

Cryo-EM Structure of a Mammalian-specific Alternative Amyloid Exon

Javier Garcia-Pardo^a, Andrea Bartolomé-Nafria^a, Antonio Chaves-Sanjuan^{b,c}, Marcos Gil-Garcia^a, Cristina Visentin^{b,d}, Martino Bolognesi^{b,c}, Stefano Ricagno^{b,d} and Salvador Ventura^{a,*}

^aInstitut de Biotecnologia i de Biomedicina (IBB) and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

^bDipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milano, Italy.

^cCRC Fondazione Romeo e Enrica Invernizzi and NOLIMITS, Università degli Studi di Milano, 20133 Milano, Italy.

^dInstitute of Molecular and Translational Cardiology, I.R.C.C.S. Policlinico San Donato, piazza Malan, 2, 20097, San Donato Milanese, Italy.

* To whom correspondence should be addressed: Salvador Ventura, Institut de Biotecnologia i de Biomedicina (IBB) and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. Phone: +34-93-586 8956; Fax: +34-93-581 1264; E-mail: Salvador.Ventura@uab.cat.

Keywords: hnRNPD, Amyloid, Alternative splicing, Exon, Low complexity domain, LGMDD3, cryo-EM structure.

ABSTRACT

hnRNPD_L is a ribonucleoprotein (RNP) involved in transcription and RNA-processing, with missense mutations causing limb-girdle muscular dystrophy-3 (LGMDD3). Mammalian-specific alternative splicing (AS) renders three natural isoforms, hnRNPD_L-2 being predominant in humans. We present the cryo-electron microscopy structure of full-length hnRNPD_L-2 amyloid fibrils, which are stable, non-toxic, and bind nucleic acids, with the RNA binding domains building a solenoidal coat around them. The amyloid core consists of a single Gly/Tyr-rich and highly hydrophilic filament containing internal water channels. The architecture and activity of hnRNPD_L-2 fibrils are reminiscent of functional amyloids, our results suggesting that LGMDD3 might be a loss-of-function disease associated with impaired fibrillation. Strikingly, the fibril core matches exon 6, absent in the soluble hnRNPD_L-3 isoform. This provides structural evidence for AS controlling hnRNPD_L assembly by precisely including/skipping an amyloid exon, a mechanism that holds the potential to generate functional diversity in RNPs.

INTRODUCTION

Human heterogeneous ribonucleoprotein D-like (hnRNPDL) belongs to a class of conserved nuclear RNA-binding proteins (RBPs) that assemble with RNA to form ribonucleoproteins (RNPs). hnRNPDL acts as a transcriptional regulator and participates in the metabolism and biogenesis of mRNA¹⁻⁴.

Three isoforms of hnRNPDL are produced by alternative splicing (AS): hnRNPDL-1, hnRNPDL-2, and hnRNPDL-3⁵ (**Fig. 1a** and **Extended Data Fig. 1**). hnRNPDL-2 is a 301 residues protein and the predominant isoform in human tissues⁶. It consists of two consecutive globular RNA recognition motifs (RRM1 and RRM2), followed by a C-terminal low-complexity domain (LCD) that maps at residues ~201–285 and a nuclear localization sequence (PY-NLS) comprising residues 281-301. hnRNPDL-1 is a longer isoform of 420 amino acids containing an additional Arg-rich N-terminal LCD⁵; it is less abundant than hnRNPDL-2 and is present mainly in brain and testis⁶. hnRNPDL-3 is a shorter and minor isoform of 244 amino acids that lacks the N- and C-terminal LCDs but conserves the PY-NLS⁷.

A point mutation in hnRNPDL exon 6 causes autosomal dominant limb-girdle muscular dystrophy-3 (LGMDD3)⁸⁻¹⁰, a rare disease characterized by slowly progressive proximal muscle weakness^{11,12}. This mutation changes the conserved Asp259 in the C-terminal LCD to either Asn or His. Similarly, mutation of specific Asp residues to Asn or Val in the LCDs of hnRNPA1 and hnRNPA2 are linked to amyotrophic lateral sclerosis (ALS) and multisystem proteinopathy (MSP)^{13,14}. However, unlike ALS and MSP patients, where hnRNPA1 and hnRNPA2 accumulate in cytoplasmic inclusions in muscular fibers^{14,15}, LGMDD3 patients do not exhibit nuclear or cytoplasmic protein inclusions⁸.

Structures of the amyloid fibrils formed by hnRNPA1 and hnRNPA2 LCDs have been recently determined^{16,17}. For both fibrils, the PY-NLS was embedded within the ordered core. This led to suggest that, under physiological conditions, binding of the import receptor karyopherin- β 2 (Kap β 2) to PY-NLS¹⁸ would impede fibrillation, whereas its exposure under pathological conditions would allow amyloid formation¹⁶. However, in these studies, constructs of the LCD, alone or fused to a fluorescent protein, were used to form the fibrils, and it is unknown whether the observed assemblies would match those of the natural full-length proteins' fibrils.

Here we present the cryo-electron microscopy (cryo-EM) structure of the fibrils formed by the full-length hnRNPD-2 isoform at an overall resolution of 2.5 Å. These fibrils are stable and bind oligonucleotides, with the associated RRM domains building an exposed solenoidal coat wrapping the structured fibril. The fibrils' core is formed by a single highly hydrophilic filament encompassing LCD residues 226-276, including Asp259. Modeling suggests that disease-associated mutations at this residue may have limited impact on fibril stability. Importantly, the fibril core does not include the PY-NLS. The hnRNPD-2 fibril core precisely matches exon 6 (residues 223-279), which is alternatively spliced (AS) in a mammalian-specific manner¹⁹. These AS events are frequent in RNPs, especially at their Y/G-rich LCDs¹⁹ and they have been proposed to be a way to regulate protein function by controlling the formation of high-order assemblies²⁰; our results provide structural evidence supporting this hypothesis.

Overall, we describe the first structure of a full-length RNP in its fibrillar functional conformation, providing novel insight into the molecular bases of LGMDD3, and illustrating how AS can control RNPs assembly by including/excluding amyloidogenic exons at their LCDs.

RESULTS

hnRNPDLS localize at the nucleus and exon 6 is key for their compartmentalization.

Human hnRNPD L AS renders three naturally occurring transcripts (**Fig. 1a** and **Extended Data Fig. 1**). When transfected into a HeLa hnRNPD L knockout (KO) cell line (**Extended Data Fig. 2**), the three ectopically expressed isoforms accumulated in the nucleus (**Fig. 1b** and **Extended Data Fig. 3**), consistent with sharing a functional PY-NLS⁷. hnRNPD L-1 and hnRNPD L-2 exhibited a granulated nuclear distribution and were excluded from the nucleolar regions. In contrast, hnRNPD L-3 was homogeneously distributed in both the nucleolus and the nucleoplasm. This indicates that exon 6 (residues 223-279), shared by hnRNPD L-1 and hnRNPD L-2 and absent in hnRNPD L-3, is responsible for the differential intranuclear compartmentalization of the isoforms, delimiting regions with higher and lower protein concentration.

