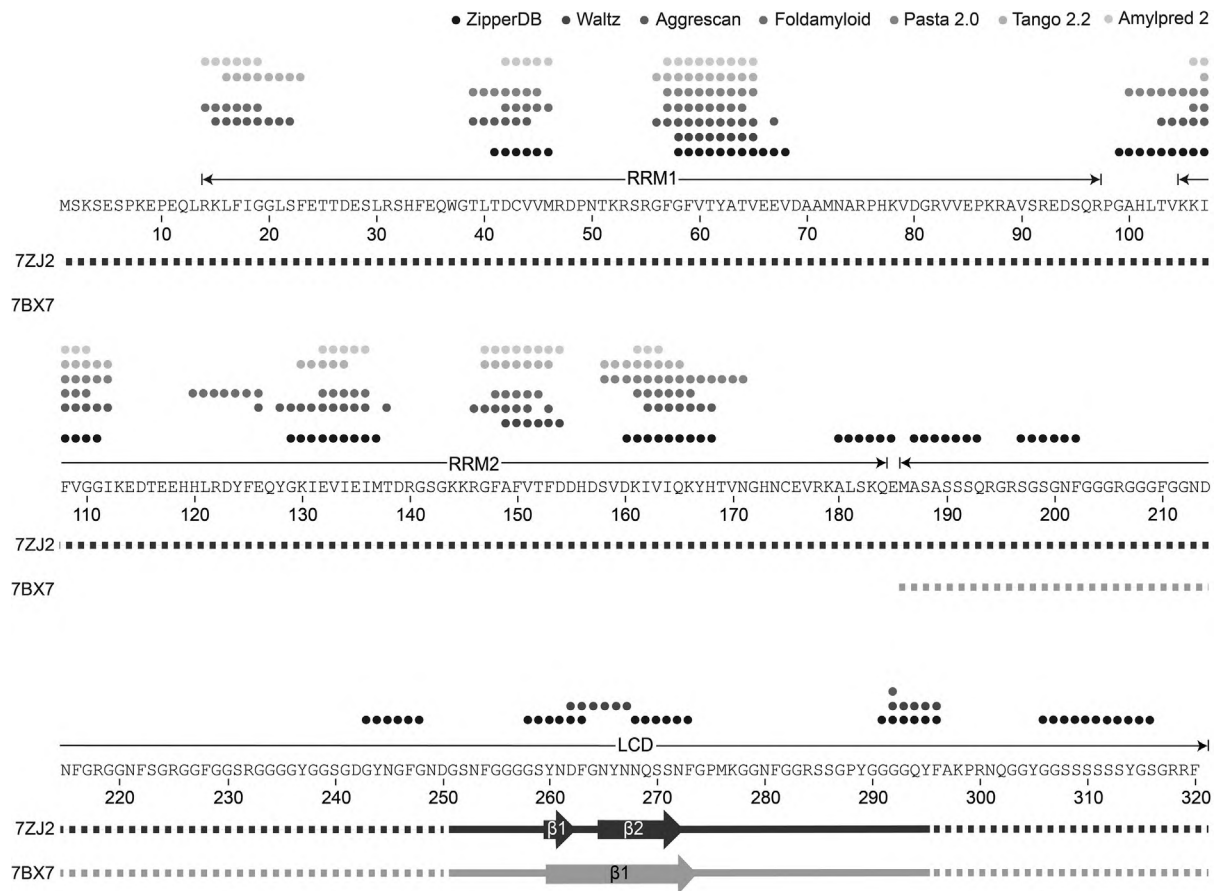
4

Figure 4. Location of the fibril cross-β structure in the hnRNPA1 sequence. Residue specific prediction of aggregation prone regions within full-length hnRNPA1 protein using seven different prediction tools. The dotted and continuous lines refer to the fibril structures of full-length (black, Protein Data Bank (PDB) entry 7ZJ2) and N-terminally truncated hnRNPA1 protein (grey, PDB entry 7BX7).[18] Dotted lines: residues not seen in the 3D map; continuous line: residues seen in the 3D map. Arrows above the lines indicate β-strands. The β-strands were drawn according to the respective publication and differences in their position may thus reflect different definitions. For ZipperDB, every residue in the predicted aggregation prone six-residue sequence segment were counted as hits and are assigned with a dot.

shows low specificity. One of the segments predicted by ZipperDB (residues Gly243 to Gly248) was shown to adopt a so-called low-complexity aromatic-rich kinked segment structure as a peptide crystal.[26] The apparently best compromise between true and false positives was obtained with WALTZ, which showed 4 true positives within the cross-β structure and only 21 false positive residues within the remaining parts (ordered and disordered) of the fibril protein (Figure 4).

