

ARTICLE

<https://doi.org/10.1038/s41467-019-10355-1>

OPEN

Tau local structure shields an amyloid-forming motif and controls aggregation propensity

Dailu Chen^{1,2,6}, Kenneth W. Drombosky^{1,6}, Zhiqiang Hou¹, Levent Sari^{3,4}, Omar M. Kashmer¹, Bryan D. Ryder^{1,2}, Valerie A. Perez^{1,2}, DaNae R. Woodard¹, Milo M. Lin^{3,4}, Marc I. Diamond¹ & Lukasz A. Joachimiak^{1,5}

Tauopathies are neurodegenerative diseases characterized by intracellular amyloid deposits of tau protein. Missense mutations in the tau gene (*MAPT*) correlate with aggregation propensity and cause dominantly inherited tauopathies, but their biophysical mechanism driving amyloid formation is poorly understood. Many disease-associated mutations localize within tau's repeat domain at inter-repeat interfaces proximal to amyloidogenic sequences, such as ³⁰⁶VQIVYK³¹¹. We use cross-linking mass spectrometry, recombinant protein and synthetic peptide systems, in silico modeling, and cell models to conclude that the aggregation-prone ³⁰⁶VQIVYK³¹¹ motif forms metastable compact structures with its upstream sequence that modulates aggregation propensity. We report that disease-associated mutations, isomerization of a critical proline, or alternative splicing are all sufficient to destabilize this local structure and trigger spontaneous aggregation. These findings provide a biophysical framework to explain the basis of early conformational changes that may underlie genetic and sporadic tau pathogenesis.

¹Center for Alzheimer's and Neurodegenerative Diseases, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ²Molecular Biophysics Graduate Program, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ³Green Center for Molecular, Computational and Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ⁴Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ⁵Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ⁶These authors contributed equally: Dailu Chen, Kenneth W. Drombosky. Correspondence and requests for materials should be addressed to L.A.J. (email: Lukasz.Joachimiak@utsouthwestern.edu)

Tauopathies comprise a group of over 20 neurodegenerative diseases in which tau protein aggregates in neurons and glia. Tau aggregation correlates strongly with the degree of dementia and neurodegeneration, especially in Alzheimer's Disease. The mechanisms by which disease-associated mutations, alternative splicing, or other events promote aggregation and pathology are not well understood. Understanding the molecular basis of tau aggregation could greatly improve diagnosis and treatment of tauopathies.

The N-terminal ~200 and C-terminal ~80 residues of tau are largely disordered, rendering this system refractory to high-resolution studies using structural biology methods¹. In contrast, the tau repeat domain (tau RD), which spans residues 243–365, is predicted to be more structured², forms the core of amyloid fibrils³, and is the minimal region to propagate tau prion strains⁴. Tau RD contains an amyloid motif (³⁰⁶VQIVYK³¹¹) (Fig. 1a) that is central to conversion between the soluble and insoluble states, as it mediates self-assembly, drives amyloid formation *in vitro*⁵ and promotes pathology *in vivo*⁶. Nuclear magnetic resonance (NMR) experiments on tau indicate that in solution the ³⁰⁶VQIVYK³¹¹ motif adopts a β -strand conformation^{2,7}. Recent cryo-electron microscopy (cryo-EM) studies of tau patient-derived fibrils have shown that ³⁰⁶VQIVYK³¹¹ mediates important contacts in these structures^{3,8}. Despite these structural studies, it is not clear how native tau avoids aggregation, nor is it clear how tau transitions from a soluble state to an aggregated assembly.

Polyanions such as heparin, nucleic acids, and arachidonic acid are commonly used to induce tau aggregation *in vitro*^{9–11}. Solution NMR experiments mapped the tau-heparin binding site to repeat 2 just prior to the ³⁰⁶VQIVYK³¹¹ motif, but how this binding event modulates tau aggregation remains unclear¹². Double electron–electron resonance experiments indicated an expansion of this region upon heparin binding⁹. Cryo-EM structures also suggested an extended conformation of tau when bound to tubulin¹³. Other work mapping the recruitment of molecular chaperones to tau indicated that many chaperones, including Hsp40, Hsp70, and Hsp90, localize around ³⁰⁶VQIVYK³¹¹¹⁴. Furthermore, unfolding of tau RD appeared to promote chaperone binding to the amyloid motif, suggesting that local conformational changes may help recruit factors to limit aggregation¹⁵. Recent data from our group indicated that soluble monomeric tau exists in at least two conformational ensembles: inert monomer (M_i), which does not spontaneously self-assemble, and seed-competent monomer (M_s), which spontaneously self-assembles into amyloid¹⁶. M_s itself adopts multiple stable structures that encode different tau prion strains¹⁷, which are unique amyloid assemblies that faithfully replicate in living systems. Based on extrapolations, the existence of an aggregation-prone monomer of tau had been previously proposed^{18,19} but our study was the first to biochemically isolate and characterize this species¹⁶. Different forms of M_s have been purified from recombinant protein, and tauopathy brain lysates^{16,17}. Using multiple low-resolution structural methods, we have mapped critical structural changes that differentiate M_i from M_s to near the ³⁰⁶VQIVYK³¹¹ motif and indicated that the repeat two and three region in tau is extended in M_s , which exposes the ³⁰⁶VQIVYK³¹¹ motif¹⁶. In contrast, intramolecular disulfide bridge between two native cysteines that flank ³⁰⁶VQIVYK³¹¹ in tau RD is predicted to form a local structure that is incompatible with the formation of amyloid²⁰. Thus, conformational changes surrounding the ³⁰⁶VQIVYK³¹¹ amyloid motif appear critical to modulate aggregation propensity. A fragment of tau RD in complex with microtubules hinted that ³⁰⁶VQIVYK³¹¹ forms local contacts with upstream flanking sequence²¹. This was recently supported by predicted models guided by experimental

restraints from cross-linking mass spectrometry¹⁶ and is consistent with independent NMR data^{22,23}.

Based on our prior work¹⁶ we hypothesized that tau adopts a β -hairpin that shields the ³⁰⁶VQIVYK³¹¹ motif and that disease-associated mutations near the motif may contribute to tau's molecular rearrangement which transforms it from an inert to an early seed-competent form by perturbing this structure. Many of the missense mutations genetically linked to tau pathology in humans occur within tau RD and cluster near ³⁰⁶VQIVYK³¹¹²⁴ (Fig. 1a, b and Table 1), such as P301L and P301S. These mutations have no definitive biophysical mechanism of action, but are nevertheless widely used in cell and animal models^{25,26}. Solution NMR experiments on tau RD encoding a P301L mutation have shown local chemical shift perturbations surrounding the mutation resulting in an increased β -strand propensity²⁷. NMR measurements have yielded important insights but require the acquisition of spectra in non-physiological conditions, where aggregation is prohibited. Under these conditions weakly populated states that drive prion aggregation and early seed formation may not be observed²⁸.

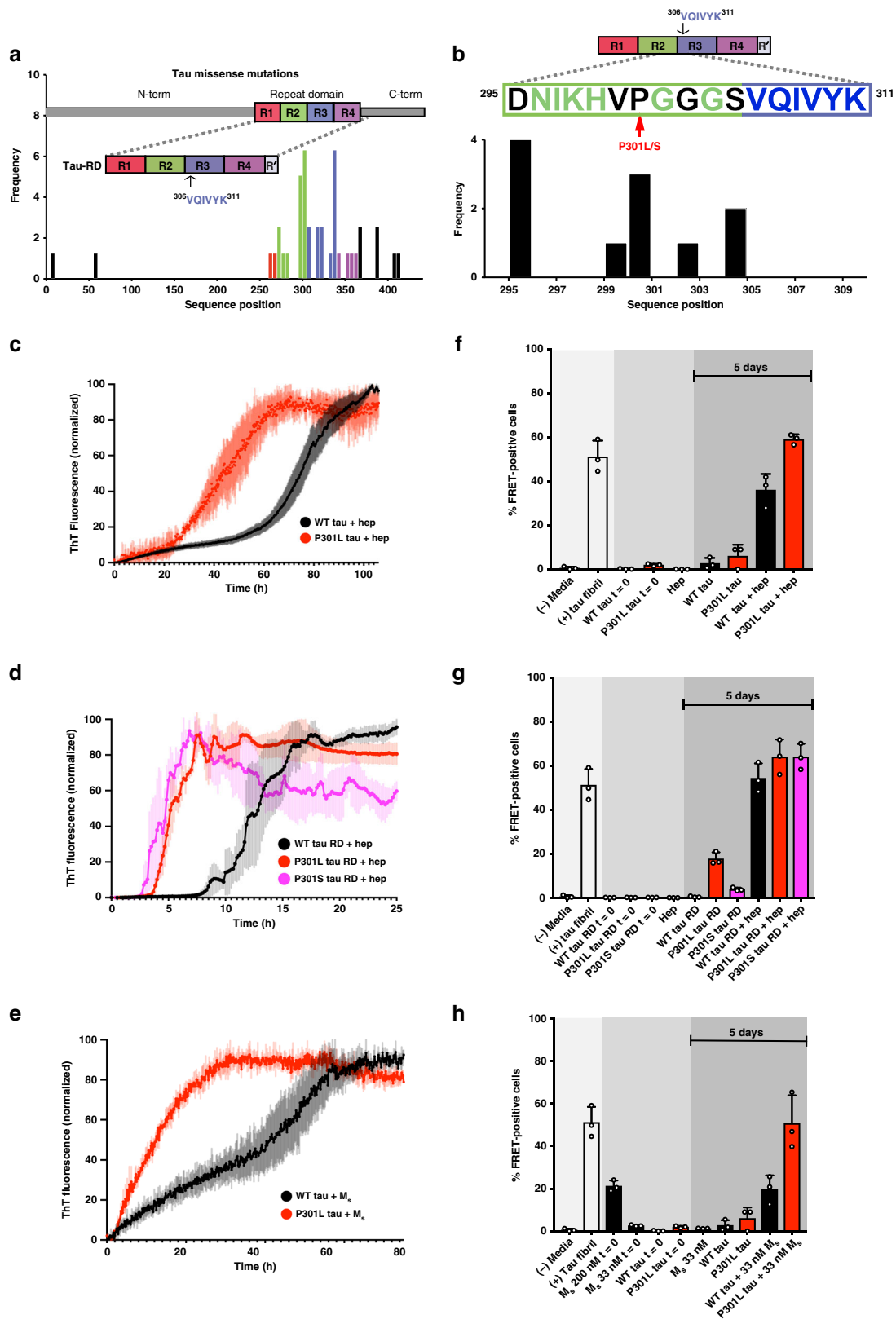
As with disease-associated mutations, alternative splicing also changes the sequence N-terminal to ³⁰⁶VQIVYK³¹¹. Tau is expressed in the adult brain primarily as two major splice isoforms: three-repeat and four-repeat²⁹. The truncated three-repeat isoform lacks the second of four imperfectly repeated segments in tau RD. Expression of the four-repeat isoform correlates with the deposition of aggregated tau tangles in many tauopathies³⁰ and non-coding mutations that increase preferential splicing or expression of the four-repeat isoform cause dominantly inherited tauopathies^{30–32}. It is not obvious why the incorporation or absence of the second repeat correlates with disease, as the primary sequences, although imperfectly repeated, are relatively conserved.

Previous reports have focused on the structure of a repeat with the assumption that each repeat functions independently within tau RD³³. These have described a relationship between the length of a repeat fragment, its propensity to spontaneously aggregate, and its seeding capacity in cells³³. However, inter-repeat interactions may also influence aggregation given that both alternative splicing and many disease-associated mutations cluster around the repeat interfaces (Fig. 1a). Our prior work suggested that wild-type tau aggregates less efficiently because the flanking sequence shields ³⁰⁶VQIVYK³¹¹¹⁶. We hypothesize that the intrinsically disordered tau protein evolved to minimize aggregation by adopting local structure that shields the ³⁰⁶VQIVYK³¹¹ amyloid motif from interactions leading to seed formation and amyloid propagation. We employed an array of *in silico*, *in vitro*, and cellular assays to elucidate the molecular interactions and physiological consequences of ³⁰⁶VQIVYK³¹¹ within tau. Our data support a model where disease-associated mutations, alternative splicing, or other factors can destabilize this local structure and expose ³⁰⁶VQIVYK³¹¹ leading to enhanced self-assembly.