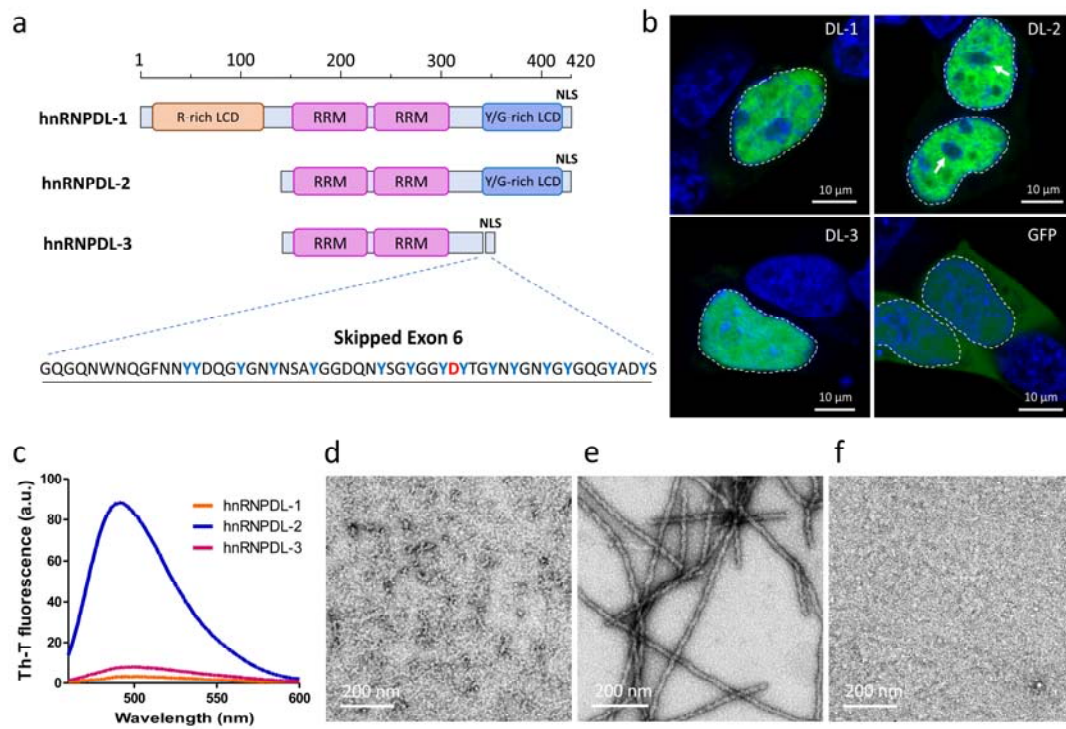


Fig. 1. hnRNPD-2 aggregates into ordered amyloid fibrils. (a) Domain organization of human hnRNPD isoforms. The RNA recognition motifs are labelled as RRM and N- and C-terminal LCDs are indicated as R-rich LCD and Y/G-rich LCD, respectively. The region and sequence of exon 6 is also indicated (EXON). PY-NLS indicates the shared NLS. b) Nuclear localization of GFP-tagged hnRNPD isoforms transiently transfected in HeLa cells. GFP distribution is shown as a control. Scale bar, 10 μ m. c) Th-T binding to hnRNPD-1 (orange line), hnRNPD-2 (blue line) and hnRNPD-3 (purple line) after incubation at 37°C, pH 7.5 and 300 mM NaCl for 2 days. d-f) Representative negative staining TEM micrographs of the incubated solutions of d) hnRNPD-1, e) hnRNPD-2 and f) hnRNPD-3. Note that only in e) individual amyloid filaments are evident.

hnRNPD-2 forms stable and non-toxic amyloid fibrils

We purified monomeric hnRNPD-1, hnRNPD-2, and hnRNPD-3 by size-exclusion chromatography. Upon incubation, hnRNPD-2 forms amyloid fibrils (**Extended Data Fig. 4**), in a reaction that is highly sensitive to the solution ionic strength (**Extended Data Fig. 5a**). At pH 7.5 and low salt (50 mM NaCl), aggregation was fast, rendering large assemblies that precipitated as the reaction progressed. In contrast, at 300 mM NaCl, the reaction exhibited characteristic sigmoidal kinetics, with the formation of highly ordered individual amyloid fibrils that strongly bind Thioflavin-T (Th-T) (**Fig. 1c** and **1e**). A further increase in ionic strength (*i.e.* 600 mM NaCl) drastically delayed and reduced hnRNPD-2 amyloid formation (**Extended Data Fig. 5a**). Thus, the fibrils formed at pH 7.5 and NaCl 300 mM were selected for structural studies. Importantly, under the same conditions, monomeric hnRNPD-1 and hnRNPD-3 did not aggregate into Th-T positive assemblies (**Fig. 1c-1f**, and **Extended Data Fig. 4**). These results point to exon 6 as the protein region responsible for amyloid

formation in hnRNPD-2. In fact, exon 6 is missing in hnRNPD-3, and the hnRNPD-1 N-terminal Arg-rich LCD may effectively counteract exon 6 amyloidogenic propensity, likely by diverting the assembly towards phase-separated condensates³.

Once formed, hnRNPD-2 fibrils are stable and insensitive to salt or temperature (**Extended Data Fig. 5b**). Remarkably, these *in vitro* formed structures are devoid of toxicity for different human cell lines up to 0,4 mg/ml (**Extended Data Fig. 6**), which is consistent with the observation that > 30 % of endogenous hnRNPD-2 accumulates into the detergent-insoluble fraction of wild type HeLa cells (**Extended Data Fig. 7**).

Cryo-EM structure determination of hnRNPD-2 amyloid fibrils.

We used cryo-EM to investigate the molecular structure of hnRNPD-2 amyloid fibrils (**Fig. 2a**). Two-dimensional (2D) classification yielded one single species of twisted fibrils. About 54,500 fibril segments from 1,114 micrographs were selected for helical reconstruction in RELION 3.1^{21,22}, allowing us to obtain a three-dimensional (3D) density map of hnRNPD-2 at an overall resolution of 2.5 Å (**Fig. 2b, Fig. 2c and Extended Data Fig. 8**). The fibril consists of a single protofilament where hnRNPD-2 subunits stack along the fibril (Z) elongation axis. The fibril forms a left-handed helix with a full pitch of ~357 Å, a helical rise of 4.82 Å, and a helical turn of -4.86°, featuring a typical cross-β amyloid structure (**Fig. 2c-2e**). A molecular model could be unambiguously built *de novo* for the fibrils core, spanning hnRNPD-2 residues 226-276 (**Fig. 2d and 2e**) and matching strikingly the region of exon 6 (residues 223-279), which is absent in the non-amyloidogenic hnRNPD-3 isoform (**Fig. 1a**). The fibril core hosts six individual β-strands (β1 to β6) connected by β-turns and loops, yielding a

sort of M-shaped fold, with approximate overall footprint of 55 Å x 40 Å (**Fig. 2c-2e**).

Data collection and refinement statistics are summarized in **Supplementary Table 1**.

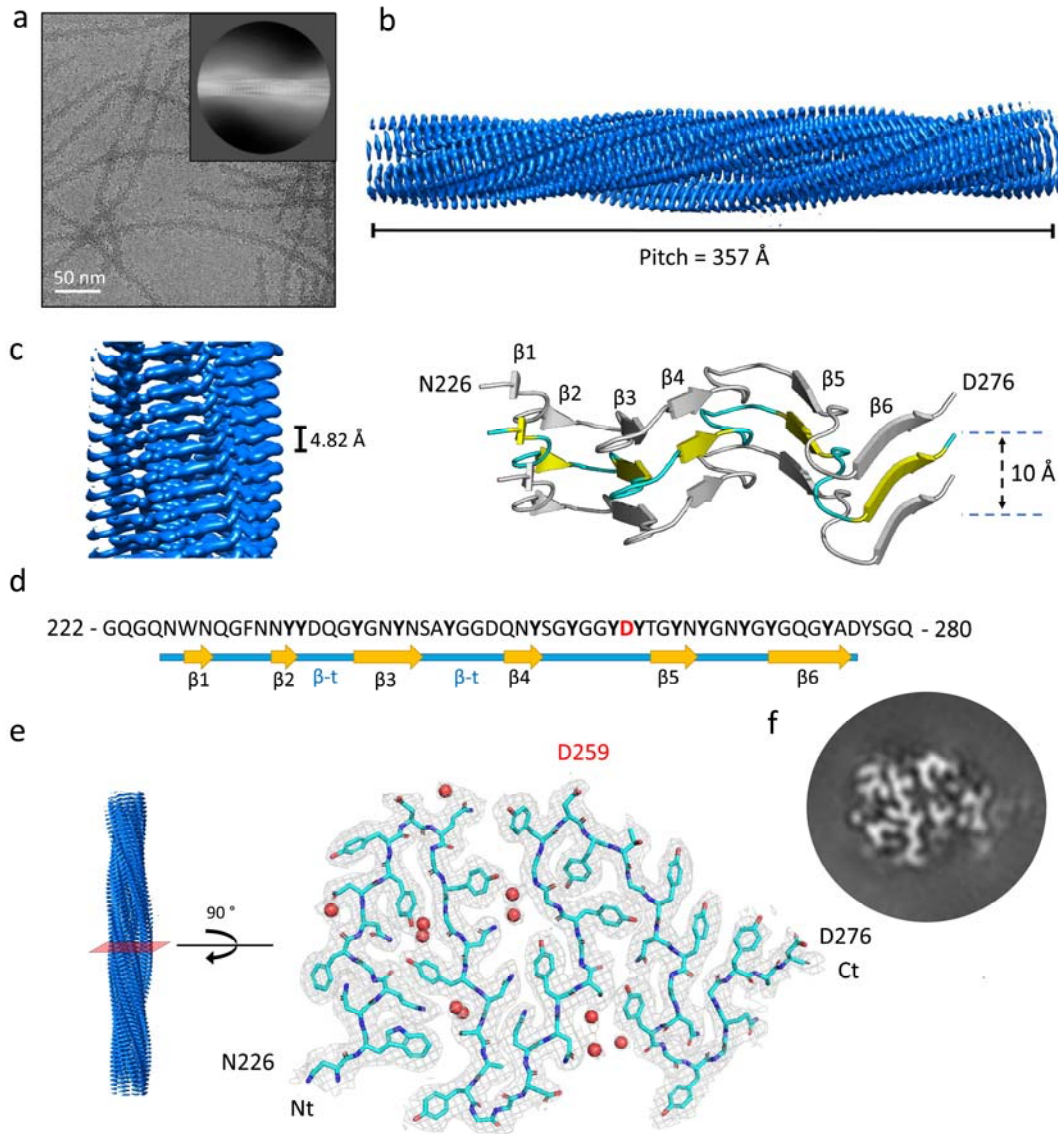


Fig. 2. Structure of hnRNPD-2 amyloid filaments. a) Cryo-electron micrograph of hnRNPD-2 filaments. The inset shows a representative reference-free 2D class average image of the hnRNPD-2 filament. b) Side view of the three-dimensional cryo-EM density map showing the structured core of an individual hnRNPD-2 filament. The fibril pitch is indicated. c) Left panel, detailed view of the cryoEM density map showing the layer packing in the hnRNPD-2 filament. Right panel, rendered side view of the secondary structure elements

accounting for three stacked rungs comprising residues N226-D276. In b) and c), the filament rise/subunit is indicated in Å. d) Sequence of hnRNPD L exon 6. The observed β -strands that build the hnRNPD L-2 fibrils amyloid core are indicated. e) cryo-EM density map of a layer of hnRNPD L-2 amyloid core. Fitting of the atomic model for residues N226-D276 is shown on top. f) Representative 3D class average image of the hnRNPD L-2 amyloid fibril.

