MECHANISTIC INVESTIGATION INTO THE REGULATION OF AMYLOID MOTIFS

IN TAU AGGREGATION AND DISEASE

APPROVED BY SUPERVISORY COMMITTEE

Lukasz Joachimiak, Ph.D., Supervisor

Milo Lin, Ph.D., Chair

Marc Diamond, Ph.D.

Michael Rosen, Ph.D.

Jose Rizo-Rey, Ph.D.

DEDICATION

I dedicate this dissertation to my family, my mentors and many who had offered me help throughout the journey.

I thank my mom, Jian and my stepdad, Richard, who I know I could always fall back on.

I thank my mentor, Lukasz for his mentorship, support and the always opened door for any obstacles I encounter.

I thank my family in China, Aunt Jing and Ping, Uncle Yupu and Aijun, and my grandparents, whose confidence in me got me to where I am today. I also thank my family in the US for being inclusive and caring.

I thank all the members of the Joachimiak lab for being kind and accommodating. I thank Marc, Milo, Jose and Mike for their valuable insights and suggestions on my research.

I thank Sofia for great time on our hiking and ski trips. I appreciate Jaime for being my role model in research. I thank Vish for being always willing to troubleshoot my computational problems. I also thank Sushobna for our food and dessert explorations in DFW.

I thank the late Ze for inspiring me to work hard and pursue my goals fiercely.

I thank Kane for his support during the job interview season.

I thank Alex for getting me through the rough times during qual.

I thank the Andy from the Proteomics Core for being supportive.

I thank Amy Haughey for being the best program advisor.

MECHANISTIC INVESTIGATION INTO THE REGULATION OF AMYLOID MOTIFS IN TAU AGGREGATION AND DISEASE

by

DAILU CHEN

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2022

MECHANISTIC INVESTIGATION INTO THE REGULATION OF AMYLOID MOTIFS IN TAU AGGREGATION AND DISEASE

Publication No. _____

Dailu Chen, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2022

Supervising Professor: Lukasz Joachimiak, Ph.D.

PREFACE

Amyloid formation of tau protein is a unifying theme in a multitude of neurodegenerative diseases, collectively called tauopathies. Missense mutations in the tau gene (*MAPT*) correlate with aggregation propensity and cause dominantly inherited tauopathies, but the molecular mechanism of how they promote tau assembly into amyloids is poorly understood. Many disease-associated mutations localize within tau's repeat domain proximal to amyloidogenic sequences, such as ^{306}VQIVYK^{311}. We use computational modeling,

recombinant protein and synthetic peptide systems, cross-linking mass spectrometry, and cell models to investigate the biophysical mechanisms behind aggregation of the P301L/S and S320F mutants. We conclude that the aggregation prone ^{306}VQIVYK^{311} motif forms metastable compact structures with its upstream sequence that modulates aggregation propensity. We report that disease-associated mutations such as P301L/S at the inter-repeat interface, isomerization of a critical proline, or alternative splicing are all sufficient to destabilize this local structure and trigger aggregation. In the study of S320F, we again demonstrate the role of local protective structures in the regulation of tau aggregation driven by amyloid motif and uncover that the S320F mutation allosterically exposes ^{306}VQIVYK^{311} by retaining one of the protecting motifs in a stabilized local hydrophobic cluster. We show that rational design of this nonpolar cluster centered on position 320 based on tauopathy fibril structures maintains a spontaneous aggregation phenotype revealing new principles that govern tau aggregation. We uncover a nuanced balance of local protective structures that sequester amyloid motifs and how introduction of a hydrophobic mutation redistributes these interactions to drive spontaneous aggregation. Tau presents as an aggregation-resistant monomer and only in the presence of an inducer such as heparin, RNA or other polyanion will tau aggregate in vitro. Our studies use disease-causing mutations as a bridge to explain the basis of early conformational changes that may underlie genetic and sporadic tau pathogenesis. Our findings provide molecular insights into regulation of tau assembly, and we anticipate deeper knowledge of this process will begin control of tau aggregation into discrete structural polymorphs.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Chen, D., Drombosky, K. W., Hou, Z., Sari, L., Kashmer, O. M., Ryder, B. D., Perez, V. A., Woodard, D. R., Lin, M. M., Diamond, M. I., Joachimiak, L. A. (2019). Tau local structure shields an amyloid-forming motif and controls aggregation propensity. Nat Commun, 10(1), 2493. doi:10.1038/s41467-019-10355-1