Results

P301 mutations promote aggregation *in vitro* and in cells.

Missense mutations that change proline 301 to leucine or serine cause dominantly inherited tauopathy³⁴ and are associated with neurodegeneration in model systems^{26,35}, though the biophysical mechanism is not understood. We studied changes in aggregation propensity driven by mutations at P301 in full-length (FL) tau (2N4R; amino acids 1–441) and tau repeat domain (tau RD; amino acids 244–380) (Supplementary Table 1). First, we monitored aggregation of FL wild-type (WT) tau and mutant (P301L) tau using a Thioflavin T (ThT) fluorescence assay induced with stoichiometric amounts of heparin. We observed that P301L tau



($t_{1/2} = 41.6 \pm 0.5$ h) aggregated more rapidly compared with WT tau ($t_{1/2} = 75 \pm 0.3$ h) (Fig. 1c and Supplementary Data 1). Next, we compared changes in heparin-induced aggregation of the tau RD, comparing WT, P301L, and P301S mutants. We again observed that the two mutants aggregated faster (P301L tau RD, $t_{1/2} = 5.2 \pm 0.1$ h; P301S tau RD, $t_{1/2} = 3.9 \pm 0.1$ h) than WT tau

RD (WT tau RD, $t_{1/2} = 12.5 \pm 0.2$ h) (Fig. 1d and Supplementary Data 1). Consistently, we found that mutations at position 301 (from proline to either leucine or serine) increased aggregation rates by approximately twofold compared with WT in both FL tau and tau RD constructs. Thus, in vitro, tau RD recapitulates key aspects of aggregation observed in FL tau.

Fig. 1 Tauopathy mutations cluster to inter-repeat regions and promote aggregation. **a** Disease-associated mutation frequency found in human tauopathies. Most mutations are found within the repeat domain (tau RD) (repeat 1 = red; repeat 2 = green; repeat 3 = blue; repeat 4 = purple). Amyloidogenic sequence $^{306}\text{VQIVYK}^{311}$ is shown in the inset cartoon. **b** Detailed mutation frequencies found near the $^{306}\text{VQIVYK}^{311}$ amyloid motif. **c** FL WT tau and mutant P301L tau at a 4.4 μM concentration were mixed with stoichiometric amounts of heparin (4.4 μM), and allowed to aggregate in the presence of ThT at room temperature. Control WT and P301L tau in the absence of heparin yielded no detectible ThT signal change (less than twofold ratio to background signal) over the course of the experiment (see Supplementary Data 1). ThT fluorescence was normalized to the maximum for each condition. **d** WT tau RD and mutant P301L and P301S tau RD at a 4.4 μM concentration were each mixed with equimolar amounts of heparin (4.4 μM), and allowed to aggregate in the presence of ThT at room temperature. Control WT, P301L, and P301S tau RD in the absence of heparin yielded no detectible ThT signal change (less than twofold ratio to background signal) over the course of the experiment (see Supplementary Data 1). **e** WT FL tau and mutant P301L tau at a 4.4 μM concentration were mixed with sub-stoichiometric M_s tau seed (33 nM) and allowed to aggregate in the presence of ThT at room temperature. Control WT and P301L tau in the absence of M_s yielded no detectible ThT signal change (less than twofold ratio to background signal) over the course of the experiment (see Supplementary Data 1). All ThT experiments were carried out in triplicate. The data are shown as the average with standard deviation and are colored according to mutation. **f–h** After 120 h of in vitro incubation, proteins from previous ThT experiments were transduced into tau biosensor cells via lipofectamine (Methods). FRET signal from each condition (tau RD-CFP/tau RD-YFP) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition. Error bars represent a 95% CI of each condition

Table 1 List of AlzForum disease-associated mutations

Name	Amino-acid sequence
Tau RD AlzForum Mutations ^a	R1 : 244 Q T APVMPD L KN- V K S K I G S TEN L K H Q P G G K 274 R2 : 275 V Q I I N K K L D L S N -V Q S K C G S K D N I K H V P G G S 305 R3 : 306 V Q I V Y K P V D L S K - V T S K C G S L G N I H H K P G G Q 336 R4 : 337 V E V K S E K L D F K D R V S K I G S L D N I T H V P G G N 368

^aSites of mutation are shown in bold

The inert conformation of monomeric tau (M_i) requires co-factors, such as heparin, to spontaneously aggregate in vitro, whereas the seed-competent monomer (M_s), derived from recombinant protein or Alzheimer's patient brain material, readily self-assembles to form amyloid¹⁶. Previously we determined that M_s converts FL tau into fibrils at sub-stoichiometric ratios, in contrast to the stoichiometric amounts necessary in heparin-containing reactions¹⁶. In this study, we evaluated the aggregation propensity of the P301L mutant compared with WT when incubated in the presence of recombinantly produced M_s . We incubated FL tau with sub-stoichiometric amounts of M_s (1:133) and monitored aggregation using ThT. In comparison, we observed that M_s -seeded P301L tau self-assembled more rapidly (P301L tau, $t_{1/2} = 8.5 \pm 0.6$ h) than the WT protein (WT tau, $t_{1/2} = 40 \pm 1.1$ h) (Fig. 1e and Supplementary Data 1). P301L tau aggregated faster than WT tau with a fourfold increase in rate after seeding by M_s . Independent of induction—heparin or M_s —P301L assembled into ThT-positive aggregates more rapidly. Moreover, tau appeared to be more sensitive to M_s seeded aggregation compared with heparin, given the sub-stoichiometric ratios needed for robust aggregation. The effectiveness of M_s to seed aggregation of M_i may be explained by a direct templating of M_i to M_s at the amyloid motif region, interface of repeat 2 and 3, which we previously characterized to be more exposed in M_s ¹⁶. Mutations at the P301 may exacerbate aggregation by unfolding the region surrounding the amyloid motif $^{306}\text{VQIVYK}^{311}$, thereby producing a more compatible conformation for the similarly expanded aggregation-prone M_s seed.

To test the structural compatibility of aggregates formed by in vitro tau models, we employed tau biosensor HEK293 cells that stably express tau RD (P301S) fused to cyan or yellow fluorescent proteins²⁵. These cells sensitively report a fluorescence resonance energy transfer (FRET) signal (tau RD-CFP/tau RD-YFP) only when aggregated in response to tau amyloid seeds, and are unresponsive to aggregates formed by other proteins, such as huntingtin or α -synuclein³⁶. Each sample formed amyloid fibril morphologies confirmed by transmission electron microscopy,

except for samples not incubated with heparin or M_s and the low-concentration M_s , where no large ordered structures were found (Supplementary Figure 1). The tau biosensor cells responded to FL tau fibrils created by exposure to heparin and showed an increase in seeding activity for the P301L mutant compared with WT fibrils (Fig. 1f and Supplementary Data 2). Next, we compared seeding for the tau RD heparin-induced fibrils and again found that P301L and P301S mutants produced higher seeding activity relative to WT (Fig. 1g and Supplementary Data 2). At last, the seeding activity for the M_s -induced FL tau fibrils showed a twofold higher activity for P301L compared with WT (Fig. 1h and Supplementary Data 2). WT FL tau and tau RD control samples (no heparin or M_s) did not produce seeding activity in cells, whereas P301 mutants, both FL and tau RD, showed hints of seeding activity despite not yielding positive ThT signal in vitro (Supplementary Data 1), perhaps owing to the formation of oligomers not captured by ThT. As expected, 33 nM M_s control exhibited seeding activity at the onset and did not change after 5 days, but overall signal was low owing to the low concentrations used in the aggregation experiments. Interestingly, WT tau induced with 33 nM M_s seeded at similar levels to concentrated control (200 nM) M_s samples highlighting efficient conversion of WT tau into seed-competent forms (Fig. 1h and Supplementary Data 2). Thus, P301 mutations promote aggregation in vitro and in cells across different constructs. Importantly, these effects are conserved between FL tau and tau RD.

Mutations at P301 destabilize native tau structure. To determine how the P301L mutation drives conformational changes, we employed cross-linking mass spectrometry (XL-MS) in a heat denaturation experiment. XL-MS defines amino-acid contacts in proteins and can thus guide the determination of structure for large complexes, transient interactions, and dynamics of intrinsically disordered proteins^{16,37,38}. To preclude the formation of intermolecular cross-links between monomers, low-concentration samples of WT, P301L, and P301S tau RD (Supplementary Table 1) were incubated at different temperatures, reacted with

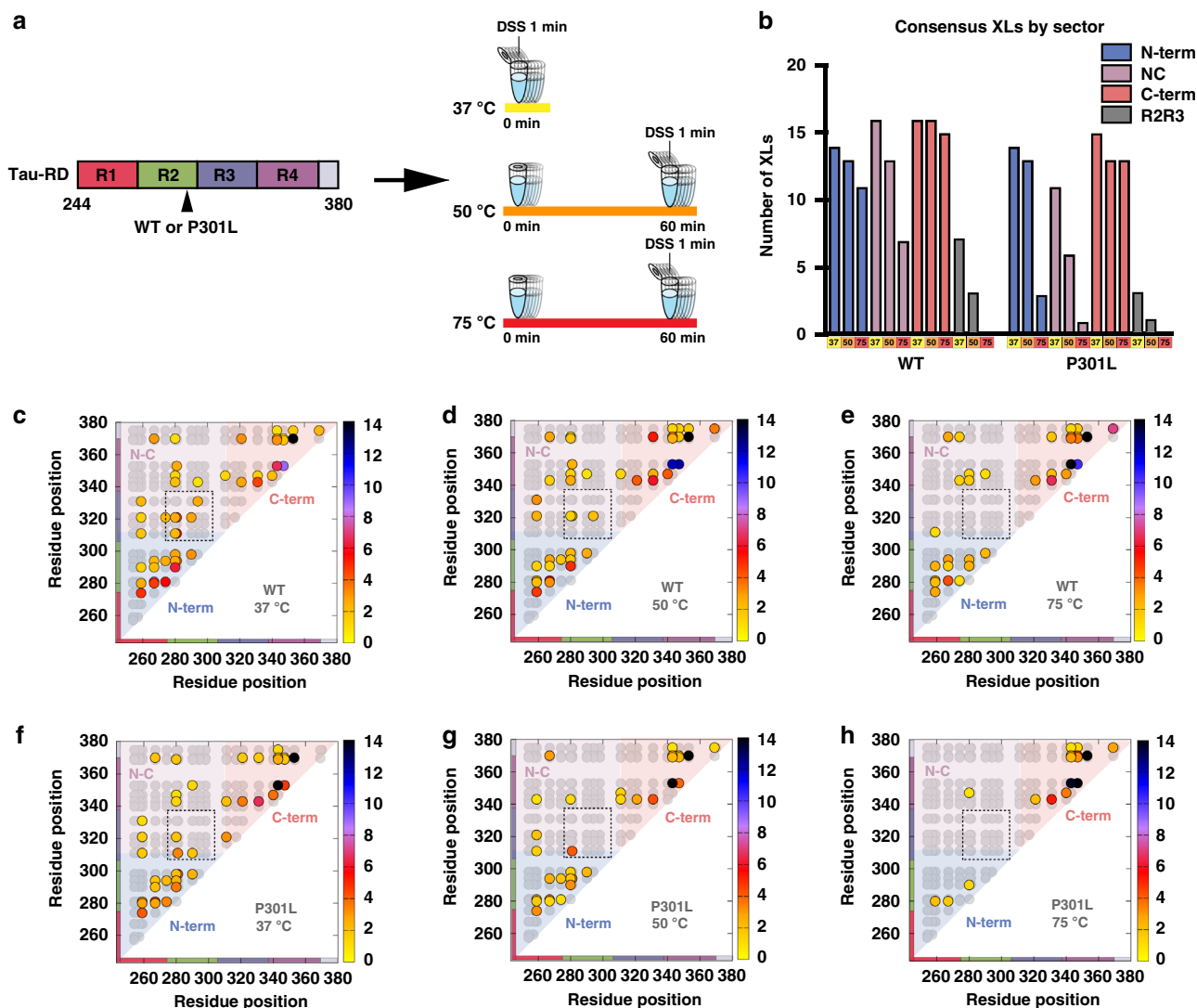


Fig. 2 Tau RD encodes global and local structure. **a** Cartoon schematic of tau RD used for XL-MS studies colored according to repeat domain. Recombinant WT and P301L tau RD were heated at 37 °C, 50 °C or 75 °C for 1 hour, then chemically cross-linked using DSS. After cross-linking, trypsin fragmentation, and LC-MS/MS analysis were performed. Each sample was carried out in five technical replicates. **b** Total consensus cross-links parsed by temperature and location in WT and P301L tau RD: within N-terminus (blue; residues 243–310; N-term), within C-terminus (orange; residues 311–380; C-term), span N- and C-terminus (magenta; between residues 243–310 and 311–380; N-C) and between repeat 2 and repeat 3 (R2R3) (gray; between residues 275–305 and 306–336). **c–e** Consensus cross-links (circles) are shown in contact maps color coded by average frequency across replicates. The theoretical lysine pairs are shown in the background as gray circles. Cross-link contacts within the N-term (blue), C-term (red), and across N- to C-term (purple) are shown as sectors. The x and y axis are colored according to repeat number as in Fig. 1. The dashed boxes define inter-repeat cross-links observed between repeat 2 and repeat 3. **f–h** Same as **c–e** above, except with tau RD that contains a P301L mutation