Structural features of hnRNPD L-2 amyloid fibrils.

The core of hnRNPD L-2 amyloid fibrils, as well as exon 6, map to the C-terminal Y/G-rich LCD, comprising residues ~201–285. Accordingly, the filament amino acid composition is skewed towards Gly (27 %) and Tyr (25 %), which comprise > 50% of the residues in the structure, and neutral-polar residues, with Asn, Gln, and Ser summing to 32 %. Thus, each layer of the amyloid core is essentially stabilized through intra-subunit hydrogen bonds involving buried polar side chains, main-chain peptide groups, and ordered solvent molecules (**Fig. 3a**). The individual protein subunits pack on each other through an extensive hydrogen-bonding network holding the in-register β -strands together, in the typical cross- β arrangement (**Extended Data Fig. 9a**). Additionally, the Tyr aromatic side chains build ladders of π - π stacking interactions along the fibril elongation direction, stabilizing the inter-subunit association interface. Tyr residues are approximately equally distributed between the filament interior and the surface. The core region N-end contains two additional aromatic residues, W227 and F231, which also form individual ladders of π - π stacking interactions along the fibrillar axis; surprisingly, F231 is exposed on the fibril surface, and not buried within the core. Other hydrophobic residues such as Val, Leu, and Ile are absent, whereas they are frequently observed in the structure of pathological fibrils²³. Gly residues are dispersed throughout the sequence, with their peptide units also involved in hydrogen bonding. The main chain of each individual subunit stacked in the

fibril does not lay in a plane, with a 10 Å distance between the top of β 1 strand and the edge of the turn immediately after β 4 (**Fig. 2c**). As a result, each subunit (*i*) not only interacts with the layer directly above (*i* + 1) and below (*i* – 1), but also with layers (*i* + 3) and (*i* – 3). This is best exemplified by the formation of a strip of hydrogen bonds involving the side chains of Q237 in layer *i* and Y258 in layer *i*-3 (**Extended Data Fig. 9b** and **Fig. 9c**).

Unlike most amyloid structures, the conformation of hnRNPD-2 residues N226-D276 outlines pores that define two internal water channels. The major of such channels spans the fibril's entire length and contains additional densities that correspond to two ordered water molecules per layer, that are well-coordinated with nearby polar groups of Y239, N241, and with peptide N-atoms of G256 (**Extended Data Fig. 9d**). According to our experimental estimates, the cryo-EM density around the channel shows one of the highest resolutions within the entire fibril structure, indicating high local stability (**Extended Data Fig. 8a**). On the average, each fibril layer encloses ten ordered water molecules.

hnRNPD-2 fibrils are mainly composed of hydrophilic amino acids and display more polar surfaces than the fibrils of typical pathogenic proteins (**Extended Data Fig. 10**). The solvent-accessible surface area (SASA) of the hnRNPD-2 fibril upper layer is 3458 Å², 48 % of which is covered by polar atoms, a value that is significantly lower than in disease-associated fibrils, which display more hydrophobic end surfaces (**Supplementary Table 2**). The SASA of internal hnRNPD-2 layers is, on average, 1294 Å², reflecting a burial of 63% of the surface relative to the end solvent-exposed layers, 57 % of exposed atoms being polar. This endorses the lateral surface of the fibril with a high hydrophilic character (**Extended Data Fig. 10a**). In contrast, in the lateral surfaces of pathogenic fibrils, non-polar atoms are predominant (**Supplementary Table**

3). The proportion of exposed polar atoms in the inner layers of hnRNPD-2 is also higher than in the hnRNPA1 and hnRNPA2 LCDs fibrils. Overall, hnRNPD-2 appears to assemble into one of the most hydrophilic fibrils described so far, providing donor and acceptor groups for potential interactions.

We calculated the total solvation free energy gain per layer upon burial in the assembly, and the average solvation energy per residue of hnRNPD-2 fibrils (**Extended Data Fig. 11 and Supplementary Table 4**). These values are significantly lower than those of disease-associated fibrils, indicating that despite their irreversibility, hydrophilic hnRNPD-2 fibrils are less stable.

Importantly, in contrast to the hnRNPA1 and hnRNPA2 LCD fibrils^{16,17}, the PY-NLS in hnRNPD-2 fibrils is adjacent to, but not part of the structural core (**Extended Data Fig. 12**), suggesting that binding of Kap β 2 would not necessarily hamper fibrillation.

Disease-causative hereditary mutations in the hnRNPD-2 fibril structure.

Two missense mutations in hnRNPD exon 6, D259N and D259H, are linked to LGMDD3⁸⁻¹⁰. D259 is strictly conserved in the hnRNPD C-terminal LCD of vertebrates (**Fig. 3b**) and maps at the end of the loop connecting β 4 to β 5 in the hnRNPD-2 fibril (**Fig. 3c**). This residue is solvent exposed, with 54% of its surface accessible in the inner fibril layers, yielding a negatively charged ladder along the fibril surface (**Fig. 3c and 3d**). Apart from D259, the fibril core contains three additional exposed Asp residues, D236, D249, and D276, and no positively charged amino acid, which results in a calculated single-layer pI of 3.3 and a highly anionic patch extended along the fibrillar axis. This distinguishes hnRNPD-2 from hnRNPA1 and hnRNPA2 LCDs fibril cores, with calculated pIs of 6.0 and 8.4, respectively.

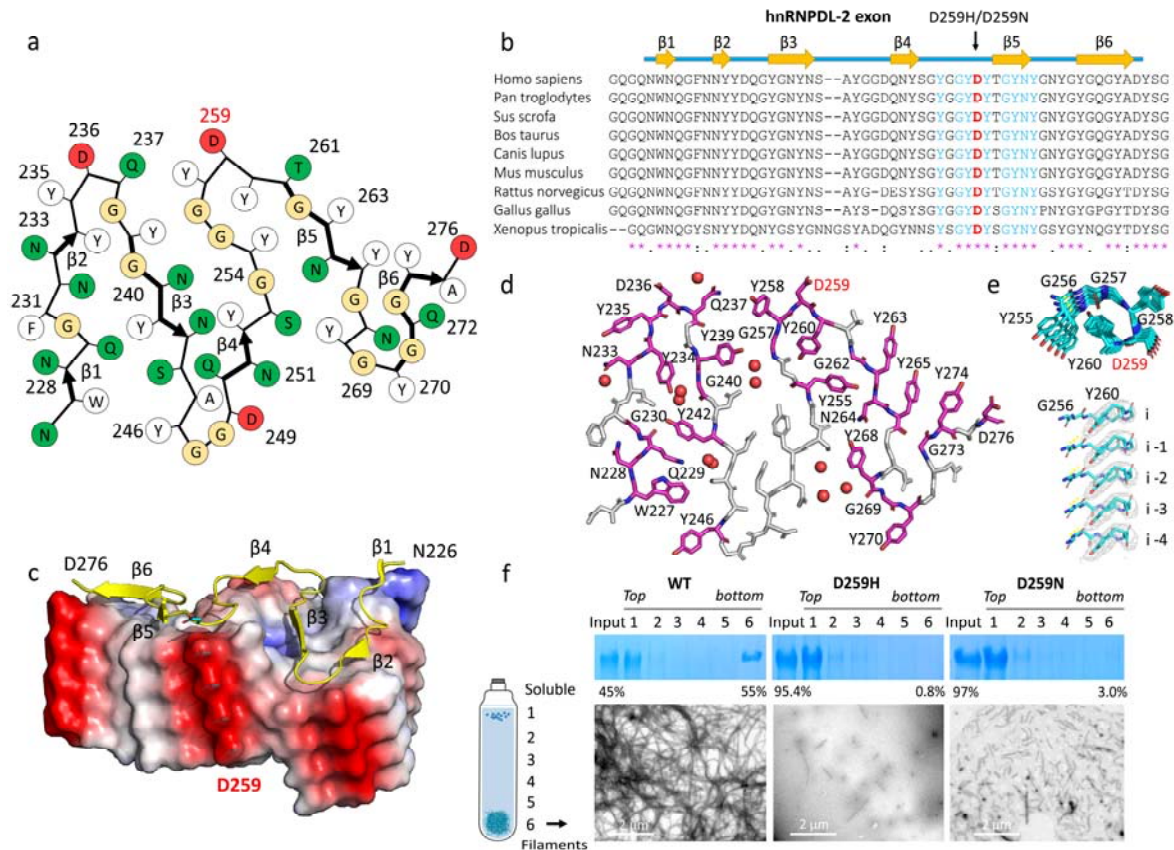


Fig. 3. Overall structure of the hnRNPD-2 fibril core. a) Schematic representation of one cross-sectional layer of the hnRNPD-2 fibril core. The location of the β -strands is indicated with thicker arrows. Polar and hydrophobic residues are colored in green and white, respectively. Glycine residues are colored in yellow. b) Sequence alignment of hnRNPD-2 orthologues showing evolutionary conservation of residues within exon 6. The mutated D259 and surrounding conserved residues are shown in red and blue, respectively. All the conserved residues are indicated with a purple asterisk below the sequence alignment. c) Surface representation showing the electrostatic surface potential distribution of the hnRNPD-2 fibril with ribbon representation of one subunit on top. Negative and positive charge potentials are colored in red and blue, respectively. The location of the six β -strands (β 1 to β 6) is indicated. The side chain of D259, mutated in LGMDD3, is shown as blue sticks, and labeled. d) Stick model of one hnRNPD-2 fibril rung showing the position of all evolutionary conserved