In conclusion, the tested tools to predict the amyloidogenic potential of a sequence do not specifically predict the sequence segments that form the fibril cross-β structure of the full-length hnRNPA1 protein. Related conclusions were obtained previously with experimentally verified amyloid fibril structures from other fibril proteins, such as the fibrils obtained from immunoglobulin light chains and serum amyloid A protein. In all cases, it was found that the predictions of the

highly amyloidogenic regions did match well the experimentally verified β-sheet structure in the amyloid fibril.[27–29] These conclusions with amyloid fibrils from relatively large-sized proteins (>50 amino acids) contrast to observations made with short peptide fragments (<15 amino acids) corresponding to segments with a high intrinsic aggregation propensity where many studies show that these segments avidly form cross-β structures as amyloid fibrils and peptide crystals.[26,30–32]

Location of the disease-related mutations in the fibril structure

The structural importance of the LCD, which we reveal here for the full-length hnRNPA1 fibrils with cryo-EM, is indicated further by observations that several ALS- and MSP-associated mutations are located within residues Gly251 to Tyr295.[2,3,13,17] Examples hereof include the disease-associated

hnRNPA1 mutations Asp262Asn, Asp262Val, Asn267Ser, Pro288Ala and Pro288Ser (Supplementary Figure 7(a)). At least two of the affected residues occur at strategic positions of the fibril architecture. Asn267 is located in the center of the fibril where it forms part of the interface between the two protein stacks, while Asp262 helps to define the fibril protein fold by forming an intermolecular salt bridge with Arg284 (Supplementary Figure 7 (a)). The mutational changes, Asp262Asn and Asp262Val, remove the charge at residue 262 and lead to a substantial change of the residue's chemical properties. Indeed, Asp262Asn, Asp262Val and Asn267Ser hnRNPA1 fibrils are morphologically different from wild WT hnRNPA1 fibrils, as demonstrated by TEM (Supplementary Figure 7(b)). The same observation is made with fibrils formed from Pro288Ser hnRNPA1 protein, although residue Pro288 does not seem to be involved in the formation of crucial interactions within the fibril structure (Supplementary Figure 7(b)). The observation of structural effects upon mutation differs from previous results obtained with fibrils formed from shorter constructs (residues Met286-Phe320) of hnRNPA1 protein, where the mutational variants Asp262Val and Asp262Asn were reported not to substantially affect the fibril structure.[18] According to our data, however, each mutation leads to a visibly different fibril morphology as defined by the fibril width and the distance between adjacent crossovers (Supplementary Figure 7(b)).

Interestingly, we find that the full-length hnRNPA1 protein forms a fibril core that is essentially indistinguishable from the one formed from
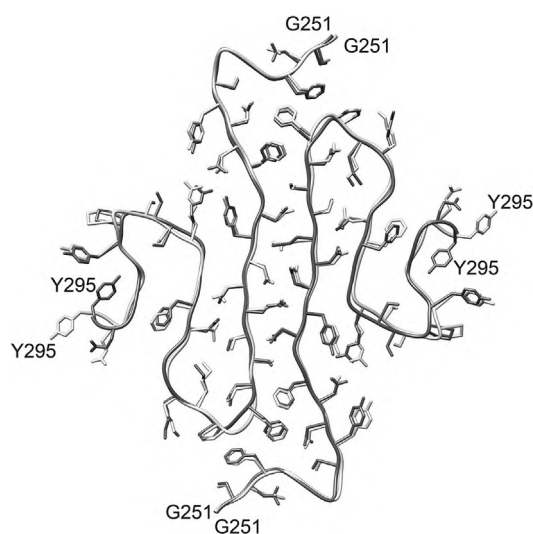
truncated hnRNPA1 protein (Figure 5). The Root mean square deviation (RMSD) of the two structures is 0.688 Å. This similarity is remarkable given that the two fibrils were formed not only by different protein constructs but also under different conditions of fibril formation. Full-length hnRNPA1 fibrils were formed at 25 °C in 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-KOH, 150 mM KCl, 5% (v/v) glycerol, 20 mM glutathione, pH 7.4, whereas the fibrils from the truncated protein were grown by its incubation at 16 °C in 50 mM tris(hydroxymethyl)aminomethane-HCl, 100 mM NaCl, pH 7.5.[18] That is, different protein constructs and different conditions of fibril formation did not substantially affect the formed fibril structure. This finding contrasts sharply to observations made with other fibril proteins, such as from α-synuclein, TAR DNA-binding protein 43 and Aβ peptide, where even slight change in the conditions of fibril formation or in the length of the protein construct resulted in clearly different fibril morphologies.[33–35]