disuccinimidyl suberate (DSS; a primary amine crosslinker) for 1 min and quenched (Fig. 2a). The cross-linked protein monomers were confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (Supplementary Figure 2a). Cross-linked samples were trypsin digested, analyzed by mass spectrometry and the spectra were searched using Xquest³⁹ to identify intramolecular protein contact pairs (Methods and Supplementary Data 3). In each data set, the cross-links reported represent consensus data across five independent samples with a low false discovery rate (FDR) (Methods, Supplementary Figure 3 and Supplementary Data 4). XL-MS of recombinant WT tau RD acquired at 37 °C revealed three clusters of interactions that localize within the N-terminus (residues 243–310; N-term), within the C-terminus (residues 311–380; C-term) and span N- and C-termini (N-C; between residues 243–310 and 311–380) (Fig. 2b, c). Importantly, the experimentally observed cross-links

represent only a small subset of all theoretical lysine pairs suggesting that tau RD samples have discrete modes of contacts (Fig. 2c, gray circles). Heating the samples to 50 °C or even 75 °C decreased the number of N-C long-range and N-term short-range contacts identified (Fig. 2b, d, e and Supplementary Figure 2b). The data acquired for WT tau RD in physiological conditions are consistent with a loose metastable structure comprised of weak long-range and short-range contacts that are sensitive to temperature.

In contrast, XL-MS of recombinant tau RD with the P301L mutation revealed an increased susceptibility to heat denaturation. At 37 °C, the cross-links found in P301L tau RD (Fig. 2f) were similar in pattern to WT, except for fewer N-C-terminal long-range contacts (Fig. 2b and Supplementary Figure 2b, c). However, samples incubated at 50 °C and 75 °C revealed a significant reduction in both long-range and short-range

contacts in P301L tau RD compared with WT (Fig. 2b). The loss of short-range contacts, both the total number of cross-links and the abundance of each cross-link, were detected particularly within the N-terminal sector, which sits upstream of P301 (Fig. 2b, f–h and Supplementary Figure 2b, c). In contrast, the C-term local contacts sample many theoretical lysine–lysine pairs and remained relatively constant across temperatures for both WT and P301L tau RD, possibly suggesting a higher degree of disorder that is independent of the mutation site. Moreover, a stepwise loss of the inter-repeat cross-links between repeat 2 and 3 was seen in the heat denaturation of WT tau RD and was more pronounced with the P301L mutation, indicating an unfolding of local structure at the interface of repeat 2 and 3, encompassing $^{306}\text{VQIVYK}^{311}$ (Fig. 2b, gray bars; Fig. 2c–h, inset box). Thus, although the number of cross-links identified was comparable between WT and P301L tau RD at 37 °C, P301L tau RD retained approximately half as many cross-links as WT when heated. Similar cross-linking profiles were observed for the P301S tau RD sample (Supplementary Figure 2d). Thus, the lack of thermostability in P301 mutated tau RD as compared with WT tau RD suggests that the P301 mutations may lower the threshold for tau to enter into an aggregation-prone conformation.

Tau RD models show local structure in inter-repeat elements.

To gain insight into what types of local structures the inter-repeat elements can form, we used ROSETTA to predict structures of tau RD. We built 5000 models using two parallel strategies in ROSETTA: *ab initio* that employed fragment libraries derived from experimental structures⁴⁰ and CS-ROSETTA, which leveraged available chemical shifts for tau RD to produce fragment libraries⁴¹. Both approaches led to a diversity of models consistent with experimentally determined radii of gyration⁴² (Supplementary Figure 4a–d and Supplementary Data 5). Analysis of each structural ensemble showed a propensity to form hairpin-like structures across R1R2, R2R3, R3R4, and R4R' repeat interfaces centered on the $^{271}\text{PGGG}^{274}$, $^{301}\text{PGGG}^{304}$, $^{333}\text{PGGG}^{336}$, and $^{366}\text{PGGG}^{369}$ sequences (Supplementary Figure 4e, Supplementary Figure 5a–d, and 6 and Supplementary Data 5). Previously published solution NMR data have shown that the PGGG sequences in tau can adopt type II β -turns⁷, and the $^{301}\text{PGGG}^{304}$ sequence preceding $^{306}\text{VQIVYK}^{311}$ is compatible with the formation of a β -hairpin. We illustrated the R2R3 $^{306}\text{VQIVYK}^{311}$ -

containing fragment derived from low energy expanded models produced by each method (Supplementary Figure 4c, d). The $^{306}\text{VQIVYK}^{311}$ -containing interface has the highest frequency of disease-associated mutations, particularly P301L and P301S (Fig. 1a). Other potential amyloid-forming regions, such as the aggregation-prone $^{275}\text{VQIINK}^{280}$ (Supplementary Figure 6), is also preceded by $^{271}\text{PGGG}^{274}$ and predicted to form a β -hairpin (Supplementary Figure 4e and Supplementary Figure 5), however, it is absent in recent cryo-EM structures of tau aggregates^{3,43}. Mapping known missense mutations onto the *ab initio* β -hairpin structure at the R2R3 interface (Supplementary Figure 4f), we hypothesized that this cluster of disease-associated mutations could destabilize the β -hairpin secondary structure, thus exposing the amyloid motif $^{306}\text{VQIVYK}^{311}$ and enabling aggregation. This model is compatible with recent cryo-EM findings that indicate a disengagement of the $^{306}\text{VQIVYK}^{311}$ N-terminal flanking sequence in a fibril structure³. Thus, we focused our studies on the R2R3 motif of tau that contains $^{306}\text{VQIVYK}^{311}$.

P301L promotes extended forms of tau. In silico modeling corroborated recent biochemical findings¹⁶ and suggested a minimal sequence necessary to form a collapsed structure around $^{306}\text{VQIVYK}^{311}$. To understand how these structures might self-assemble, we employed molecular dynamics (MD) simulations of two tau peptide fragments comprising the minimally structured fragment centered around the R2R3 interface ($^{295}\text{DNIKHVPPGGGSVQIVYK}^{311}$): R2R3-WT and R2R3-P301L (Supplementary Table 2). To enable sufficient sampling of oligomer structures, we employed an unbiased algorithm based on a recently developed symmetry-constraint approach⁴⁴. The trimer conformations obtained in simulations are depicted on a root mean square deviation (RMSD) matrix for both the R2R3-WT (Fig. 3a) and the R2R3-P301L mutant peptide fragments (Fig. 3b). For the R2R3-WT peptide fragment, we observe a dominant population of trimeric conformations composed of hairpins, whereas the P301L disease-associated mutation stabilizes an extended fibrillar form. The energy basin for the R2R3-WT peptide fragment is predicted to be 5–6 kJ/mol lower in a collapsed state than an extended state, whereas the R2R3-P301L peptide fragment is 3 kJ/mol lower in an extended state than a collapsed state (Fig. 3c and Supplementary Data 6). In addition, the free-energy surface suggests an energy barrier of ~5 kJ/mol

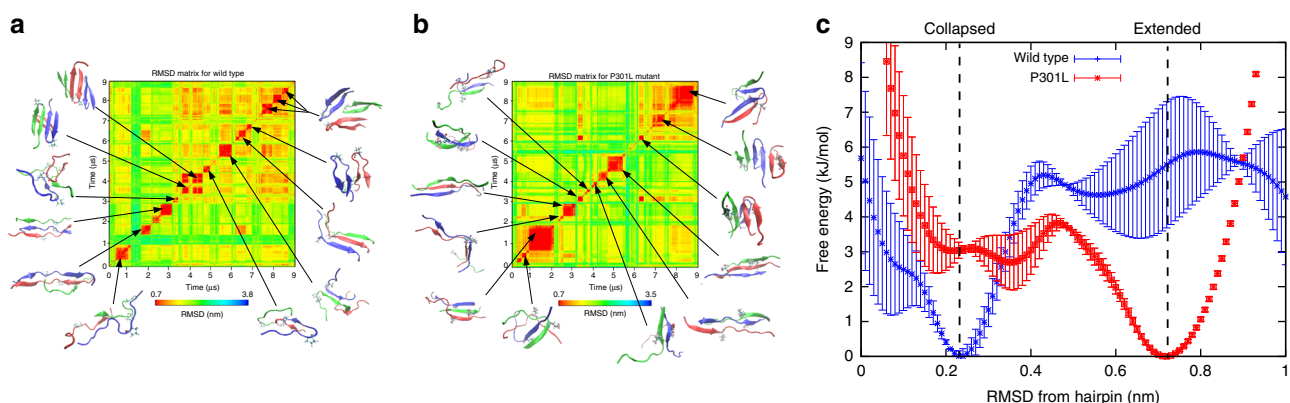


Fig. 3 Wild-type and mutant peptides differentially populate collapsed and extended conformations. **a** Trimer conformations obtained from MD simulations of WT peptide fragment (R2R3-WT) with the sequence $^{295}\text{DNIKHVPPGGGSVQIVYK}^{311}$. Two-dimensional root mean-squared-differences (RMSD's) are calculated between all pairs of conformations visited during MD simulations. Snapshots of trimeric structures are depicted for selected metastable basins, with each peptide monomer represented by a different color. **b** The same analysis as in **a**, but for the P301L substituted trimer. **c** The free-energy surface as a function of deviation from a canonical hairpin structure. Two distinct basins, corresponding to collapsed and extended sub-ensembles, are found in WT and P301L peptide fragment, respectively. Error bars represent a 95% CI of each RMSD bin

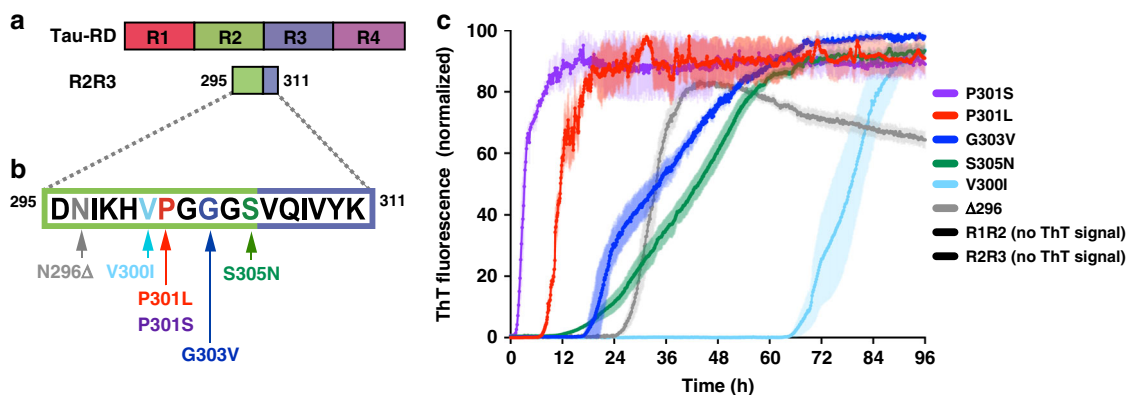


Fig. 4 Tauopathy mutations drive aggregation propensity. **a** Schematic of tau RD and the derived peptides representing the minimal structural element around ³⁰⁶VQIVYK³¹¹. **b** WT and mutant peptides were disaggregated, resuspended to 200 μM, and allowed to aggregate in the presence of ThT at room temperature. The WT R2R3 and R1R2 fragment peptides yielded no detectible ThT signal change (less than twofold ratio to background signal) over the course of the experiment (see Supplementary Data 1). ThT signals are shown as average of triplicates with standard deviation, are colored according to mutation and are normalized to the maximum for each condition

to convert the R2R3-WT peptide fragment from collapsed to extended. That same barrier however is <1 kJ/mol for the R2R3-P301L peptide fragment, suggesting a faster rate of kinetic conversion from collapsed hairpin to extended fibrillar form. Thus, MD predicts that the P301L mutation promotes amyloid assembly by destabilizing monomeric hairpin structures.