residues (colored pink) e) Upper panel, top view of the Y255 to Y260 segment showing the interlayer interactions between Y260 and G256. Lower panel, close-up view perpendicular to the fibril axis showing the Y260 packing and their interactions with the main chain of G256. The side chains of Y260 are shown as sticks (colored in cyan) over the cryo-EM map shown as a grey mesh. f) Effect of LGMDD3-associated mutations on hnRNPD2L-2 fibril formation. The WT hnRNPD2L and D259H and D259N mutant proteins were incubated under the same conditions (see Materials and Methods section for details) and subjected to separation on a glycerol cushion followed by SDS-PAGE analysis of the recovered fractions (top to bottom, 1–6). WT hnRNPD2L formed extensive insoluble amyloid aggregates that migrated to the bottom after centrifugation, while both disease-related mutations showed a pronounced reduction in the relative number of amyloid fibrils. Bottom panel, representative NS-TEM micrographs of amyloid filaments (bottom fraction, 6) from WT hnRNPD2L-2 and disease-related mutants D259H and D259N.

As in hnRNPD2L-2, mutation of a conserved Asp at the fibrillar core of hnRNPA2 LCD to Val (D290V) or to Asn and Val in hnRNPA1 LCD (D262N/V) are disease-associated. Virtual mutations of the respective amyloid cores^{16,17} indicated that they would render more stable fibrils by removing charge repulsions along the structure, explaining the presence of inclusions of these RNPs variants in the tissues of ALS or MSP patients¹⁴. However, this mechanism would not necessarily apply to hnRNPD2L-2; firstly, because the mutation causing more severe and earliest LGMDD3 onset, D259H, does not neutralize the negative charge but reverts it, and secondly because the D259N mutation still leaves three other Asp residues exposed to solvent, the core surface remaining highly acidic, with a calculated pI of 3.4. Indeed, modeling the impact of the D259H mutation in the hnRNPD2L-2 fibril predicts that it would be destabilizing, whereas the stabilization predicted for D259N is lower than for the D262N/V or D290V mutations in hnRNPA1 and hnRNPA2 fibrils, respectively (**Supplementary Table 5**).

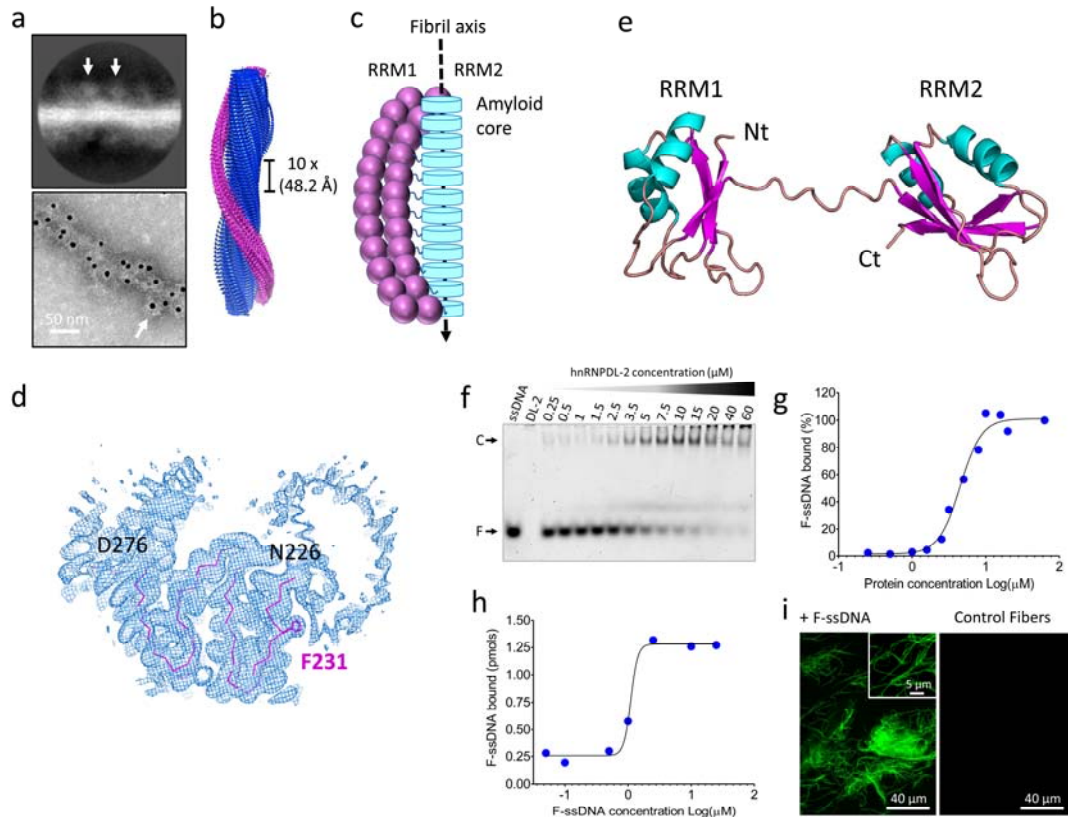
It has been recently proposed that mutations promoting the conversion of low-complexity amyloid-like kinked segments (LARKS) into steric zippers underlie irreversible aggregation and disease²⁴. The impact of the D290V mutation in hnRNPA2 LCD seems to respond to this mechanism²⁴. However, in hnRNPD1, D259H/N mutations do not increase the propensity to form steric zippers above the detection threshold (**Extended Data Fig. 13**). These analyses suggested that in LGMDD3, disease-associated variants may not necessarily act by increasing the fibril stability or amyloid propensity.

We produced and purified monomeric hnRNPD1-2 D259H and D259N variants and incubated them under the same conditions used to produce the wild-type (WT) fibrils described in the previous section. When the species in the incubated samples were separated on a glycerol cushion, it could be observed that the formation of high-molecular weight species was dramatically reduced in the disease-associated mutants, compared with WT (**Fig. 3f**). Moreover, TEM imaging of the high-molecular-weight fraction evidenced that, in contrast to the copious and long fibrils present in WT, the mutants exhibited scarce and significantly shorter fibrils (**Fig. 3f**). WT hnRNPD1-2 exhibited a strict bimodal distribution, being present only in the low- and high-molecular weight fractions, indicating that they constitute the predominant metastable states of hnRNPD1-2. These results question an amyloid origin of LGMDD3 and are consistent with the absence of mutated protein inclusions in the atrophied muscle of patients⁸.

hnRNPD1-2 amyloid fibrils bind nucleic acids

hnRNPD1-2 contains two N-terminal tandem RRM domains (RRM1 and RRM2) that are thought to be functional RNA/DNA-binding motifs²⁵ (**Fig. 1a and**

Extended Data Fig. 14). As shown in **Fig. 4a and 4b**, the RRM1s were visible in hnRNPD-2 fibrils 2D classes, when a small number of segments were averaged, as additional fuzzy globular densities around the filament core; such densities are averaged



out in the 3D reconstruction due to irregular locations of the RRM1s along the fibril. Importantly, the RRM1s surrounding the fibrillar core could be evidenced by immunogold labeling in TEM images (**Fig. 4a**). We then propose a model whereby the structured fibril core, built by exon 6 residues, is decorated by a fuzzy coat of flexible RRM1s (**Fig. 4c**)

Fig. 4. Structure of the RNA/DNA-binding domains from hnRNPD-2 and their location in the amyloid filaments. a) Top panel, representative 2D class average image showing hnRNPD-2 fibrils with an approximate solenoid coat of globular domains. The arrow indicates the location of the density assigned to the RRM1s. Bottom panel, negative-stain electron

microscopy (EM) micrographs of hnRNPD-2 filaments bound to nanogold-antibodies (~10 nm, white arrow) targeting the location of the hnRNPD-2 globular domains. The N-terminal 6xHis tag was labelled using nanogold-conjugated secondary antibodies against anti-6xHis antibodies produced in mouse. b) 3D unsharpened density map reconstruction of hnRNPD-2 fibril. Densities for the amyloid core and putative RRM domains are colored in blue and purple, respectively. c) Schematic diagram showing the proposed organization of the RRM domains around the fibril core. d) Cross-sectional view of the unsharpened cryo-EM map of hnRNPD-2 fibril, with the superimposed hnRNPD-2 amyloid core in a ribbon representation. The side chain of F231 is represented. e) Model structure of the N-terminal RNA-binding domains RRM1 and RRM2 from hnRNPD-2 generated with AlphaFold2²⁶. f) Electrophoretic mobility shift assay (EMSA) of soluble hnRNPD-2 with a Fluorescein-labelled oligonucleotide (F-ssDNA). The 7-mer ssDNA was incubated with the soluble form of hnRNPD-2 at the indicated protein concentrations. g) Binding affinity of soluble hnRNPD-2 to the 7-mer fluorescent ssDNA (F-ssDNA) determined by the EMSA assay. h) Binding of the 7-mer fluorescent ssDNA (F-ssDNA) to hnRNPD-2 amyloid filaments. i) Representative confocal microscopy image of hnRNPD-2 amyloid fibrils bound to fluorescent ssDNA (+F-ssDNA). Control fibrils without ssDNA are shown as control condition.