Conclusion

Using cryo-EM, we have obtained the amyloid fibril structure from full-length hnRNPA1 protein (Figure 1). We found that the fibril core is formed from a segment of the LCD, while the more N- and C-terminal segments of the protein are structurally heterogenous and are not seen in our 3D map (Figures 1 and 2). Our structure matches a recently obtained structure of a N-terminally truncated fragment of hnRNPA1 protein (Figure 5) that lacked the two RRM domains.[18] Our analysis now shows that these domains do not contribute to the formation of the core of full-length hnRNPA1 fibrils, although they exhibit the highest intrinsic amyloidogenic potential of the fibril protein sequence (Figure 4). This finding shows that it is currently difficult, if not impossible, to correctly anticipate which sequence segments will form the fibril cross-β structure. Our data therefore support the idea that the amyloid cross-β structure represents a generic state of the polypeptide chain, which is primarily adopted due to an intrinsic structural prediction of the polypeptide backbone,[36,37] while the amino acid side chains and their specific interactions are more relevant to determine the packing of the cross-β structural elements into a stable fibril.[30]

Material and Methods

Protein purification and fibril formation

WT hnRNPA1 and its mutagenic variants Asp262Asn hnRNPA1, Asp262Val hnRNPA1, Asn267Ser hnRNPA1 and Pro288Ser hnRNPA1 were expressed and purified from *E. coli* as glutathione *S*-transferase (GST)-tagged proteins.

Figure 5. Comparison of the fibril core in full-length and N-terminally truncated hnRNPA1 fibrils. Super-imposition of one molecular layer of the fibril core of full-length hnRNPA1 (light gray, PDB entry 7ZJ2) and N-terminally truncated hnRNPA1 protein (dark gray, taken from the PDB entry 7BX7).[18]

Expression constructs were generated in pDuet to contain a tobacco etch virus (TEV)-cleavable site, resulting in a GST-TEV-hnRNPA1 construct.[38,39] GST-TEV-hnRNPA1 was overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Agilent) and purified under the native conditions using a glutathione-sepharose column (GE Healthcare) according to the manufacturer's instructions. Proteins were eluted from the glutathione sepharose column with assembly buffer (AB; 40 mM HEPES-KOH, 150 mM KCl, 5% (v/v) glycerol, 20 mM glutathione, pH 7.4, Supplementary Figure 8). The protein was centrifuged for 10 min at 16,100 g, and the supernatant was separated from the pellet to remove protein aggregates. The protein concentration was determined using the Bradford assay (Bio-Rad) and bovine serum albumin as a standard. Aggregation was initiated by addition of TEV protease (Invitrogen) to GST-TEV-hnRNPA1 (5 μM) in AB. Fibrils were assembled by incubation at 25 °C for 24 h with agitation at 1,200 r.p.m. using an Eppendorf Thermomixer.

TEM analysis of negatively-stained specimens

Formvar and carbon-coated 200 mesh copper EM grids (Electron Microscopy Sciences) were glow discharged for 40 s at 40 mA using PELCO easiGlow glow discharge cleaning system (TED PELLA). 3.5 μl of fibril sample were placed on the grid. After 1 min, the sample was blotted with filter paper (Whatman) and the grids were washed three times with 10 μl water followed by three washing with 10 μl 2% (w/v) uranyl acetate solution. The grids were allowed to dry for 10 min, before they were analyzed at 120 kV with a JEOL1400F TEM equipped with CMOS camera (TVIPS).