Tau amyloid formation is governed by flanking residues. In tau RD, ³⁰⁶VQIVYK³¹¹ is necessary for amyloid formation^{5,6}. In solution, the ³⁰⁶VQIVYK³¹¹ hexapeptide aggregates spontaneously and rapidly as measured by ThT fluorescence intensity, whereas the upstream N-terminal sequence ²⁹⁵DNIKHV³⁰⁰ does not aggregate (Supplementary Figure 6). To experimentally test the prediction of a local hairpin structure encompassing ³⁰⁶VQIVYK³¹¹, we employed a mutagenesis study on synthetic peptide systems that recapitulate the minimal hairpin sequence (Supplementary Table 2). Consistent with the prediction from MD simulation, the R2R3-WT peptide fragment did not aggregate readily, with no ThT detected within 96 h (Fig. 4a, c and Supplementary Data 1). By contrast, single disease-associated mutations (Fig. 4b and Supplementary Data 1) substituted into the R2R3 peptide fragment were sufficient to promote spontaneous amyloid formation: R2R3-P301S ($t_{1/2} = 4.1 \pm 1.3$ h), R2R3-P301L ($t_{1/2} = 7.2 \pm 0.2$ h), R2R3-N296Δ ($t_{1/2} = 31.9 \pm 0.2$ h), R2R3-G303V ($t_{1/2} = 32.1 \pm 0.7$ h), R2R3-S305N ($t_{1/2} = 41.2 \pm 0.2$ h), and R2R3-V300I ($t_{1/2} = 77.8 \pm 1.3$ h) (Fig. 4c and Supplementary Data 1). Each of these peptides was confirmed to form amyloid-like fibril morphologies by transmission electron microscopy, except for the WT R2R3 peptide fragment where no large structures were found (Fig. 5b–h).

To test the structural compatibility of peptide aggregates formed by in vitro tau models, we again employed tau biosensor cells²⁵. The tau biosensor cells responded to all disease-associated tau peptide fragments that aggregated spontaneously in vitro, but not to the wild-type R2R3 peptide fragment (which did not aggregate in vitro) (Fig. 5a and Supplementary Data 7). Qualitatively, biosensor cells retained their diffused tau localization when untreated or exposed to a wild-type R2R3 peptide fragment but formed fluorescent puncta when cultured with aggregated mutant peptides (Fig. 5i–p). Interestingly, the biosensor cells responded to disease-associated mutant peptides with varying degrees of sensitivity and created distinct aggregate morphologies. This is consistent with amyloid structures that act as distinct templates and form the basis of tau prion strains^{4,45}.

Thus, the R2R3 peptide fragment model system responds to mutations in vitro and in cells similarly to the FL tau and tau RD system, suggesting that local conformational changes in tau can be recapitulated using shorter fragments.

Tau splice variants reveal different aggregation propensity. Tau is expressed in the adult brain as six major splice isoform types that include either three or four repeated segments within RD (Fig. 6a). 3R tau lacks the second of four imperfect repeats. 4R tau correlates strongly with aggregation in most tauopathies³⁰ and mutations that increase splicing of the 4R isoform cause dominantly inherited tauopathies^{30–32}. We examined whether this splice isoform affects the propensity of ³⁰⁶VQIVYK³¹¹-mediated aggregation owing to the different composition of upstream flanking sequence. We constructed a series of peptide fragments to encompass the R1R3 interface (Fig. 6b). This wild-type peptide fragment R1R3 mimicking a 3R splice isoform did not spontaneously aggregate (Supplementary Figure 7 and Supplementary Data 1). Surprisingly, an R1R3 peptide fragment with a corresponding P301L mutation (R1R3-P270L) also did not aggregate (Fig. 6, Supplementary Figure 7 and Supplementary Data 1). We hypothesized that the R1-leading sequence stabilizes the amyloid motif ³⁰⁶VQIVYK³¹¹, resulting in the aggregation resistance in the presence of disease-associated mutations.

The R1-leading sequence ²⁶⁴ENLKHQPGGGK²⁷³ differs from R2 ²⁹⁵DNIKHVPGGGG³⁰⁴ at four amino-acid positions. To identify which amino acid(s) governed R1's stronger inhibitory effects, we constructed 16 peptides with a P301L mutation to represent every combinatorial sequence between the two leading strands and measured their aggregation kinetics (Fig. 6b, Supplementary Figure 7 and Supplementary Data 1). We identified a general trend where the R2R3-P301L peptide fragment aggregates in hours with zero or one R1 substitutions. With two R1 substitutions, the R2R3-P301L peptide aggregation was delayed roughly an order of magnitude to tens of hours. With three R1 substitutions, the R2R3-P301L peptide fragment aggregation was further delayed to hundreds of hours. With all four R1 substitutions in the peptide (R1R3-P301L), no ThT signal was observed within a week (Fig. 6b and Supplementary Figure 7). Thus, all four amino acids contributed to the ability of the R1 leading sequence to delay ³⁰⁶VQIVYK³¹¹-mediated spontaneous aggregation in a 3R splice isoform. This may explain the differential aggregation propensities of tau isoforms in human pathology.

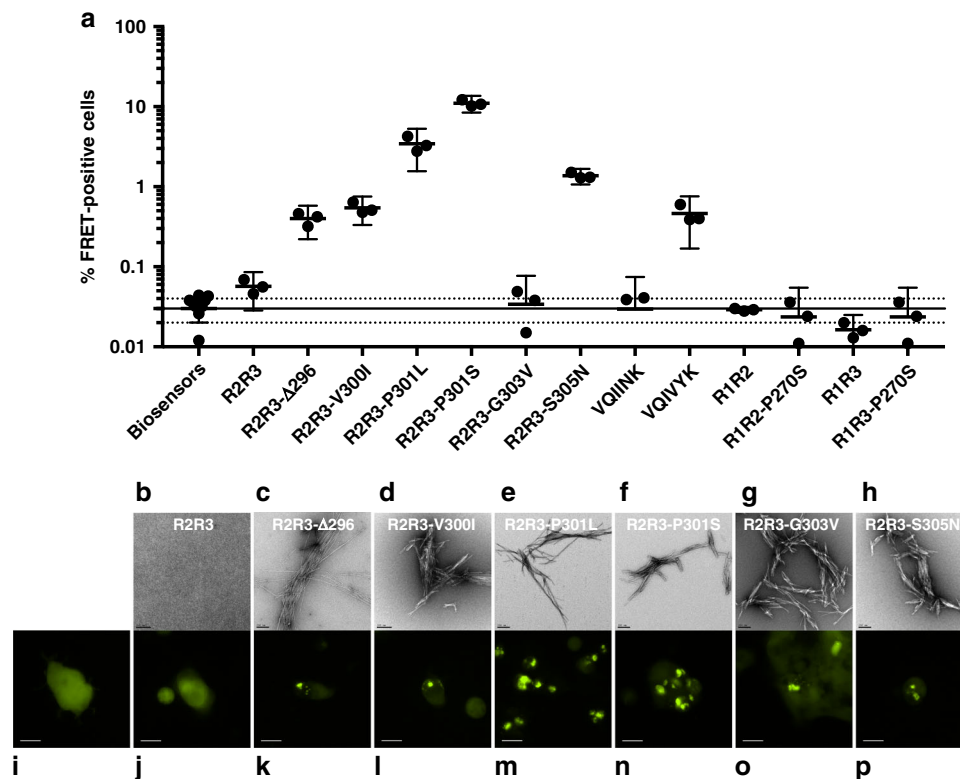


Fig. 5 Peptides form amyloid structures and seed in vivo. **a** After 96 h of in vitro incubation, peptides from previous ThT experiments (Fig. 4c) were transduced into tau biosensor cells via lipofectamine (Methods). FRET signal from each condition (tau RD-CFP/tau RD-YFP) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition. Error bars represent a 95% CI of each condition. Solid and dashed horizontal lines represent the mean and 95% error from untreated biosensor cells, respectively, for ease of statistical comparison. **b-h** Electron microscopy images of each peptide from previous ThT experiments (Fig. 4c). The black bar represents 200 nm distance in each image. **i-p** Qualitative fluorescence microscopy images of tau biosensor cells immediately prior to flow cytometry experiments. **i** shows a representative image of untreated biosensor cells. **j-p** each shows a representative image of biosensor cells treated with samples from **b-h**, respectively. The white bar represents 10 μ m distance in each image

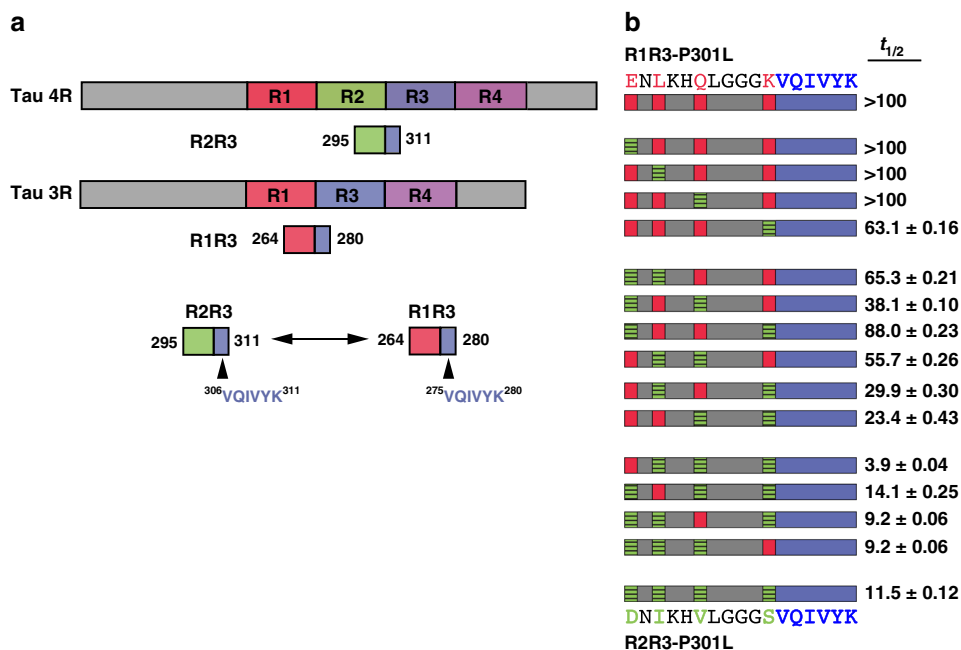


Fig. 6 Alternative splicing modulates aggregation propensity. **a** Cartoon schematic for tau 4R and 3R splice isoforms illustrate the difference in primary amino-acid sequence leading into the amyloidogenic ³⁰⁶VQIVYK³¹¹ motif. **b** A full combinatorial panel of R2R3-P301L and R1R3-P301L chimeras were aggregated in vitro. ³⁰⁶VQIVYK³¹¹ is shown in blue, amino acids common between the splice isoforms are shown in gray, amino acids unique to an R3 isoform are colored red, amino acids unique to an R4 isoform are colored green. The aggregation kinetics, represented as $t_{1/2}$ in hours with 95% CI, are listed in the right-side column alongside its respective peptide

Stabilizing β -hairpin blocks P301L-mediated aggregation. Our model predicted shielding of the $^{306}\text{VQIVYK}^{311}$ motif in tau via local β -structure, and thus we hypothesized that artificially stabilizing the termini of this local structure would promote a more inert, closed conformation. To test this, we flanked the R2R3-P301L peptide fragment with a tryptophan zipper (Trp-R2R3-P301L-Trp, Supplementary Table 2 and Fig. 7a), which stabilizes a β -hairpin structure ~ -2.5 to -7 kJ/mol⁴⁶. Consistent with our model, the Trp-R2R3-P301L-Trp peptide fragment does not spontaneously aggregate in vitro (Fig. 7b and Supplementary Data 1).