In this respect, it should be noted that residue F231 appears exposed to solvent, giving rise to a hydrophobic ladder along the fibril hydrophilic surface. Closer inspection of the density map, however, suggests that the N-terminal arm of the fibril core, including F231, interacts with other regions of the hnRNPD-2 assembly, likely the fuzzy RRM domains (**Fig. 4d**), that would preclude F231 entropically disfavored interaction with the solvent.

Previous studies have shown that hnRNPD is a protein that actively participates in transcription and AS regulation¹⁻⁴. To confirm that soluble hnRNPD-2 can bind nucleic acids, we performed Electrophoretic Mobility Shift Binding Assays

(EMSA). We found that the soluble protein binds to a 7-mer fluorescently labeled ssDNA oligonucleotide with an apparent dissociation constant (K_d) of $\sim 4.5 \mu\text{M}$ (**Fig. 4f** and **4g**), whereas the affinity of the protein for an equivalent RNA sequence was lower (**Extended Data Fig. 15**).

We incubated the same ssDNA oligonucleotide with preformed hnRNPD-2 amyloid fibrils. As shown in **Fig. 4h**, the fibrils significantly bind ssDNA in a concentration-dependent manner, showing a K_d of $\sim 1.1 \mu\text{M}$. This interaction was confirmed by confocal microscopy, as incubated hnRNPD-2 amyloids appeared highly fluorescent due to the incorporation of the fluorescein-labeled ssDNA (**Fig. 4i**). Thus, hnRNPD-2 amyloid fibrils retain the ability to bind nucleic acids, in keeping with a potential functional role for these self-assembled structures. We have shown that the amyloid fibril surface is strongly acidic; we therefore expect the globular RRM domains that decorate the fibril to be responsible for such activity.

DISCUSSION

hnRNPD-2 is the major hnRNPD isoform in human tissues⁶ and, as we show here, the only one forming ordered amyloid fibrils under physiologic-like conditions. This property can unequivocally be attributed to exon 6, which also accounts for the granular and heterogeneous protein distribution in the cell nucleus. Importantly, hnRNPD-1 does not form fibrils, although it phase separates under the same conditions³. Thus, hnRNPD constitutes a notable exception to the general rule considering liquid-liquid phase separation (LLPS) and amyloid formation as two interconnected phenomena^{27,28}. This might be a way to evolve two structurally different assemblies, condensates and fibrils, each associated with an hnRNPD isoform, skipping potential pathogenic transitions between these conformational states.

hnRNPD_L-2 fibrils differ from those of other RNPs

To our knowledge, the *in vitro* cryo-EM hnRNPD_L-2 fibrillar structure we report here is the first presented for a full-length RNP, since previous ones were solved departing from the LCD alone^{16,29,30} or the LCD fused to a fluorescent protein^{17,31}. Thus, we think it should better reflect the *in vivo* fibrillar packing of this protein family. Indeed, the hnRNPD_L-2 fibril structure exhibits significant differences relative to those of hnRNPA1 and hnRNPA2 LCDs, even though the respective full-length soluble forms share the same overall molecular architecture^{32,33}. In particular, the hnRNPD_L-2 amyloid core matches a vertebrate conserved exon, whereas this is not the case for hnRNPA1 and hnRNPA2 (**Extended Data Fig. 16**). Secondly, the hnRNPD_L-2 fibril is significantly more hydrophilic and acidic on its surface, a property directly impacting on the interaction with other (macro)molecules. Thirdly, hnRNPD_L-2 fibrils are irreversible and cannot be disassembled at 90° C, whereas those of hnRNPA1 and hnRNPA2 LCDs are reversible^{16,17}. This difference cannot be explained in terms of the solvation free energy of the buried cores (**Supplementary Table 4**). It might respond to the fact that our amyloid includes the complete protein, where contacts between the RRM and the core may also contribute to stability. Indeed, reversibility of hnRNPA1 and hnRNPA2 fibrils is expected since, for them, LLPS and fibril formation are interrelated, and they can potentially transition between the two states. Such connection does not apply in the case of hnRNPD_L isoforms, and because hnRNPD_L-2 does not phase separate, no back-transition to liquid droplets is possible. The equilibrium should be established between the amyloid and monomeric states, which fits well with our sedimentation assays.

A final and important difference between hnRNPA1, hnRNPA2, and hnRNPD_L-2 fibril structures is that, in the first two proteins, the PY-NLS residues that bind to

importin Kap β 2 establish stabilizing interactions and are buried inside the fibril core^{16,17}; this does not occur in hnRNPD-2, where PY-NLS is adjacent but external to the amyloid core (not mapped by density). The accessibility of PY-NLS in both monomeric and fibrillar states of hnRNPD-2 is consistent with the observation that, in humans, Kap β 2 co-localizes with both condensed and diffuse hnRNPD nuclear regions⁸. This does not exclude that Kap β 2 might still modulate hnRNPD-2 fibril formation, sterically interfering with the building of the amyloid core; however, this action would be mechanistically different from those exerted on hnRNPA1 and hnRNPA2, where Kap β 2 also acts as disaggregase³⁴.

hnRNPD-2 fibrils properties are reminiscent of functional amyloids

hnRNPD-2 fibrils share similarities with the 3D structures of functional amyloids. Like most of them³⁵⁻³⁷, hnRNPD-2 fibrils do not exhibit polymorphism. This suggests that they may represent a global free energy minimum, allowing us to speculate that the same structure would be adopted in the cell and that it represents the functional conformation of the assembled state. In contrast, pathogenic fibrils are mostly polymorphic²³. The hnRNPD-2 fibril core is hydrophilic, and the structure is stabilized by hydrogen-bonding networks, between residues, and with water molecules, together with Tyr side chains π - π stacking and dipole-dipole interactions. The eminent polar nature of the interactions that hold up the hnRNPD-2 fibril and the abundance of flexible Gly residues would allow protein chains to explore the conformational space efficiently towards the final structure, without being trapped into polymorphic local minima by stable hydrophobic interactions, as it often occurs for disease-associated fibrils²³. In addition, the hydrophilic nature of hnRNPD-2 fibril surfaces may lay

behind their observed lack of toxicity, precluding interactions with hydrophobic cellular membranes and their subsequent disruption.

hnRNPD_L-2 fibrils consist of a single filament, whereas 74 % of the available amyloid structures have two or more protofilaments³⁸. This is not surprising, since most fibril interfaces involve hydrophobic interactions between individual protofilaments³⁸ and, as said, non-polar amino acids are virtually absent in the hnRNPD_L-2 fibril core. In addition, the surrounding RRM_s coat would impair any filament-to-filament lateral association.

The structure of hnRNPD_L-2 fibrils is reminiscent of that of HET-s prion, where globular domains also hang from a single filament fibril in a solenoidal fashion³⁵. A single filament allows for the decoration of amyloid cores with regularly spaced globular domains following the helical twist of the fibril, something hardly compatible with a multi-protofilament assembly. Indeed, the fibrils of hnRNPA2¹⁷ and FUS³¹ LCDs fused to fluorescent proteins also exhibit a single protofilament, consistent with this being the preferred disposition when globular domains are adjacent to LCDs in the sequence. Instead, the fibrils of hnRNPA1 LCD alone involved two filaments¹⁶.

The most substantial evidence of hnRNPD_L-2 fibrils' functionality is their ability to bind oligonucleotides, especially ssDNA, with an affinity equal to or higher than the soluble counterpart. This indicates that the RRM domains wrapping around the structured fibrils are folded and functional.