Platinum side shadowing and analysis

The specimens for platinum side shadowing were prepared by glow discharging formvar and carbon-coated 200 mesh copper EM grids (Electron Microscopy Sciences) for 40 s at 40 mA using PELCO easiGlow glow discharge cleaning system (TED PELLA). 3.5 μl of the fibril sample were placed on the grids and incubated for 1 min before the grids were blotted with filter paper (Whatman). The grids were washed three times with 10 μl water and allowed to dry for 15 min. Afterwards, grids were coated with a 1 nm thick layer of platinum which was evaporated at an angle of 30° with a Balzers TKR 010 system. The grids were analyzed using a JEOL1400F TEM which was operated at an accelerating voltage of 120 kV and equipped with a CMOS camera (TVIPS).

Cryo-EM

The C-flat holey carbon grids (CF 1.2/1.3-4C, Electron Microscopy Sciences) were glow-discharged using PELCO easiGlow glow discharge cleaning system (TED PELLA) and added with 3.5 μl of fibril sample. The sample was incubated on the grid for 30 s, blotted on both sides for 8 s at a humidity > 95% and then plunge-frozen in liquid ethane using FEI Vitrobot MarK-3 (Thermo Fisher Scientific). The plunge-frozen grids were screened with a JEM 2100F (JEOL) TEM at an accelerating voltage of 200 kV and equipped with a CMOS camera (TVIPS). The image data for the cryo-EM reconstruction were recorded with a Titan Krios TEM operated at accelerating voltage of 300 kV. Images were taken with a K2 quantum detector (Gatan). The data acquisition parameters are listed in the Supplementary Table 1. The fibril width and crossover distances were measured using Fiji.[40]

Helical reconstruction

The recorded multi-frame tiff files were converted to mrcs files and gain corrected using the IMOD software.[41] The converted mrcs files were aligned and motion-corrected using Motioncor2.[42] CTFFind 4.1[43] was used to estimate the contrast transfer function of the motion-corrected micrographs. The helical 3D reconstruction of the fibrils was done with Relion 3.1.[44] 54,904 particles with a box size of 350 pixels and inter-box distance of 13.7 Å (~3.76%) were extracted from manually picked fibrils in 2,624 micrographs. The reference free 2D classes were generated to evaluate the homogenous subsets of the extracted particles. Initially, a featureless cylinder was used as a reference map for the 3D classification. The best class with 54,408 particles was re-extracted with a box size of 256 pixels and an inter-box distance of 13.7 Å (~5.4%). Due to the cross-sectional similarity of the 3D class average with the PDB entry 7BX7 we used the 3D map corresponding to 7BX7 (Electron Microscopy Data Bank, EMDB, entry EMD-30235)[18] as a reference map for the subsequent 3D classification. The re-extracted particles were subjected to 3D classification using a single class and a regularization parameter value of 60 followed by a 3D auto-refinement. The beam tilt, aberrations, defocus per particle and astigmatism per micrograph were adjusted by contrast transfer function refinement tool of RELION to enhance the resolution of the 3D constructed map. The adjusted particles were polished by the RELION's implementation of Bayesian polishing and auto-refinement. The final 3D map with a soft-edge mask was obtained. The reconstruction details are listed in Supplementary Table 1.

Model building and refinement

The post-processed, masked 3D map was sharpened by the autosharpen tool in Phenix[45] and forwarded to model building in Coot.[46] The fibril model for truncated hnRNPA1 fibrils (PDB entry

7BX7) was used as a starting model. A single polypeptide chain was manually fitted into the density map in Coot. The structural refinement was done in Coot by imposing Ramachandran and beta-sheet restraints. The steric clashes between atoms, rotamer and Ramachandran outliers were analyzed by the validation report generated using the validation tool in Phenix and subsequently corrected in Coot. Once a reasonable main- and side-chain density fit was achieved for one single polypeptide chain, a fibril stack composed of twelve polypeptide chains was built using the pdbsymm tool implemented in Situs.[47] The described iterative refinement process and validation was repeated for fibril stack until a satisfactory density to model was achieved. The secondary structure assignment was done manually for the fibril model. The beta-strands were defined by the residues which lie in the β-sheet region in the Ramachandran plots and involved in backbone hydrogen bonding (bond distance <3 Å) along the fibril axis. The model building and refinement details are listed in Supplementary Table 2.