To ensure that this effect was not a result of adding bulky tryptophan residues, we constructed control peptide fragments that contain only the N-term (Trp-R2R3-P301L) or the C-term (R2R3-P301L-Trp) portion of the tryptophan zipper sequence (Fig. 7a). Both half-sequence controls spontaneously aggregated, implying that a tryptophan at either position is insufficient to block aggregation (Fig. 7b). Only a fully intact tryptophan zipper that stabilizes a β -hairpin conformation ameliorates aggregation

propensity. Alternative methods to stabilize a β -hairpin architecture, such as introducing isoelectric interactions, also delayed aggregation: peptides containing two additional aspartic acid on the N-terminus and two lysine on the C-terminus (R2R3-IEZip, Supplementary Table 2) retarded the R2R3-P301L peptide fragment aggregation over an order of magnitude ($t_{1/2} = 7$ h to $t_{1/2} > 70$ h, Supplementary Figure 8 and Supplementary Data 1).

To test this effect in cells, we engineered tau RD (P301S) biosensor cells encoding tryptophan zipper motifs that flank the R2R3 element. These biosensors had a significantly diminished capacity to be seeded; R2R3-P301S peptide fragment aggregates triggered aggregation in $11 \pm 1\%$ of tau biosensor cells, but only $0.36 \pm 0.12\%$ of the tryptophan zipper stabilized biosensor cells (Supplementary Figure 9 and Supplementary Data 7).

Proline 301 *cis-trans* isomerization modulates aggregation.

Many proteins in the cell utilize proline isomerization as a molecular switch, such as heat shock protein activation⁴⁷ or cell cycle regulation⁴⁸. In some proteins, proline isomerization directly

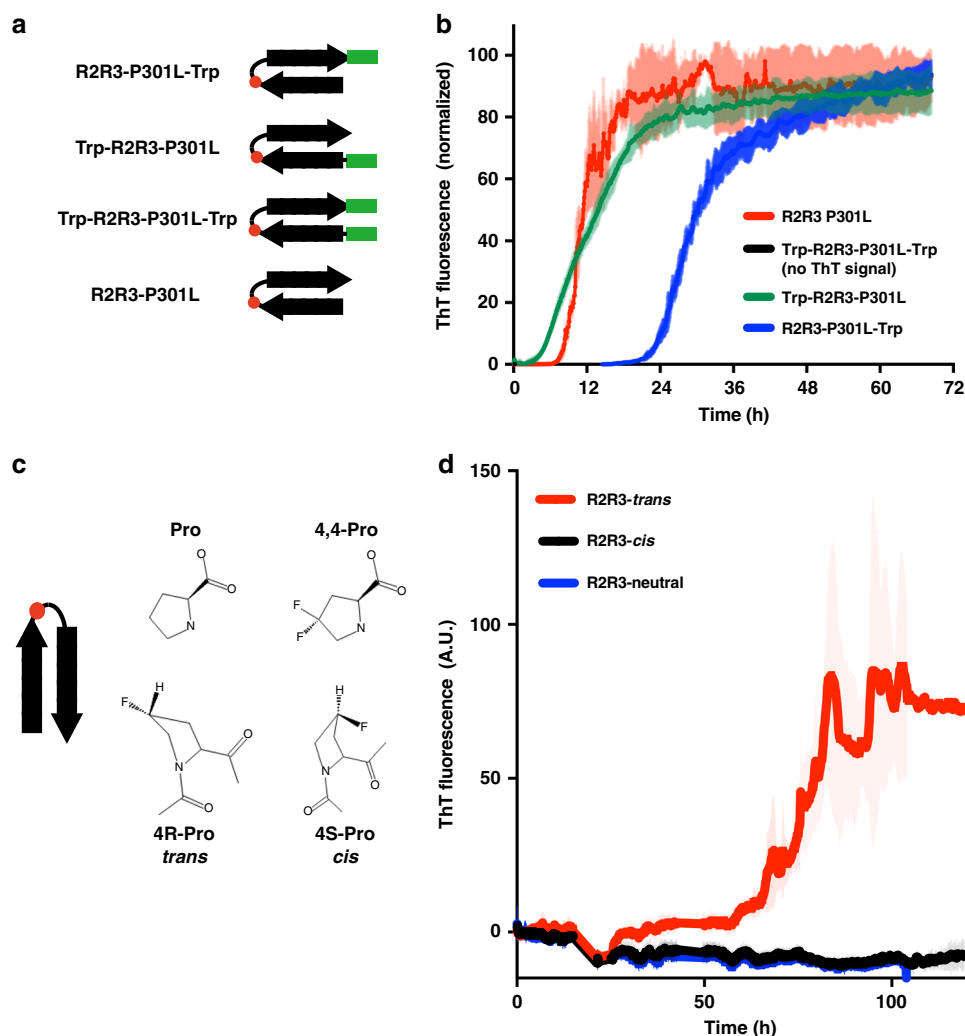


Fig. 7 Enhancing β -hairpin structure rescues spontaneous aggregation phenotypes. **a** Cartoon schematic representation of the tryptophan zipper motif (green bar) and controls used to stabilize a β -hairpin structure in an R2R3-P301L peptide fragment (Supplementary Table 2). **b** Aggregation reactions of the tryptophan zipper peptide and controls measured by ThT fluorescence. The Trp-R2R3-P301L-Trp fragment peptide yielded no detectible ThT signal change (less than twofold ratio to background signal) over the course of the experiment (see Supplementary Data 1) ThT signals are shown as average of triplicates with standard deviation and were normalized to the maximum for each condition. **c** Schematic of proline and fluorinated proline analogs used to generate *cis* and *trans* proline conformers at the position corresponding to P301 (red circle) in peptide models. **d** ThT aggregation reactions of the *cis*, *trans*, and neutral proline analogs substituted into the R2R3 peptide fragment. ThT signals are an average of six independent experiments with standard deviation shown

induces or mitigates aggregation into amyloid^{49–51}. Proline isomerization events in tau have been proposed to play a role in aggregation and disease⁴⁹, but P301 isomerization has not been linked to tau aggregation and pathology. With the fact that serine or leucine substitutions at P301 proximal to ³⁰⁶VQIVYK³¹¹ drastically alter aggregation propensity, we hypothesized that P301 plays a crucial role inducing a β -turn in a PGGG motif, which mediates a collapsed structure. To test whether isomerization of P301 could influence spontaneous amyloid formation, we constructed a series of R2R3 peptide fragments with proline analogs that preferentially populate either: (1) a *cis* rotamer (2S,4S)-fluoroproline; (2) a *trans* rotamer (2S,4R)-fluoroproline; or (3) an analog that easily interconverts between *cis* and *trans* (4,4)-difluoroproline (Fig. 7c, Supplementary Table 2 and Supplementary Data 1). Only the R2R3-Trans peptide spontaneously aggregated (Fig. 7d and Supplementary Data 1), indicating the potential for proline isomerization events in tau pathogenesis.

Discussion

Here, we establish the molecular and functional basis for how a series of prominent tau mutations drive aggregation. Integrating experimental and computational approaches, we independently and directly probed the local structural changes within tau. We identified metastable local structures within the inter-repeat junction of tau RD (the repeat 2–3 interface), which encompasses the amyloidogenic ³⁰⁶VQIVYK³¹¹ motif. This R2R3 interface becomes less stable when a disease-associated mutation is present, such as P301L, which is commonly employed in cell and animal models of tauopathy. Thus, P301L and similar mutations decrease the threshold for local structural expansion, especially in the presence of stressors (heat, seeds, heparin, or high concentration). This in turn is predicted to enhance the conversion of tau into a seed-competent form¹⁶. Thus, the proposed model rationalizes the fundamental molecular mechanisms of aggregation for P301L and at least five other mutations, explains why P301L spontaneously aggregates in animal and cellular models, and defines how splice isoforms of tau and proline isomerization at P301 may contribute to aggregation. Ultimately, these insights may inform the mechanisms of tauopathy in human disease and potential molecular targets for therapeutic development.

In vitro induction of tau aggregation is typically achieved by the addition of polyanionic molecules such as heparin, arachidonic acid, or nucleic acids^{10,11,52}. It is thought that heparin binding to tau expands the local conformation of the repeat 2 and repeat 3 regions, thereby exposing amyloidogenic sequences for subsequent aggregation^{12,16,52}. This process, however, requires stoichiometric amounts of polyanion and is not a physiological condition, as heparin is not present intracellularly. Our recent work has elucidated a seed-competent form of tau monomer that can promote tau aggregation. This seed-competent monomeric tau is found in AD patient brains and is likely the incipient species contributing to pathology¹⁶. We find that sub-stoichiometric amounts of M_s (1:133) enhance the rate of WT tau aggregation relative to heparin. Parallel experiments with P301L tau show an even more dramatic enhancement. Our data support that the ³⁰⁶VQIVYK³¹¹ motif is preferentially exposed in M_s or P301L mutant in contrast to normal tau where it is relatively shielded. Thus, the marked sensitivity of P301L to seeds can be explained by an increased exposure of the aggregation-prone ³⁰⁶VQIVYK³¹¹ sequence. These data suggest that M_s functions catalytically to convert normal tau into aggregates. Thus, the proposed seeding mechanism of M_s may be generalized to tauopathies that are not caused by mutations.

Ensemble averaging methods, such as NMR, have had limited success in understanding the solution conformations of tau under physiological conditions. They have revealed secondary structure

propensities of key regions and proposed the existence of local contacts^{2,7,22,23,53}. However, capturing more transient or low population local conformations has been difficult. This is confounded by poor signal to noise, requiring long acquisition times at high concentrations, and non-physiological temperatures to suppress protein aggregation. As such, capturing transient but important local structural signatures have been challenging with classical structural biology methods. Both experiment and simulation have shown that weak local structure may play key roles in limiting aggregation of globular proteins during translation and that these structural elements may play even bigger roles in intrinsically disordered proteins^{54,55}. Thus, local structures that bury proximal amyloid sequences may be a general evolutionary design principle that controls aggregation.

Our study has suggested that local structure encompassing the amyloid motif ³⁰⁶VQIVYK³¹¹ regulates aggregation of tau and that the P301L mutation increases susceptibility to conformational changes that expose the ³⁰⁶VQIVYK³¹¹ amyloid motif. Although these differences are subtle, we observe that P301L-mediated structural rearrangements only manifest under moderate stress conditions (i.e., heat, seed). Hence, as compared with NMR, real-time assays, such as XL-MS that kinetically traps conformations are more appropriate to detect metastable sub-populations. These data may explain the elusiveness of a biophysical basis of the cluster of pathogenic mutations near ³⁰⁶VQIVYK³¹¹. Simulations predict that repeat interfaces could encode local structures that are compatible with a β -hairpin and that the P301L mutation, dramatically shifted the equilibrium away from collapsed hairpins to extended fibril-like conformations. Our findings are consistent with published NMR data—PGGG sequences in tau can adopt type II β -turns⁷ and that the P301L mutation increases local β -strand propensity²⁷. Thus, our work supports the structural and functional findings that metastable local structures in tau are destabilized by disease-associated mutations.