LGMDD3-associated mutations impair hnRNPD_L-2 fibrillation

The functionality of hnRNPD_L-2 fibrils opens a new scenario for understanding the connection between D259 mutations and LGMDD3 and why, in contrast to ALS and MSP patients bearing similar mutations in hnRNPA1 and hnRNPA2 LCDs¹⁵, patients suffering this muscular dystrophy do not exhibit protein inclusions⁸. For

hnRNPA1 and hnRNPA2, theoretical calculations indicated that Asp substitutions would stabilize the fibrils^{16,17} and facilitate LARKS to steric zippers transitions²⁴. This would thermodynamically shift any potential droplet/fibril equilibrium towards the fibrillar state, reducing reversibility. For hnRNPD2, this equilibrium does not apply, and the WT fibrils are already irreversible. The same calculations indicate that D259H/N mutations do not significantly stabilize the fibril core and that no LARKS to steric zipper transition occurs. Indeed, the mutant proteins exhibit a low propensity to fibrillate, and the fibrils are shorter and less organized. This is consistent with the lack of aggregates in patients' muscular tissue⁸, suggesting that we might face a loss-of-function disease. It could be that hnRNPD2 fibrils cannot be efficiently formed in the affected muscle, and the soluble protein cannot compensate for their activity. Alternatively, inefficient fibrillation pathways of mutants might involve intermediates that are either degraded by the protein quality control machinery, decreasing the pool of active protein, or instead accumulate in myocytes, exerting toxicity. Indeed, knockdown of zebrafish hnRNPD2 (85 % identity with human hnRNPD2) using antisense oligonucleotides resulted in dose-dependent disorganization of myofibers, causing body shape defects and restricted and uncoordinated movements, which is consistent with a loss-of-function myopathy⁸.

Our results align with the recent evidence that even for hnRNPA1, disease manifestation is not always associated with increased fibrillation, and variants with LCD mutations displaying a low ability to form fibrils are also pathogenic³⁹. Interestingly, these hnRNPA1 variants cause vacuolar rimmed myopathy, histologically similar to LGMD3. This suggests that RNP-associated diseases might not respond to a unique molecular mechanism, but somewhat different sequential/structural perturbations might elicit cellular dysfunction and degeneration.

Mammalian-specific AS controls hnRNPD assembly

The most intriguing and unique feature of the hnRNPD-2 fibril is the perfect overlap between the amyloid core and exon 6. AS patterns have diverged rapidly during evolution, with exons that were ancestrally constitutive in vertebrates evolving to become alternatively spliced in mammals, expanding the regulatory complexity of this lineage⁴⁰. These evolutionary changes impact all members of the hnRNP D family, to which hnRNPD belongs. Accordingly, the exon 6 sequence is conserved among vertebrates (**Fig. 3b**), an additional evidence of its functionality, and alternatively spliced in mammals^{7,19}. Mammalian-specific AS is especially frequent at the Y/G-rich LCDs of RNPs¹⁹, suggesting that regulation of the number of GY motifs in these regions confers fitness benefit²⁰. Furthermore, elimination of these repeats through exon skipping results in dominant-negative RNPs that bind nucleic acids but cannot form multimeric complexes through Tyr-dependent interactions, which significantly modifies their gene regulation activity and nuclear patterning²⁰. This differential behavior is often attributed to the longer isoforms' ability to undergo LLPS²⁰. However, in proteins of the Rbfox family, the splicing activity is contingent on the formation of Th-T positive fibrous structures, which is mediated by its LCD in a Tyr-dependent manner⁴¹, although it is unknown if Rbfox fibers correspond to cross- β amyloids. Here we provide high-resolution structural evidence supporting the role of mammalian specific-AS in controlling the assembly and nuclear distribution of RNPs, which in the particular case of hnRNPD occurs by precisely including/skipping a conserved amyloid-prone exon.

Conclusion

Overall, this work presents the first detailed cryo-EM structure of a full-length RNP in its fibrillar functional conformation. The structure of hnRNPD-2 exon 6 in its

amyloid form provides novel insights into the molecular bases of LGMDD3 and the mechanism of AS-controlled RNPs assembly in mammals.

MATERIALS AND METHODS

Cell culture, Plasmids and Cell lines

The human HeLa (ATCC CCL-2) and SH-SY5Y (ATCC CRL-2266) cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) or minimum essential medium α (MEM- α) medium, respectively. Media were supplemented with 10% (v/v) Fetal Bovine Serum (FBS). Both cell lines were grown under a highly humidified atmosphere of 95% air with 5% CO₂ at 37 °C. The HeLa hnRNPD L KO cell line was generated as described elsewhere³.

For recombinant protein expression, the genes encoding the three isoforms of hnRNPD L (hnRNPD L-1, hnRNPD L-2 and hnRNPD L-3) were inserted into pETite (Lucigen corporation) vector with a His-SUMO N-terminal tag. For subcellular localization experiments, the same genes were cloned into pEGFP-C3 (Clontech). In all the cases, the correctness of the DNA sequence was verified by sequencing.

Cell transfection and immunoblotting

Adherent cells were transfected with linear polyethylenimine (PEI; Polysciences) in a 1:3 DNA:PEI ratio. Cells were collected after 48 h, lysed in M-PER mammalian protein extraction reagent (Thermo-Fisher) with 1/1000 of the EDTA-free protease inhibitor cocktail Set III (Calbiochem), and centrifuged for 30 min at 15,000 \times g at 4°C. The soluble and insoluble fractions were analyzed by SDS-PAGE and immunoblotting onto PVDF membranes (EMD Millipore) were performed using standard protocols⁴².

hnRNPD L proteins were detected using different primary antibodies (anti-hnRNPD L antibody, HPA056820 from Sigma-Aldrich or with an anti-GFP monoclonal antibody, sc-9996 from Santa Cruz Biotechnology). The primary antibodies were detected with HRP-labelled secondary antibodies, followed by chemiluminescence detection using Immobilon Forte Western HRP substrate (Sigma-Aldrich).

Protein expression and purification

Protein expression was performed in *E. coli* BL21(DE3) cells, induced with 0.5 mM IPTG at a $OD_{600}=0.5$. After incubation for 3 hours at 37°C and 250 rpm, the cells were harvested by centrifugation for 15 min at 4000 $\times g$. Cell pellets were resuspended in Binding Buffer (50 mM HEPES, 1 M NaCl, 5% glycerol and 20 mM imidazole, pH 7.5), lysed by sonication, and centrifugated at 30000 $\times g$ for 30 min at 4°C. The supernatant was filtered through a 0.45 μm filter and loaded into a HisTrapTM FF Ni-column equilibrated with Binding Buffer. The bound protein was eluted with an imidazole gradient starting from 0 to 100% of the Elution Buffer (50 mM HEPES, 1 M NaCl, 5% glycerol, 500 mM imidazole, pH 7.5). Afterwards, fractions containing purified proteins were pooled and loaded into a HiLoadTM 26/600 SuperdexTM 75 pg column equilibrated with a 50 mM HEPES pH 7.5 buffer containing 1M NaCl and 5% glycerol. Finally, the proteins were concentrated using a 10 K Amicon (Merck-Millipore), flash-frozen in liquid nitrogen and stored at -80°C until use.

Fibril formation and *in vitro* aggregation kinetics

For aggregation experiments, the purified proteins corresponding to the three different hnRNPD L isoforms were loaded into a PD-10 desalting column SephadexTM G-25 M and the buffer was exchanged to 50 mM HEPES, 150 mM NaCl, pH 7.5. Samples were

diluted to a final protein concentration of 50 μ M in 50 mM HEPES, 300 mM NaCl, pH 7.5. Then, the aggregation reactions were incubated at 37 $^{\circ}$ C for 2 days with 600 rpm agitation in a sealed Eppendorf.

Th-T binding to hnRNPD L aggregates was measured by recording Th-T fluorescence in the range between 460 and 600 nm after excitation with wavelength of 445 nm using a Jasco FP-8200 spectrofluorimeter. The final Th-T and protein concentrations were 25 μ M and 10 μ M, respectively. All the samples were diluted in 50 mM HEPES buffer at pH 7.5 with 300 mM NaCl and this same buffer alone was used as a control.

The aggregation kinetics of hnRNPD L isoforms were monitored in 96 well plates by following increments in Th-T signal of 25 μ M protein samples diluted in 50 mM HEPES, 300 mM NaCl, pH 7.5. Plates were incubated at 37 $^{\circ}$ C under constant shaking (100 rpm) using a Spark (TECAN) Fluorescence microplate reader. The Th-T fluorescence of each well was measured every 30 min by exciting with a 445 nm filter and collecting the emission with a 480–510 nm filter. Fluorescence emission of all the proteins in the absence of Th-T and the signal of the buffers alone with Th-T were determined as control conditions. To test the effect of NaCl on the aggregation kinetics of hnRNPD L-2, aggregation reactions with different NaCl concentrations (50, 300 and 600 mM) were also prepared.

Ultracentrifugation

To perform ultracentrifugation experiments, the samples were centrifuged at 120,000 \times g for 2 h at 4 $^{\circ}$ C in a 20% (v/v) glycerol cushion. After centrifugation, six fractions (70 μ l each) were collected from top to bottom, without disturbing the layers, and analyzed by SDS-PAGE. The total protein concentration of each fraction was

determined by Bradford using a commercial Coomassie Protein assay reagent (Thermo Fisher).