Prediction of aggregation-prone regions

The residue specific aggregation propensity of hnRNPA1 was determined using seven different computational programs: ZipperDB, LARKSdb, Waltz, Aggrescan, FoldAmyloid, PASTA 2.0, TANGO 2.2 and Amylpred 2. In case of ZipperDB, if a residue with Rosetta energy below the threshold of -23 kcal/mol was found then the corresponding residue and the following five residues were counted as hits. For Aggrescan, residues with an a^4v value of greater than -0.02 were counted as hits. In the case of Amylpred 2, the consensus method was used to assign a residue as a hit. For FoldAmyloid, 5 or more successive residues with a score greater than 21.4 were counted as hits. In PASTA 2.0, thresholds were kept on the region with 90% specificity which means a residue was counted as a hit if the energy was below the cutoff value of -2.8 PASTA energy units. For TANGO 2.2, an amino acid was counted as a hit that have a score above 5%. In the case of WALTZ, residue with a score greater than 0.00 was counted as a hit.

MD simulations

All-atom MD simulations were used to characterize the conformational dynamics of the fibril structure. The cubic simulation box had a size of 102.75 Å. The box was filled with 33,308 water molecules. The system was neutralized with 0.15 M NaCl, leading to a system size of 140,386 atoms. The force-field parameters for the peptides were taken from Amber99sb-star-ildn.[48] The TIP4P-Ew model[49] was used for the water molecules. For NaCl, the Mamatkulov-Schwierz force field parameters[50] were used. The simulations were performed at three different temperatures (T) with a fixed particle number (N) and volume (V), using the Gromacs simulation package, version 2018.[51,52] Periodic boundary conditions were applied, and the particle-mesh Ewald method was used for the periodic treatment of Coulombic interactions.[53] Bonds to hydrogen atoms were constrained using LINCS,[54] and a 2 fs time step was used. Close Coulomb real space interactions were cut off at 1.2 nm and the Lennard-Jones interactions were cut off after 1.2 nm, respectively. The long-range dispersion corrections for energy and pressure (P) were applied to account for the errors stemming from truncated Lennard-Jones interactions. To equilibrate the system at 300 K, first an energy minimization with the steepest descent algorithm was performed. The system was equilibrated for 1 ns, first in the NVT, where N, V and T are constant (canonical ensemble). Afterwards, it was equilibrated in the NPT, where N, P and T are constant (isothermal–isobaric ensemble). For both, the Berendsen thermostat and barostat with $\tau_T = 0.1$ ps and $\tau_P = 0.1$ ps was used. Afterwards, this system was used to equilibrate at 350 K by running for 1 ns in NVT with the Berendsen thermostat. To equilibrate at 400 K, the system at 350 K was used and equilibrated for 1 ns in NVT via the Berendsen thermostat. For the NVT production run, we performed 200 ns simulations employing the velocity rescaling thermostat with stochastic term[55] with a time constant of $\tau_T = 0.1$ ps. The RMSF was calculated from the production run discarding the first 50 ns for equilibration. For the RMSF and RMSD, the experimental structure was used as reference. The maximum distance between two adjacent residues within one stack was calculated using the production run after 50 ns.

Accession Numbers

The reconstructed cryo-EM map was deposited in the EMDB with the accession codes EMD-14739. The coordinates of the fitted atomic model were deposited in the PDB under the accession code 7ZJ2. The following previously published coordinates were used in Figures 4 and 5: PDB entry 7BX7.

Author Contribution

K.S., S.B., D.S., C.R. S.W. and A.G. carried out experiments. A.G., C.L., J.S and L.G. contributed reagents and materials. K.S, S.B., D.S., C.R., A. G., N.S., J.S., M.S., L.G. and M.F. analyzed data. M.F. designed research. K.S. and M.F. wrote the paper. All authors could comment on the manuscript.

DATA AVAILABILITY

Data will be made available on request.

DECLARATION OF COMPETING INTEREST

All authors except J.S. declare no conflict of interest. J.S. is a consultant for Dewpoint Therapeutics, ADRx, and Neumora. J.S. is an advisor and shareholder for Confluence Therapeutics.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168211.