Guided by our simulations, we predicted that a local fragment spanning the interface between repeat 2 and 3 should encode a minimal structure necessary to replicate this aggregation phenomenon. We examined whether structural perturbations influenced aggregation propensity in a peptide model system that captures this local structural element. The WT tau interface peptide model containing ³⁰⁶VQIVYK³¹¹ did not aggregate spontaneously; however, single point substitutions of six disease-associated mutations immediately N-terminal to ³⁰⁶VQIVYK³¹¹ consistently induced spontaneous aggregation. Given that destabilization of local structure around ³⁰⁶VQIVYK³¹¹ promotes aggregation, stabilizing local structure should rationally mitigate aggregation. By promoting a β -hairpin structure via tryptophan zipper motifs or by using isoelectric forces, a P301L-containing tau peptide had an inhibited propensity to aggregate.

Our data support the hypothesis that local forces are key to preventing aggregation of tau by maintaining specific local structures. Tau is generally considered to be an intrinsically disordered protein, and therefore long-range contacts are unlikely to play a significant role in stability. Published NMR experiments support local structure formation of these regions in tau. Spectra of tau RD (K18; amino acids 244–372) overlaps with a N- and C-terminally expanded tau RD (K32; amino acids 198–394) and even with the splice isoform of tau RD missing repeat 2 (K19; amino acids 244–372 with 275–306 deleted)^{7,53}, suggesting that adding residues and even deleting an entire repeat have minimal effects on the local structure. Thus, the conformations of local structures in tau are disproportionately more important to its properties compared with structured proteins. This suggests that peptide fragment models are a valid surrogate and can encapsulate the most relevant endogenous structural elements for investigating aggregation of tau.

Disease-associated mutations found near tau's amyloid motif, such as P301L or P301S, have no definitive biophysical mechanism but are nevertheless widely used in cell and animal models of disease^{25,26}. Using our peptide, tau RD and FL tau model systems, we observe that key mutations dramatically alter aggregation rates on similar time scales *in vitro* and seed in cell models. We therefore provide an explanation for the toxic-gain-of-function for several mutations.

Previous reports have studied intra-repeat interactions, with the assumption that each repeat functions independently within tau RD³³. Their peptide models have shown a relationship between the length of a peptide fragment and the seeding capacity of tau to define repeat 3 as the minimal sequence necessary to act as a fully functional seed³³. Our model defines a minimal regulatory sequence that limits spontaneous aggregation and suggests that inter-repeat structural elements modulate aggregation propensity. The composition of these inter-repeat sequences, governed by missense mutations, directly impacts the stability of local structures and aggregation propensity. It is tempting to speculate that local structure surrounding each of the four inter-repeat regions plays independent roles in the exposure of amyloid sequences. This modular nature of the tau RD region may explain how these independent regions can lead to distinct structures that define tau prion strains⁴. A more comprehensive structure–function analysis of other repeat interfaces may help explain how each inter-repeat element contributes to the formation of different tau structures.

The expression levels of the two major isoform types of tau in the central nervous system—3R and 4R—are similar in the adult brain²⁹. However, the 3R:4R ratio of aggregate deposits is disproportionately shifted toward 4R in most tauopathies³⁰. Mutations in tau that affect alternative splicing and generate excess 4R isoforms correlate with some genetic tauopathies^{31,32}. The N-terminal flanking sequences leading into ³⁰⁶VQIVYK³¹¹ differ by four amino acids between the two isoforms. We find that these two isoforms have drastically different aggregation propensities in the presence of disease-associated mutations ($t_{1/2} = 7$ h vs $t_{1/2} > 100$ h, respectively). Chimeras of R1R3/R2R3 transition from aggregation-resistant to aggregation-prone as they lose R1 N-terminal flanking character. The ability of an R1-leading strand to mitigate ³⁰⁶VQIVYK³¹¹ aggregation may explain why 4R tau correlates more closely with pathology. Thus, inter-repeat contacts may explain aggregation propensities of tau isoforms in disease. Encouraging data for a tau vaccine targeting a ³⁰⁰HXPGGG³⁰⁴ sequence suggests it is possible to utilize inter-repeat regions to select between pathogenic and non-pathogenic conformations of tau⁵⁶.

Studying the missense mutations in tau has generated valuable disease models^{26,35}; however, the majority of human tauopathies have no observed genetic mutation in tau³⁴. Critical proline residues N-terminal to the amyloid motif can isomerize into *cis* or *trans* rotamers spontaneously or through unidentified cellular mechanisms. We observe that P301 *cis* and *trans* rotamers have

distinct aggregation propensities *in vitro*. In fact, the aggregation kinetics for a *trans* rotamer of P301 are on par with some well-defined disease mutants (N296Δ, V300I). The concept of proline isomerization triggering aggregation into amyloid is not novel, as this is an accepted mechanism of β2-microglobulin aggregation in kidney dialysis amyloidosis⁵⁷. Other proline residues outside of the tau repeat domain have also been proposed to undergo proline isomerization⁴⁹. Our proposed model suggests a possible mechanism whereby WT tau aggregation could be controlled *in vivo*: specific prolyl isomerization events—possibly triggered by cellular proline isomerases—could trigger spontaneous aggregation by modulating inter-repeat structural elements.

We propose that sequences N-terminal to tau's amyloid motif forms local contacts consistent with a β-hairpin-like compact structure. This shields the amyloid motif and mitigates aggregation (Fig. 8). This represents a simple yet comprehensive model of tau aggregation that unifies key observations throughout tau literature. Algorithms that identify potential amyloid-nucleating regions, such as TANGO, have indicated that ~75% of aggregation nucleating regions in the human proteome use two or more “gatekeeper” residues, with proline being the most-common single gatekeeping residue⁵⁸. These gatekeeping residues are more likely than average to be the site of disease-associated missense mutations and are consistent with our identification of gatekeeping residues near tau's amyloid motif. Thus, local flanking sequences and their structural contacts may play an important role in mitigating aggregation propensity in tau and likely other intrinsically disordered proteins. Finally, the identification and characterization of metastable compact structures encompassing ³⁰⁶VQIVYK³¹¹ may itself prove to be a valuable therapeutic target. One might be able to shift the structural rearrangement of tau amyloid motif from exposed (aggregation-prone) to buried (inert) using small molecules, antibodies, or cellular co-factors. Our results indicate that subtle changes in local structure have immense functional ramifications; therefore, small molecules that shift this structural equilibrium modestly may have significant benefits.

Methods

Recombinant full-length tau and tau RD production. We utilized several forms of recombinant tau. The pet28b-tau plasmid encoding full-length WT tau was a kind gift from Dr. David Eisenberg (UCLA). The P301L mutation was introduced using QuikChange (Stratagene) with primers shown in Supplementary Table 3. Each plasmid was transformed into BL21-Gold (DE3) cells. Cells were grown in 1 × Terrific Broth media to OD600 1.4 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 3 h at 37 °C. The cells were harvested and lysed in 50 mM Tris, 500 mM NaCl, 1 mM β-mercaptoethanol, 20 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 7.5, using an Omni Sonic Ruptor 400 at 4 °C. The lysates were centrifuged, and the supernatant was applied to a Ni-NTA column and eluted with 50 mM Tris, 250 mM NaCl, 1 mM β-mercaptoethanol, 300 mM imidazole. Eluting fractions containing tau were desalted into 50 mM MES, 50 mM NaCl, 1 mM β-mercaptoethanol (pH 6.0) by PD-10 column GE. Exchanged fractions were applied to a HiTrap SP HP (GE) and eluted with a 50 mM–1 M NaCl gradient. Tau containing fractions were concentrated on an Amicon-15 concentrator and applied to a Superdex 200 Increase 10/300 GL (GE) and eluted into 1 × PBS (136.5 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM

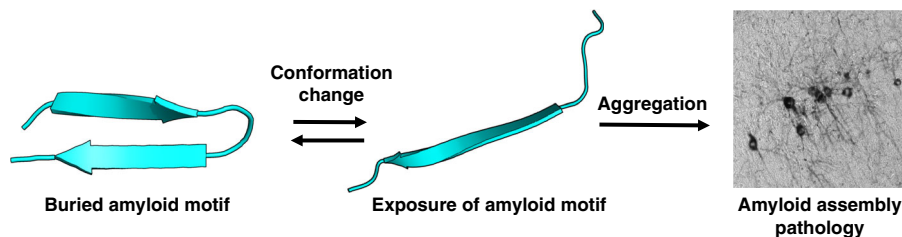


Fig. 8 Molecular model of tau amyloid domain structural rearrangement and subsequent aggregation. Naive tau monomer (left) exists with a propensity to form a relatively collapsed structure, which buries the amyloid domain ³⁰⁶VQIVYK³¹¹. In the presence of disease-associated mutations, proline isomerization events, or certain splice isoforms, the equilibrium is shifted to disfavor local compact structure. This exposes the aggregation-prone ³⁰⁶VQIVYK³¹¹ amyloid motif and enhances aggregation propensity, leading to subsequent tau pathology

KH₂PO₄, pH 7.4). pNG2-tau RD plasmid encoding tau residues 244–380 was a kind gift from Dr. David Eisenberg (ULCA). The P301L and P301S mutations were introduced using Quikchange (Stratagene) with primers shown in Supplementary Table 3. Tau RD wildtype and mutants were expressed the same way as full-length tau. The cells were harvested and lysed in 1 × BRB-80 (80 mM K-PIPES, 1 mM MgSO₄, 1 mM EGTA, pH 6.8), 0.1% β-ME, 1 mM PMSF, DNase (5 units/mL from NEB M0303), and RNase (1 unit/mL from Invitrogen AM2266), using Omni Sonic Ruptor 400 at 4 °C. The lysates were centrifuged, and the supernatant was boiled in a conical tube for 15 min in a water bath. The boiled supernatant was centrifuged at 500 rounds per minute (RPM) for 20 min. The supernatant after centrifugation was filtered using 0.22 μm filter and loaded on HiTrap SP HP (GE) and eluted with a 50 mM–1 M NaCl gradient. Tau RD containing fractions were concentrated on an Amicon-15 concentrator and applied to a Superdex 75 Increase 10/300 GL (GE) and eluted into 1 × PBS (136.5 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4). Aliquots were all stored at –80 °C in 1 × PBS. Tau seeding monomer (M_s) was produced as previously described¹⁶. Specifically, 16 μM WT tau was incubated with heparin (Amsbio) at a 1:1 ratio for 1 h at 37 °C in 1 × PBS. The reaction was resolved on a Superdex 200 Increase 10/300 GL (GE) equilibrated in 1 × PBS. The M_s peak eluted at 12 mL, the concentration of the fraction was measured, the sample aliquoted and flash frozen in liquid nitrogen.

ThT fluorescence aggregation assays. Wild-type or mutant FL tau and tau RD protein was diluted in 1 × PBS with 5 μM β-mercaptoethanol and boiled at 95 °C for 5 min. A final concentration of 4.4 μM heparin (Amsbio) or 33 nM M_s seed was added to 4.4 μM tau or tau RD protein in a 60 μL volume mixed with 25 μM ThT and aliquoted into a 96-well clear bottom plate. Peptides were disaggregated as previously described⁵⁹. In brief, peptides were resuspended in a 1:1 mixture (v/v) of TFA (Pierce) incubated at room temperature (RT) for 1 h. In a chemical fume hood, the peptide solution was dried under a stream of nitrogen gas, and then immediately placed under vacuum to remove any residual volatile solvents. The peptide residue was resuspended in 2 × PBS to a 200 μM concentration to adjust the peptide to buffered reaction conditions. In total, 25 μM ThT was added to 200 μL of 200 μM peptide in a 96-well clear bottom plate. All conditions were done in triplicates (except for the R2R3-IEZip experiment) at RT. ThT kinetic scans were run every 5 min on a Tecan M1000 plate reader at 446 nm Ex (5 nm bandwidth), 482 nm Em (5 nm bandwidth). Blank wells containing buffer and ThT were subtracted from experimental values. Samples producing signal to background (ThT only) with ratios only > 2:1 were considered and these values were normalized to the maximum amplitude in each condition. The data were plotted, and the *t*_{1/2} values were derived from a non-linear regression model fitting in GraphPad Prism.