Immunolabelling and negative staining electron microscopy

Aliquots of hnRNPD-2 amyloid fibrils were mixed with an anti-6xHis (MA1-21315, Sigma-Aldrich) primary antibody (1:10 dilution) in 50 mM HEPES, 150 mM NaCl, pH 7.5 buffer and the sample was incubated 18 h at 4°C. After incubation, an anti-mouse 10 nm colloidal gold linked secondary antibody (Cell Signaling Technology) was added to the samples and incubated for 1 hour at room temperature. For negative staining transmission electron microscopy (NS-TEM), 5 μ l of each sample was incubated on a EMR 400 mesh carbon-coated copper grids (Micro to Nano Innovative Microscopy Supplies) for 5 minutes. After incubation, the grids were washed with 10 μ l MQ water and stained for 1 min with 5 μ l of 2% (w/v) uranyl acetate. The excess solutions from each step were removed with filter paper. Each grid was allowed to dry before inspection using a JEOL JEM-1400 Electron Microscope operating at 120 kV with a CCD GATAN 794 MSC 600HP camera. 3–5 micrographs were recorded for each sample at two different nominal magnifications (*i.e.* $\times 6000$ and $\times 10000$) and an estimated defocus of about $\pm 1-4 \mu\text{m}$.

Cryo-EM sample preparation and data collection

For cryo-EM, sample vitrification was carried out using a Mark IV Vitrobot (Thermo Fisher Scientific). 3 μ l hnRNPD-2 amyloid fibrils diluted in MQ water at a final concentration of 0.25 mg/mL were applied to a C-Flat 1.2/1.3-3Cu-T50 grid (Protochips) previously glow-discharged at 30 mA for 30 seconds in a GloQube (Quorum Technologies). Sample was incubated on grid for 60 seconds at 4°C and 100%

humidity, blotted and plunge-frozen into liquid ethane. Vitrified samples were transferred to a Talos Arctica transmission electron microscope (Thermo Fisher Scientific) operated at 200 kV and equipped with a Falcon 3 direct electron detector (Thermo Fisher Scientific). A total of 1114 movies were collected using EPU 2.8 (Thermo Fisher Scientific) in electron counting mode with an applied dose of $40 \text{ e}^-/\text{\AA}^2$ divided in 40 frames at a magnification of 120kX. All the micrographs were acquired with a pixel size of $0.889 \text{ \AA}/\text{pixel}$ and a defocus range of -1.0 to -2.2 \mu m .

Helical reconstruction

Best movies (without the presence of artifacts, crystalline ice, severe astigmatism or with obvious drift) were imported in RELION 3.1 for further processing following a helical reconstruction pipeline²². All the movies were motion-corrected and dose-weighted using MOTIONCOR2⁴³. Contrast transfer function (CTF) estimation was performed on aligned, unweighted sum power spectra every $4 \text{ e}^-/\text{\AA}^2$ using CTFFIND4⁴⁴. Micrographs with a resolution estimate of 5 \AA or better were selected for further analysis. Fibrils were manually picked, and the segments were successively extracted using a box size of 380 pixel and inter-box distance of 33.2 \AA , yielding a total of 158493 segments. Reference-free 2D classification was performed to identify homogeneous segments for further processing.

An initial model was generated using a single large class average similarly as described elsewhere⁴⁵. In brief, the cross-over distance was measured from the class average image to be $\sim 179 \text{ \AA}$. The initial model was then re-scaled and re-windowed to match the unbinned particles with a 380-pixel box and low-pass filtered to 10 \AA . A 3D classification with four classes, a regularization value of $T_{\text{reg}} = 4$ and imposing a helical

rise of 4.82 \AA and a helical twist of -4.86° was used to select the best class containing 54490 particles.

Particles were aligned with a 3D auto-refine job using a 10 \AA low-pass filtered map from the previous 3D classification, considering the 30% central part of the box, and refining the helical twist and rise. Bayesian polishing^{46,47} and CTF refinement⁴⁶ were performed to increase the resolution of the final reconstruction. Finally, the refined 3D reconstructions were sharpened using the standard post-processing procedures in RELION 3.1. Based on the gold-standard Fourier shell correlation = 0.143 criteria, the best map for the hnRNPD-2 fibril showed a final resolution of 2.5 \AA . All the details of data acquisition and processing are given in **Table S1**. The hnRNPD-2 structure, and cryo-EM map have been deposited in the Protein Data Bank (PDB ID: 7ZIR) and Electron Microscopy Data Bank (EMDB-14738), respectively. Raw images and raw processing data are available at the Electron Microscopy Public Image Archive - EMBL-EBI (EMPIAR-11064).

Model building and refinement

The hnRNPD-2 amyloid fibril model was manually built into the 2.5 \AA sharpened map of the hnRNPD-2 fibril using COOT. In brief, an initial polyalanine model was incorporated into the 2.5 \AA sharpened map of the hnRNPD-2 filament. Then, Ala residues were mutated to Tyr where the electron density was clear enough. Based on the location of Tyr and spacing pattern, the comprising residues 226 to 276 of hnRNPD-2 were successfully assigned and modeled. Real-space refinement was accomplished with the phenix.real_space_refine module from Phenix⁴⁸. Molecular graphics and structural analyses were performed with Pymol⁴⁹ and ChimeraX⁵⁰. Statistics for the final model

are provided in **Supplementary Table 1** and the corresponding atomic coordinates have been deposited in the Protein Data Bank (PDB accession number: 7ZIR).

Fibril stability, solvent accessible surface area and solvation free energy gain calculations

The free energy of mutation calculations were determined using the module `ddg_monomer` from Rosetta⁵¹. Delta delta G ($\Delta\Delta G$) energy values for wt hnRNPD-2 and mutants (D259H, D259N and D259V) fibril structures were determined following a standard high-resolution protocol as described elsewhere¹⁶. In brief, we first performed energy minimization on the hnRNPD-2 fibril structure containing five successive rungs. The resultant restraint file from this step was used in the subsequent processing. Standard Van der Waals and solvation energy parameters (*i.e.* cut off value of 9 Å) were applied. For comparative purposes, the $\Delta\Delta G$ values for wt and relevant hnRNPA1 or hnRNPA2 mutant fibrils were also determined using the same Rosetta protocol.

The exposed surfaces of hnRNPD-2, hnRNPA2, hnRNPA1, hSAA1, Ab-42 and α -synuclein fibrillar structures were calculated using PDBePISA⁵² as the difference between the total and the buried surface. The percentage of exposed polar surface was calculated using the total and apolar surface values predicted by GetArea⁵³.

The ΔG^{int} values were calculated for each fiber and for each residue within the amyloid cores using PDBePISA⁵². ΔG^{int} indicates the solvation free energy gain for the structure upon assembly formation (in kcal/mol).

***In vivo* cell imaging**

For *in vivo* cell imaging experiments, adherent HeLa cells were grown on 35 mm glass-bottom culture dishes at a density of 1×10^5 cells per dish. After 24 h incubation, the

cells were transiently transfected with GFP-tagged versions of the hnRNPDL-1/-2/-3 isoforms and linear polyethylenimine (PEI; Polysciences, Eppelheim, Germany) in a 1:3 DNA:PEI ratio. After 24 h incubation, the cells were washed with fresh medium (2×1 mL), and then cell nuclei were stained with Hoechst (Invitrogen) for 10 min at 37°C and 5% CO₂. *In vivo* cell imaging was performed at 37°C using a Leica TCS SP5 confocal microscope and a 63 ×1.4 numerical aperture Plan Apochromat oil-immersion lens. All the confocal images were processed using Bitplane Imaris 7.2.1 software.

Gel Electrophoretic Mobility Shift Binding Assays (EMSA)

The ability of hnRNPDL-2 amyloid fibrils to bind RNA/ssDNA was studied by performing a Gel Electrophoretic Mobility Shift Binding Assays (EMSA). We prepared 10 µl reactions containing 50 nM of Fluorescein-labeled RNA/ssDNA (F-GACUAGC) and increasing amounts of soluble hnRNPDL-2 or amyloid fibrils. The samples were incubated for 24 h at RT before the complexes were resolved on an 8% polyacrylamide gel at a constant voltage. After electrophoresis, the gel was imaged using a ChemiDoc™ MP Imaging System (Bio-Rad).

Fluorescent RNA/DNA-binding Assays

The ability of hnRNPDL-2 amyloid fibrils to bind ssDNA was studied by measuring the fluorescence of the fibrils bound to Fluorescein-labelled ssDNA using a Jasco FP-8200 Spectrofluorometer (Jasco Corporation) and a fluorescence microscope (Leica Microsystems).

For the fluorometric assays, 20 µl reactions containing 10 µM of hnRNPDL-2 amyloid fibrils and increasing amounts of Fluorescein-labeled ssDNA (F-GACUAGC) at the following concentrations (0.05, 0.1, 0.5, 1, 2.5, 5, 10, 25, 50 and 100 µM) were

prepared. After incubation for 24 h at RT, the samples were centrifuged at $16,000 \times g$ for 30 min. The pellet containing the hnRNPD-2 amyloid fibrils bound to the ssDNA was washed with 50 μ l of 50 mM HEPES buffer (pH 7.4) with 150 mM NaCl and sonicated. Fluorescence emission of the ssDNA bound to fibrils was measured at an excitation and emission wavelength of 480 nm and 520 nm, respectively. Control samples with only fluorescent ssDNA were used as controls conditions.