Abbreviations:
hnRNPA1, Heterogeneous nuclear ribonucleoprotein A1; ALS, Amyotrophic lateral sclerosis; FTLD, Frontotemporal lobar degeneration; MSP, Multisystem proteinopathy; RRM, RNA recognition motif; LCD, Low-complexity domain; NLS, Nuclear-localization signal sequence; Kapβ2, Karyopherin-β2; SG, Stress granule; WT, Wildtype; cryo-EM, cryo-Electron microscopy; 2D, Two-dimensional; 3D, Three-dimensional; FSC, Fourier shell correlation; TEM, Transmission electron microscope; MD, Molecular dynamics; RMSF, Root mean square fluctuation; RMSD, Root mean square deviation; HEPES, 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; GST, Glutathione S-transferase; TEV, Tobacco etch virus; AB, Assembly buffer; EMDB, Electron microscopy data bank; PDB, Protein data bank; T, Temperature; N, Particle number; V, Volume; P, Pressure; EMBL, European molecular biology laboratory; PK, Proteinase K

References

1. Clarke, J.P., Thibault, P.A., Salapa, H.E., Levin, M.C., (2021). A comprehensive analysis of the role of hnRNP A1 function and dysfunction in the pathogenesis of neurodegenerative disease. *Front. Mol. Biosci.* **8**, 217.

2. Kim, H.J., Kim, N.C., Wang, Y.D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., et al., (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473.

3. Beijer, D., Kim, H.J., Guo, L., O'Donovan, K., Mademan, I., Deconinck, T., van Schil, K., Fare, C.M., et al., (2021). Characterization of HNRNPA1 mutations defines diversity in pathogenic mechanisms and clinical presentation. *JCI Insight* **6**, e148363.

4. Kapeli, K., Martinez, F.J., Yeo, G.W., (2017). Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum. Genet.* **136**, 1193–1214.

5. Beusch, I., Barraud, P., Moursy, A., Cléry, A., Allain, F.H. T., (2017). Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *Elife* **6**, e25736

6. Jean-Philippe, J., Paz, S., Caputi, M., (2013). hnRNP A1: The Swiss army knife of gene expression. *Int. J. Mol. Sci.* **14**, 18999–19024.

7. He, Y., Smith, R., (2008). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell. Mol. Life Sci.* **66**, 1239–1256.

8. Harrison, A.F., Shorter, J., (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474**, 1417–1438.

9. Calabretta, S., Richard, S., (2015). Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **40**, 662–672.

10. Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, M.C., Wang, F., Dreyfuss, G., (1996). A novel receptor-mediated nuclear protein import pathway. *Cell* **86**, 985–994.

11. Bonifaci, N., Moroianu, J., Radu, A., Blobel, G., (1997). Karyopherin β2 mediates nuclear import of a mRNA binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5055–5060.

12. Lee, B.J., Cansizoglu, A.E., Süel, K.E., Louis, T.H., Zhang, Z., Chook, Y.M., (2006). Rules for nuclear localization sequence recognition by Karyopherinβ2. *Cell* **126**, 543–558.

13. Liu, Q., Shu, S., Wang, R.R., Liu, F., Cui, B., Guo, X.N., Lu, C.X., Li, X.G., et al., (2016). Whole-exome sequencing identifies a missense mutation in hnRNPA1 in a family with flail arm ALS. *Neurology* **87**, 1763–1769.

14. Gui, X., Luo, F., Li, Y., Zhou, H., Qin, Z., Liu, Z., Gu, J., Xie, M., et al., (2019). Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nature Comm.* **10**, 1–12.

15. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., Taylor, J.P., (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133.

16. Guo, L., Kim, H.J., Wang, H., Monaghan, J., Freyermuth, F., Sung, J.C., O'Donovan, K., Fare, C.M., et al., (2018). Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell* **173**, 677–692.

17. Naruse, H., Ishiura, H., Mitsui, J., Date, H., Takahashi, Y., Matsukawa, T., Tanaka, M., Ishii, A., et al., (2018). Molecular epidemiological study of familial amyotrophic lateral sclerosis in Japanese population by whole-exome sequencing and identification of novel HNRNPA1 mutation. *Neurobiol. Aging.* **61** (255), e9–255.e16.

18. Sun, Y., Zhao, K., Xia, W., Feng, G., Gu, J., Ma, Y., Gui, X., Zhang, X., et al., (2020). The nuclear localization sequence mediates hnRNPA1 amyloid fibril formation revealed by cryoEM structure. *Nature Comm.* **11**, 1–8.

19. Goldschmidt, L., Teng, P.K., Riek, R., Eisenberg, D., (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3487–3492.