Transmission electron microscopy. An aliquot of 5 μL of sample was placed onto a glow-discharged Formvar-coated 400-mesh copper grids for 30 s, washed with distilled water, and then negatively stained with 2% uranyl acetate for 1 min. Images were acquired on a Tecnai G² spirit transmission electron microscope (FEI, Hillsboro, OR), serial number: D1067, equipped with a LaB₆ source at 120 kV using a Gatan ultrascan CCD camera.

Tau biosensor cells. Biosensor cells were plated into 96-well plates at 20,000 cells per well. For tau and tau RD experiments, after 5 days of incubation with heparin or M_s, 10 μL of 4.4 μM aggregated protein material was mixed with 1.25 μL lipofectamine and 8.75 μL Opti-MEM, incubated at RT for 30 min, and added to cell media. The “*t* = 0” samples were prepared in the same way but straight from the freezer aliquots. After 2 days, cells were harvested with 0.05% trypsin, then resuspended in Flow buffer (1 × HBSS, 1% FBS, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 × DPBS) and analyzed by flow cytometry. For peptide experiments, 10 μg of aggregated peptide material was added to 0.5 μL lipofectamine and Opti-MEM to a total volume of 10 μL, incubated at RT for 30 min, and added directly to cell media. After 3 days, cells were harvested with 0.05% trypsin, then resuspended in Flow buffer and analyzed by flow cytometry. All conditions were done in triplicates. The Trp Zip biosensor cells expressing the tryptophan zipper motifs flanking the R2R3 element in tau RD were generated as previously described²⁵. In brief, the FM5-YFP and FM5-CFP vectors were digested with NdeI (NEB) and ApoI (NEB). The P301L-Trp Zip tau RD fragment was ordered as a geneblock (IDT) (see Supplementary Table 3). Gibson assembly (NEB) was used to insert the fragment into the plasmid. To produce biosensors, HEK293 T cells were plated at a density of 150,000 cells per well in a 24-well dish. The following day, cells were transduced with tau RD P301S Trp Zip CFP or tau RD P301S Trp Zip YFP lentiviral constructs. Cells were grown in virus-containing media for 72 h before expanding. From a 10-cm dish, cells were harvested with 0.05% trypsin, resuspended in flow cytometry buffer (HBSS plus 1% FBS and 1 mM EDTA), and subjected to FACS (Sony Biotechnology). Populations of CFP and YFP dual-positive cells with a CFP:YFP median fluorescent intensity (MFI) ratio of 1:3.7 (standardized to their relative brightness) were selected to yield a FRET donor: acceptor molar ratio of 1:1. CFP or YFP single-positive cells with an equivalent MFI to dual-positive cells were selected. Following FACS and expansion, single-positive cells were maintained and used as a polyclonal line. Dual-positive cells were used to generate monoclonal lines. Here, cells were plated sparsely in a 10-cm dish and allowed to expand for 10 d, at which time cloning cylinders (Bel-Art Products) were used to isolate single clones. All stable cell lines were amplified, frozen down,

and stored in liquid nitrogen until use. The derived monoclonal biosensor cell lines were empirically tested for best FRET signal to noise, and the same monoclonal cell line was used for all experiments.

Flow cytometry. A BD LSRFortessa was used to perform FRET flow cytometry. To measure CFP and FRET, cells were excited with the 405 nm laser, and fluorescence was captured with a 405/50 nm and 525/50 nm filter, respectively. To measure YFP, cells were excited with a 488 laser and fluorescence was captured with a 525/50 nm filter. To quantify FRET, we used a gating strategy where CFP bleed-through into the YFP and FRET channels was compensated using FlowJo analysis software. The MACSQuant VYB (Miltenyi) was used to perform FRET flow cytometry. To measure CFP and FRET, cells were excited with the 405 nm laser, and fluorescence was captured with a 405/50 nm and 525/50 nm filter, respectively. To measure YFP, cells were excited with a 488 laser and fluorescence was captured with a 525/50 nm filter. To quantify FRET, we used a gating strategy similar to that previously described. In brief, CFP bleed-through into the YFP and FRET channels was compensated using MACSQuantify Software from Miltenyi Biotec. Because some YFP-only cells exhibit emission in the FRET channel, we introduced an additional gate to exclude from analysis cells that exert a false-positive signal in the FRET channel (i.e., false FRET gate). Subsequently, we created a final bivariate plot of FRET vs. CFP and introduced a triangular gate to assess the number of FRET-positive cells. This FRET gate was adjusted to biosensor cells that received lipofectamine alone and are thus FRET-negative. This allows for direct visualization of sensitized acceptor emission arising from excitation of the CFP donor at 405 nm. The integrated FRET density, defined as the percentage of FRET-positive cells multiplied by the median fluorescence intensity of FRET-positive cells, was used for all analyses. For each experiment, 20,000 cells per replicate were analyzed and each condition was analyzed in quadruplicate. Data analysis was performed using FlowJo v10 software (Treestar).

XL-MS sample preparation and mass spectrometry. Preparation of tau RD was cross-linked at a total protein concentration of 1.0 mg/mL using 100 μg of starting material. The cross-linking buffer was 1 × PBS with 3 mM DTT. Five replicates for each condition (37 °C, 50 °C, and 75 °C) were prepared. Samples for 50 °C and 75 °C conditions were equilibrated at the appropriate temperature for 1 h before cross-linking. The cross-linking reaction was initiated by adding DSS stock solution (25 mM DSS-d₀ and -d₁₂, Creative Molecules) in DMF to a final concentration of 1 mM. Samples were further incubated at 37 °C, 50 °C, or 75 °C for 1 min with 350 RPM shaking. Excess reagent was quenched by addition of Tris (pH 7.5) to 100 mM and incubation at 37 °C for 30 min, and subsequently flash frozen by liquid nitrogen and evaporated to dryness by lyophilization. Proteins were resuspended in 8 M urea, reduced with 2.5 mM TCEP (37 °C, 30 min) and alkylated with 5 mM iodoacetamide (30 min, RT, protected from light). The sample solutions were diluted to 1 M urea with 50 mM ammonium hydrogen carbonate and trypsin (Promega) was added at an enzyme-to-substrate ratio of 1:50. Proteolysis was carried out at 37 °C overnight followed by acidification with formic acid to 2% (v/v). Samples were then purified by solid-phase extraction using Sep-Pak tC18 cartridges (Waters) according to standard protocols. Samples were evaporated to dryness and reconstituted in water/acetonitrile/formic acid (95:5:0.1, v/v/v) to a final concentration of ~0.5 μg/μL. In total, 2 μL each were injected for duplicate LC-MS/MS analyses on an Eksigent 1D-NanoLC-Ultra HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system. Peptides were separated on self-packed New Objective PicoFrit columns (11 cm × 0.075 mm I.D.) containing Magic C₁₈ material (Michrom, 3 μm particle size, 200 Å pore size) at a flow rate of 300 nL/min using the following gradient: 0–5 min = 5% B, 5–95 min = 5–35% B, 95–97 min = 35–95% B and 97–107 min = 95% B, where A = (water/acetonitrile/formic acid, 97:3:0.1) and B = (acetonitrile/water/formic acid, 97:3:0.1). The mass spectrometer was operated in data-dependent mode by selecting the five most abundant precursor ions (*m/z* 350–1600, charge state 3+ and above) from a preview scan and subjecting them to collision-induced dissociation (normalized collision energy = 35%, 30 ms activation). Fragment ions were detected at low resolution in the linear ion trap. Dynamic exclusion was enabled (repeat count 1, exclusion duration 30 s).

Analysis of mass spectrometry data. Thermo.raw files were converted to the open.mzXML format using msconvert (proteowizard.sourceforge.net) and analyzed using an in-house version of xQuest⁵⁰. Spectral pairs with a precursor mass difference of 12.075321 Da were extracted and searched against the respective FASTA databases containing tau (TAU_HUMAN P10636-8) or with a P301L/S substitution. xQuest settings were as follows: Maximum number of missed cleavages (excluding the cross-linking site) = 2, peptide length = 5–50 aa, fixed modifications = carbamidomethyl-Cys (mass shift = 57.021460 Da), mass shift of the light crosslinker = 138.068080 Da, mass shift of mono-links = 156.078644 and 155.096428 Da, MS¹ tolerance = 10 ppm, MS² tolerance = 0.2 Da for common ions and 0.3 Da for cross-link ions, search in ion-tag mode. Post-search manual validation and filtering was performed using the following criteria: xQuest score > 25, mass error between –2.2 and +3.8 ppm, %TIC > 10, and a minimum peptide length of six aa. In addition, at least four assigned fragment ions (or at least three contiguous fragments) were required on each of the two peptides in a cross-link. FDRs for the identified cross-links were estimated using xprophet⁶⁰ and estimated to be 1.3–10% (Supplementary Figure 3, Supplementary Data 2). At each

temperature, the five replicate data sets were compared and only cross-links present in five of the five data sets were used to generate a consensus data set (Supplementary Data 3). Cross-link data with information of cross-linked residue positions and nseen (frequency) were visualized using customized gnuplot scripts.

Model generation of tau RD using ROSETTA. The backbone NH, N, CA, CB, and C=O chemical shift assignments for the tau fragment from 243–368 (bmrblid = 19253) were used in CS-Rosetta to generate fragment libraries for subsequent model refinement⁶¹. First, chemical shift parameters were used to predict backbone torsional angles using TALOS to generate a CS-guided fragment library representing the conformations of the protein⁶¹. For the ab initio ROSETTA calculations, the tau RD sequence was used to generate 3-mer and 9-mer fragments derived from the protein data bank using the fragment picker tool⁴⁰. The Rosetta energy function was used to assemble and iteratively refine 5000 structural models using each set of fragments^{40,41,62} (Supplementary Data 5). Ca-based RMSDs were computed in Rosetta for tau RD and hairpin fragments (Supplementary Data 5). Clustering analysis of the tau RD ensemble showed similar results yielding median RMSDs of 19.5 Å and 19.75 Å for ab initio and CS-ROSETTA simulations, respectively (Supplementary Data 5). Ensemble wide calculation of ca–ca end to end distances between residues 264–280, 295–311, 327–343 and 359–375 were carried out using a python script (Supplementary Data 5). All simulations were performed on UTSW's biohpc computing cluster. All plots were generated with gnuplot. Images were created using Pymol.

Peptide synthesis. All peptides were synthesized as ordered by Genscript with N-terminal acetylation and C-terminal amidation modifications. Peptides were purified to >95% purity by FPLC via an Agilent ZORBAX StableBond 250 mm C8 column.

Molecular dynamics simulations. Well-Tempered Metadynamics⁶³ was employed to enable accelerated conformational sampling and to construct the associated free energy surface. Metadynamics was performed on a two-dimensional space of parallel-β sheet content and anti-parallel sheet content. To increase search efficiency in oligomeric space, we have incorporated conformational symmetry constraints, which have been shown to enable sampling of multi-polymer landscapes⁴⁴. The initial dodecahedron simulation box was constructed from a trimer of a randomly unfolded structure of 295–311 by adding 7587 SPC/E explicit waters and three neutralizing Cl ions (one for each monomer). The AMBER99sb-ildn force-field⁶⁴ was used for all simulations. After an initial 1009 steepest descent steps of converged energy minimization, 10 ns of NVT and 20 ns of NPT (first 10 with Berendsen⁶⁵ and the last 10 with Parrinello-Rahman⁶⁶ barostats) equilibrations were performed. The subsequent production level trajectories are based on 5 fs time steps using hydrogen-only virtual sites⁶⁷. Production level trajectories were obtained for an NPT ensemble with Parrinello-Rahman barostat, and periodic boundary conditions with Particle Mesh Ewald (PME)⁶⁸ summation for long-range electrostatics. The tuned well-tempered metadynamics parameters are 10, 1.4 kJ/mol, and 0.3 for bias factor, Gaussian height, collective variable space Gaussian widths, respectively. The Gaussian perturbations were included into MD every 2.5 ps using the PLUMED package⁶⁹ as an external patch to Gromacs-5.0.4⁷⁰. A total of 18 μs trajectories were generated, 9 μs for wildtype and 9 μs for the P301L mutant, over a total of six independent runs. All simulations were done on UTSW's bioHPC computing cluster.