For confocal microscopy analysis, 20 μ l reactions with hnRNPD-2 amyloid fibrils at a concentration of 20 μ M were incubated for 24 h at RT in the presence of 30 μ M of Fluorescein-labelled ssDNA (F-GACUAGC). After centrifugation at $16,000 \times g$ for 40 min, the precipitated fraction was placed on a microscope slide and sealed. Confocal fluorescence images were obtained with Leica SP5 confocal microscope (Leica microsystems). hnRNPD-2 amyloid fibrils and labeled ssDNA alone were imaged as control samples.

Cytotoxicity assays

The cytotoxicity of the hnRNPD-2 amyloid fibers toward SH-SY5Y and HeLa cells was evaluated using a resazurin-based assay⁵⁴. Briefly, the cells were seeded in 96-well plates at a concentration of 3×10^3 cells per well and incubated for 24 h. Afterwards, the cells were treated with different hnRNPD-2 fiber concentrations ranging from 0.004 to 0.4 mg/mL (w/v). After 48 h of incubation, aliquots of 10 μ L of the PrestoBlue™ (Thermo Fisher) cell viability reagent solution were added to each well. After 1 h incubation at 37 °C in the absence of light and in the presence of a highly humidified atmosphere of 95% air with 5% CO₂, fluorescence emission (ex/em = 531/572 nm) of each well was measured using a fluorescence microplate reader

(PerkinElmer Victor 3V). Cell cytotoxicity was determined in terms of cell growth inhibition in treated samples and expressed as percent of the control condition.

ACKNOWLEDGEMENTS

This work was funded by European Union's Horizon 2020 research and innovation programme under GA 952334 (PhasAGE), by the Spanish Ministry of Science and Innovation (PID2019-105017RB-I00) to S.V. and by ICREA, ICREA-Academia 2015 and 2020 to S.V. J.G-P was supported by the Spanish Ministry of Science and Innovation with a Juan de la Cierva Incorporación IJC2019-041039-I. M.G-G was supported by the Spanish Ministry of Science and Innovation via doctoral grants, FPU16/02465. S.R. was funded by Fondazione ARISLA (project TDP- 43- STRUCT) and by Ricerca Corrente funding from Italian Ministry of Health to IRCCS Policlinico San Donato. The authors thank the Microscopy Service, Universitat Autònoma de Barcelona, for their exceptional technical support in electron microscopy and confocal imaging, and the University of Milan Unitech NOLIMITS Center for granting access to the cryoEM facility. We thank Prof. J. Paul Taylor for providing the hnRNPD KO HeLa cell line.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

J.G-P., M.B., S.R. and S.V. designed the conceptual framework of the study and experiments. J.G-P, A.B-N., M.G-G and C.V. performed the experiments and

contributed to data acquisition, data analysis, and preparation of manuscript figures. A.C-S., J.G.-P., and M.G-G. contributed to cryo-EM data acquisition and cryo-EM structural characterization of the amyloid fibrils. J.G-P. and S.V. wrote the manuscript with contributions and comments from all authors.

REFERENCES

1. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321 (1993).
2. Li, R.Z. et al. hnRNPDL extensively regulates transcription and alternative splicing. *Gene* **687**, 125-134 (2019).
3. Batlle, C. et al. hnRNPDL Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. *Cell Rep* **30**, 1117-1128 e5 (2020).
4. Kemmerer, K., Fischer, S. & Weigand, J.E. Auto- and cross-regulation of the hnRNPs D and DL. *RNA* **24**, 324-331 (2018).
5. Kamei, D., Tsuchiya, N., Yamazaki, M., Meguro, H. & Yamada, M. Two forms of expression and genomic structure of the human heterogeneous nuclear ribonucleoprotein D-like JKTBP gene (HNRPD). *Gene* **228**, 13-22 (1999).
6. Akagi, T. et al. Molecular characterization of a mouse heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its tissue-specific expression. *Gene* **245**, 267-73 (2000).
7. Kawamura, H. et al. Identification of the nucleocytoplasmic shuttling sequence of heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its interaction with mRNA. *J Biol Chem* **277**, 2732-9 (2002).
8. Vieira, N.M. et al. A defect in the RNA-processing protein HNRPD causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum Mol Genet* **23**, 4103-10 (2014).
9. Starling, A., Kok, F., Passos-Bueno, M.R., Vainzof, M. & Zatz, M. A new form of autosomal dominant limb-girdle muscular dystrophy (LGMD1G) with progressive fingers and toes flexion limitation maps to chromosome 4p21. *Eur J Hum Genet* **12**, 1033-40 (2004).
10. Sun, Y. et al. Limb girdle muscular dystrophy D3 HNRNPDL related in a Chinese family with distal muscle weakness caused by a mutation in the prion-like domain. *J Neurol* **266**, 498-506 (2019).
11. Liewluck, T. & Milone, M. Untangling the complexity of limb-girdle muscular dystrophies. *Muscle Nerve* **58**, 167-177 (2018).
12. Batlle, C. & Ventura, S. Prion-like domain disease-causing mutations and misregulation of alternative splicing relevance in limb-girdle muscular dystrophy (LGMD) 1G. *Neural Regen Res* **15**, 2239-2240 (2020).
13. Harrison, A.F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J* **474**, 1417-1438 (2017).
14. Kim, H.J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467-73 (2013).
15. Purice, M.D. & Taylor, J.P. Linking hnRNP Function to ALS and FTD Pathology. *Front Neurosci* **12**, 326 (2018).

16. Sun, Y. et al. The nuclear localization sequence mediates hnRNP1 amyloid fibril formation revealed by cryoEM structure. *Nat Commun* **11**, 6349 (2020).
17. Lu, J. et al. CryoEM structure of the low-complexity domain of hnRNP2 and its conversion to pathogenic amyloid. *Nat Commun* **11**, 4090 (2020).
18. Xu, D., Farmer, A. & Chook, Y.M. Recognition of nuclear targeting signals by Karyopherin-beta proteins. *Curr Opin Struct Biol* **20**, 782-90 (2010).
19. Buljan, M. et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell* **46**, 871-83 (2012).
20. Guerousov, S. et al. Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* **170**, 324-339 e23 (2017).
21. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).
22. He, S. & Scheres, S.H.W. Helical reconstruction in RELION. *J Struct Biol* **198**, 163-176 (2017).
23. Sawaya, M.R., Hughes, M.P., Rodriguez, J.A., Riek, R. & Eisenberg, D.S. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857-4873 (2021).
24. Murray, K.A. et al. Identifying amyloid-related diseases by mapping mutations in low-complexity protein domains to pathologies. *Nat Struct Mol Biol* (2022).
25. Tsuchiya, N., Kamei, D., Takano, A., Matsui, T. & Yamada, M. Cloning and characterization of a cDNA encoding a novel heterogeneous nuclear ribonucleoprotein-like protein and its expression in myeloid leukemia cells. *J Biochem* **123**, 499-507 (1998).
26. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
27. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-33 (2015).
28. Patel, A. et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-77 (2015).
29. Murray, D.T. 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).
30. Li, Q., Babinchak, W.M. & Surewicz, W.K. Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat Commun* **12**, 1620 (2021).
31. Sun, Y. et al. Molecular structure of an amyloid fibril formed by FUS low-complexity domain *iScience* **25**(2021).
32. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851-67 (2016).
33. Krecic, A.M. & Swanson, M.S. hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-71 (1999).
34. Guo, L. et al. Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. *Cell* **173**, 677-692 e20 (2018).
35. Wasmer, C. et al. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523-6 (2008).
36. Seuring, C. et al. The three-dimensional structure of human beta-endorphin amyloid fibrils. *Nat Struct Mol Biol* **27**, 1178-1184 (2020).
37. Wu, X. et al. The structure of a minimum amyloid fibril core formed by necroptosis-mediating RHIM of human RIPK3. *Proc Natl Acad Sci U S A* **118**(2021).
38. Louros, N., van der Kant, R., Schymkowitz, J. & Rousseau, F. StAMP-DB: A platform for structures of polymorphic amyloid fibril cores. *Bioinformatics* (2022).
39. Beijer, D. et al. Characterization of HNRNP1 mutations defines diversity in pathogenic mechanisms and clinical presentation. *JCI Insight* **6**(2021).

40. Barbosa-Morais, N.L., Carmo-Fonseca, M. & Aparicio, S. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res* **16**, 66-77 (2006).
41. Ying, Y. et al. Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* **170**, 312-323 e10 (2017).
42. Battle, C. et al. MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. *Commun Biol* **4**, 414 (2021).
43. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
44. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).
45. Swuec, P. et al. Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat Commun* **10**, 1269 (2019).
46. Scheres, S.H.W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94-101 (2020).
47. Zivanov, J., Nakane, T. & Scheres, S.H.W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17 (2019).
48. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
49. Schrödinger, L. The PyMOL Molecular Graphics System. 2.0 edn.
50. Pettersen, E.F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).
51. Kellogg, E.H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830-8 (2011).
52. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
53. Fraczkiwicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comp Chem* **19**, 319-333 (1998).
54. Garcia-Pardo, J. et al. Bioinspired Theranostic Coordination Polymer Nanoparticles for Intranasal Dopamine Replacement in Parkinson's Disease. *ACS Nano* **15**, 8592-8609 (2021).