20. Beerten, J., van Durme, J., Gallardo, R., Capriotti, E., Serpell, L., Rousseau, F., Schymkowitz, J., (2015). WALTZ-DB: a benchmark database of amyloidogenic hexapeptides. *Bioinformatics* **31**, 1698–1700.

21. Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., Ventura, S., (2007). AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**, 1–17.

22. Garbuzynskiy, S.O., Lobanov, M.Y., Galzitskaya, O.V., (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **26**, 326–332.

23. Walsh, I., Seno, F., Tosatto, S.C.E., Trovato, A., (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **42**, 301–307.

24. Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., Serrano, L., (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnol.* **22**, 1302–1306.

25. Tsolis, A.C., Papandreou, N.C., Iconomidou, V.A., Hamodrakas, S.J., (2013). A Consensus Method for the Prediction of 'Aggregation-Prone' Peptides in Globular Proteins. *PLoS One* **8**, e54175.

26. Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., Rodriguez, J.A., Cascio, D., Chong, L., Gonen, T., et al., (2018). Atomic structures of low-complexity protein segments reveal kinked b sheets that assemble networks. *Science* **359**, 698–701.

27. Radamaker, L., Karimi-Farsijani, S., Andreotti, G., Baur, J., Neumann, M., Schreiner, S., Berghaus, N., Motika, R., et al., (2021). Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. *Nature Comm.* **12**, 1–11.

28. Liberta, F., Loerch, S., Rennegarbe, M., Schierhorn, A., Westermark, P., Westermark, G.T., Hazenberg, B.P.C., Grigorieff, N., et al., (2019). Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nature Comm.* **10**, 1–10.

29. Rennegarbe, M., Lenter, I., Schierhorn, A., Sawilla, R., Haupt, C., (2017). Influence of C-terminal truncation of murine Serum amyloid A on fibril structure. *Sci. Rep.* **7**, 1–8.

30. Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A., Riekel, C., Grothe, R., Eisenberg, D., (2005). Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778.

31. Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., et al., (2007). Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447**, 453–457.

32. Schmidt, A., Annamalai, K., Schmidt, M., Grigorieff, N., Fändrich, M., (2016). Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6200–6205.

33. Kollmer, M., Close, W., Funk, L., Rasmussen, J., Bsoul, A., Schierhorn, A., Schmidt, M., Sigurdson, C.J., et al., (2019). Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nature Comm.* **10**, 1–8.

34. Sawaya, M.R., Hughes, M.P., Rodriguez, J.A., Riek, R., Eisenberg, D.S., (2021). The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873.

35. Mehra, S., Gadhe, L., Bera, R., Sawner, A.S., Maji, S.K., (2021). Structural and Functional Insights into α-Synuclein Fibril Polymorphism. *Biomolecules* **11**, 1419.

36. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., Dobson, C.M., (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3590–3594.

37. Fändrich, M., Dobson, C.M., (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690.

38. Johnson, B.S., Snead, D., Lee, J.J., McCaffery, J.M., Shorter, J., Gitler, A.D., (2009). TDP-43 Is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339.

39. Sun, Z., Diaz, Z., Fang, X., Hart, M.P., Chesi, A., Shorter, J., Gitler, A.D., (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol.* **9**, e1000614.

40. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., et al., (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682.

41. Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., (1996). Computer Visualization of Three-Dimensional Image Data Using IMOD. *J. Struct. Biol.* **116**, 71–76.

42. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., Agard, D.A., (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331–332.

43. Rohou, A., Grigorieff, N., (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221.

44. Scheres, S.H.W., (2020). IUCr, Amyloid structure determination in RELION-3.1. *Acta Crystallogr. D Struct. Biol.* **76**, 94–101.

45. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., et al., (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877.

46. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.

47. Wriggers, W., (2012). Conventions and workflows for using Situs. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 344–351.

48. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E., (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958.

49. Horn, H.W., Swope, W.C., Pitera, J.W., Madura, J.D., Dick, T.J., Hura, G.L., Head-Gordon, T., (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665–9678.

50. Mamatkulov, S., Schwierz, N., (2018). Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties. *J. Chem. Phys.* **148**, 074504

51. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718.

52. Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., (2008). GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447.

53. Darden, T., York, D., Pedersen, L., (1998). Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.

54. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., (1997). LINCS: a Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **18**, 1463–1472.

55. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101