Statistics. All statistics were calculated using GraphPad Prism 8.0. Three independent ThT experiments were run for each condition. The data were normalized to the highest amplitude and averages and standard deviations were plotted. Plots were fitted to a non-linear regression model, from which $t_{1/2}$ values were derived. $t_{1/2}$ error represents a 95% CI. Flow cytometry cell aggregation was conducted in three independent experiments, whose values are plotted. Error bars represent a 95% CI.

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

Data availability

ThT data for tau, tau RD, and peptide experiments is available in Supplementary Data 1. Raw biosensor FRET analysis for tau and tau RD seeding experiments is available in Supplementary Data 2. Raw cross-linking mass spectrometry data are available in Supplementary Data 3 and 4. Summary of ab initio and CS-ROSETTA ensemble and hairpin analysis available in Supplementary Data 5. Summary of MD trajectory energetics are available in Supplementary Data 6. Raw biosensor FRET analysis for synthetic peptide-seeding experiments is available in Supplementary Data 7. Other supporting data available upon reasonable request from the authors.

Code availability

All tau RD simulations were carried out with CS-ROSETTA and ROSETTA (available at <https://www.rosettacommons.org/>). All peptide MD simulations were performed using Gromacs-5.0.4 (available at <http://www.gromacs.org>). The metadynamics runs are carried out with the PLUMED-2.1.2 interface to Gromacs (available at <http://www.plumed.org>)

Received: 22 February 2019 Accepted: 8 May 2019

Published online: 07 June 2019

References

- Cleveland, D. W., Hwo, S. Y. & Kirschner, M. W. Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *J. Mol. Biol.* **116**, 227–247 (1977).
- Eliezer, D. et al. Residual structure in the repeat domain of tau: Echoes of microtubule binding and paired helical filament formation. *Biochemistry* **44**, 1026–1036 (2005).
- Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **56**, 343 (2017).
- Sanders, D. W. et al. Distinct Tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* **82**, 1271–1288 (2014).
- Sawaya, M. R. et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
- von Bergen, M. et al. Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local beta-structure. *J. Biol. Chem.* **276**, 48165–48174 (2001).
- Mukrasch, M. D. et al. Sites of tau important for aggregation populate {beta}-structure and bind to microtubules and polyanions. *J. Biol. Chem.* **280**, 24978–24986 (2005).
- Falcon, B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
- Zhang, X. et al. RNA stores tau reversibly in complex coacervates. *PLoS Biol.* **15**, e2002183 (2017).
- Ismail, T. & Kanapathipillai, M. Effect of cellular polyanion mimetics on tau peptide aggregation. *J. Pept. Sci.* **24**, e3125 (2018).
- Kuret, J. et al. Evaluating triggers and enhancers of tau fibrillation. *Microsc. Res. Tech.* **67**, 141–155 (2005).
- Zhao, J. et al. Glycan determinants of Heparin-Tau interaction. *Biophys. J.* **112**, 921–932 (2017).
- Kellogg, E. H. et al. Near-atomic model of microtubule-tau interactions. **1780**, 1–9 (2018).
- Mok, S.-A. et al. Mapping interactions with the chaperone network reveals factors that protect against tau aggregation. *Nat. Struct. Mol. Biol.* **25**, 384–393 (2018).
- Baughman, H. E. R., Clouser, A. F., Kleiv, R. E. & Nath, A. HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. *J. Biol. Chem.* **293**, 2687–2700 (2018).
- Mirbaha, H. et al. Inert and seed-competent tau monomers suggest structural origins of aggregation. *eLife* **7**, pii: e36584 (2018).
- Sharma, A. M., Thomas, T. L., Woodard, D. R., Kashmer, O. M. & Diamond, M. I. Tau monomer encodes strains. *eLife* **7**, pii: e37813 (2018).
- Elbaum-Garfinkle, S. & Rhoades, E. Identification of an aggregation-prone structure of tau. *J. Am. Chem. Soc.* **134**, 16607–16613 (2012).
- Chirita, C. N., Congdon, E. E., Yin, H. & Kuret, J. Triggers of full-length tau aggregation: a role for partially folded intermediates. *Biochemistry* **44**, 5862–5872 (2005).
- Walker, S., Ullman, O. & Stultz, C. M. Using intramolecular disulfide bonds in tau protein to deduce structural features of aggregation-resistant conformations. *J. Biol. Chem.* **287**, 9591–9600 (2012).
- Kadavath, H. et al. Folding of the Tau protein on microtubules. *Angew. Chem. Int. Ed. Engl.* **54**, 10347–10351 (2015).
- Mukrasch, M. D. et al. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* **129**, 5235–5243 (2007).
- Mukrasch, M. D. et al. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* **7**, e34 (2009).
- Alzforum Mutation Databse. <https://www.alzforum.org/mutations/mapt>. Accessed 2018.
- Holmes, B. B. et al. Proteopathic tau seeding predicts tauopathy in vivo. *Proc. Natl Acad. Sci. USA* **111**, E4376–E4385 (2014).
- Yoshiyama, Y. et al. Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron* **53**, 337–351 (2007).
- Fischer, D. et al. Structural and microtubule binding properties of tau mutants of frontotemporal dementias. *Biochemistry* **46**, 2574–2582 (2007).
- Karamanos, T. K., Kalverda, A. P., Thompson, G. S. & Radford, S. E. Mechanisms of amyloid formation revealed by solution NMR. *Prog. Nucl. Magn. Reson Spectrosc.* **88–89**, 86–104 (2015).
- Goedert, M., Wischik, C. M., Crowther, R. A., Walker, J. E. & Klug, A. Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. *Proc. Natl Acad. Sci. USA* **85**, 4051–4055 (1988).

30. Williams, D. R. Tauopathies: Classification and clinical update on neurodegenerative diseases associated with microtubule-associated protein tau. *Intern. Med. J.* **36**, 652–660 (2006).
31. Schoch, K. M. M. et al. Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse Model. *Neuron* **90**, 941–947 (2016).
32. Hutton, M. et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702–705 (1998).
33. Stöhr, J. et al. A 31-residue peptide induces aggregation of tau's microtubule-binding region in cells. *Nat. Chem.* **9**, 874–881 (2017).
34. Rizzu, P. et al. High prevalence of mutations in the microtubule-associated protein tau in a population study of frontotemporal dementia in the Netherlands. *Am. J. Hum. Genet.* **64**, 414–421 (1999).
35. Mocanu, M.-M. et al. The potential for beta-structure in the repeat domain of tau protein determines aggregation, synaptic decay, neuronal loss, and coassembly with endogenous Tau in inducible mouse models of tauopathy. *J. Neurosci.* **28**, 737–748 (2008).
36. Furman, J. L., Holmes, B. B., Diamond, M. I. Sensitive Detection of Proteopathic Seeding Activity with FRET Flow Cytometry. *J. Vis. Exp.* **106**, e53205 (2015).
37. Leitner, A. et al. The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Struct./Fold. Des.* **20**, 814–825 (2012).
38. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).
39. Rinner, O. et al. Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318 (2008).
40. Ovchinnikov, S., Park, H., Kim, D. E., DiMaio, F. & Baker, D. Protein structure prediction using Rosetta in CASP12. *Proteins* **86**, 113–121 (2018).
41. Lange, O. F. et al. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl Acad. Sci. USA* **109**, 10873–10878 (2012).
42. Mylonas, E. et al. Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry* **47**, 10345–10353 (2008).
43. Falcon, B., Zhang, W., Murzin, A. G., Murshudov, G. & Holly, J. Structures of filaments from Pick's disease reveal a novel tau protein fold **561**, 137–140 (2018).
44. Lin, M. M. Leveraging symmetry to predict self-assembly of multiple polymers. *Chem. Phys. Lett.* **683**, 347–351 (2017).
45. Kaufman, S. K. et al. Tau prion strains dictate patterns of cell pathology, progression rate, and regional vulnerability in vivo. *Neuron* **92**, 796–812 (2016).
46. Kier, B. L., Shu, I., Eidschink, L. A. & Andersen, N. H. Stabilizing capping motif for beta-hairpins and sheets. *Proc. Natl Acad. Sci. USA* **107**, 10466–10471 (2010).
47. Vogel, M., Bukau, B. & Mayer, M. P. Allosteric regulation of Hsp70 chaperones by a proline switch. *Mol. Cell* **21**, 359–367 (2006).
48. Gustafson, C. L. et al. A slow conformational switch in the BMAL1 transactivation domain modulates circadian rhythms. *Mol. Cell* **66**, 447–457.e7 (2017).
49. Pastorino, L. et al. The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. *Nature* **440**, 528–534 (2006).
50. Torbeev, V. Y. & Hilvert, D. Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate 2-microglobulin amyloid assembly. *Proc. Natl Acad. Sci. USA* **110**, 20051–20056 (2013).
51. Lim, J. et al. Pin1 has opposite effects on wild-type and P301L tau stability and tauopathy. *J. Clin. Investig.* **118**, 1877–1889 (2008).
52. Zhang, X. et al. RNA stores tau reversibly in complex coacervates. *bioRxiv* 111245, <https://doi.org/10.1101/111245> (2017).
53. Mukrasch, M. D. et al. The 'jaws' of the tau-microtubule interaction. *J. Biol. Chem.* **282**, 12230–12239 (2007).
54. Meng, W., Lyle, N., Luan, B., Raleigh, D. P. & Pappu, R. V. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc. Natl Acad. Sci. USA* **110**, 2123–2128 (2013).
55. Ding, F., Jha, R. K. & Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure* **13**, 1047–1054 (2005).
56. Kontseikova, E., Zilkka, N., Kovacech, B., Skrabana, R. & Novak, M. Identification of structural determinants on tau protein essential for its pathological function: novel therapeutic target for tau immunotherapy in Alzheimer's disease. *Alzheimers Res. Ther.* **6**, 45 (2014).
57. Eichner, T. & Radford, S. E. A generic mechanism of B2-microglobulin amyloid assembly at neutral pH involving a specific proline switch. *J. Mol. Biol.* **386**, 1312–1326 (2009).
58. Reumers, J., Maurer-Stroh, S., Schymkowitz, J. & Rousseau, F. D. Protein sequences encode safeguards against aggregation. *Hum. Mutat.* **30**, 431–437 (2009).
59. O'Nuallain, B. et al. Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. *Methods Enzymol* **413**, 34–74 (2006).
60. Walzthoeni, T. et al. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9**, 901–903 (2012).
61. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
62. Raman, S. et al. NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018 (2010).
63. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
64. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
65. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
66. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
67. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).
68. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
69. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
70. Abraham, M. J. et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

Acknowledgements

This work was supported by grants from the Tau Consortium and NIH grants awarded to 1R01NS071835 (M.I.D.), the Effie Marie Cain Endowed Scholarship (L.A.J.) and the Heising-Simons Foundation (M.M.L.). We appreciate the help of the Structural Biology Laboratory and Proteomics Core Facility at the University of Texas Southwestern Medical Center.

Author contributions

K.D., D.C., M.I.D. and L.A.J. conceived and designed the overall study. D.C. performed in vitro protein assays, cell models, cross-link mass spectrometry, and ROSETTA simulations. K.D. performed in vitro peptide assays, flow cytometry, and cell models. Z.H. developed the tau in vitro seeded aggregation assays. O.M.K. assisted with tau and tau RD purifications. V.A.P. helped with cell seeding assays. B.D.R. and D.R.W. performed electron microscopy. L.S. and M.M.L. performed MD simulations. K.D., D.C. and L.A.J. wrote the manuscript, and all authors contributed to its improvement.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-10355-1>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019