Hou, Z., **Chen, D.**, Ryder, B. D., & Joachimiak, L. A. (2021). Biophysical properties of a tau seed. Sci Rep, 11(1), 13602. doi:10.1038/s41598-021-93093-z

Mirbaha, H., **Chen, D.**, Morazova, O. A., Ruff, K. M., Sharma, A. M., Liu, X., Goodarzi, M., Pappu, R. V., Colby, D. W., Mizaei, H., Joachimiak, L. A., Diamond, M. I. (2018). Inert and seed-competent tau monomers suggest structural origins of aggregation. Elife, 7. doi:10.7554/eLife.36584

Mirbaha, H., **Chen, D.**, Mullapudi, V., Terpack, S. J., White, C. L., Joachimiak, L. A., & Diamond, M. I. (2022). Seed-competent tau monomer initiates pathology in PS19 tauopathy mice. bioRxiv, 2022.2001.2003.474806. doi:10.1101/2022.01.03.474806

CHAPTER ONE
Introduction

TAU AND DISEASE

Sequence and Structure of Tau

Protein aggregation is often correlated with diseases, especially neurodegenerative diseases, such as tauopathies. Tauopathies are a group of neurodegenerative diseases featured by pathological aggregation of tau protein into insoluble amyloid fibrils in the brain. The most common tauopathy is the Alzheimer's disease (AD) for which to-date no effective FDA approved drugs are available to prevent the disease progression. There are 6 isoforms of tau in the central nervous system (CNS), all involved in different tauopathies and at various degrees. The 6 isoforms have sequence length of ranges from 352 to 441 residues. The longest isoform, isoform 2, which is often called the full-length tau, is composed of 441 residues and contains 2 N-terminal domains, a Proline-rich domain, a repeat domain (RD) composed of 4 imperfect repeats, and a C-terminal domain. The other 5 isoforms differ from isoform 2 from the number of N-terminal domains and the number of repeats in the repeat domain it harbors. The N-terminal, the Proline-rich domain, and the C-terminal tail are found intrinsically disordered, while the RD is relatively less disordered. Nuclear magnetic resonance (NMR) secondary chemical shifts indicate while the majority of regions in tau appears as random coil, the RD shows the highest propensity for β-structures (Mukrasch, 2007).

Tau Aggregation and Amyloid Motifs

The RD composes the core of tau amyloid fibrils. In tau isoform 2, there are 2 hexapeptide motifs (^{275}VQIINK^{280} and ^{306}VQIVYK^{311}) at the beginning of repeat 2 (R2) and repeat 3 (R3) centering the aggregation-prone region (Wischik, 1988). These two hexapeptide motifs adopt β-sheet structures upon aggregation into amyloid fibrils (Daebel, 2012). Although tau is involved in aggregation in vivo, due to its high solubility, tau is difficult to aggregate spontaneously in vitro, unless induced by a poly-anionic cofactor such as heparin or RNA (Wang, 2007). The RD alone aggregates more readily than the full-length tau, but it still requires an aggregation inducer to initiate self-assembly (Stöhr, 2017). However, the 2 hexapeptide motifs by themselves are sufficient to self-assemble rapidly in vitro without any aid (Sawaya, 2007), herein referred to as amyloid motifs. This indicates that regions outside of the amyloid motifs can be inhibitory of aggregation driven by the amyloid motifs. Double Electron-Electron Resonance data detected an expansion of local conformation at the regions of the 2 hexapeptide motifs in tau upon binding of heparin (Eschmann, 2017). This suggests the unfolding of flanking regions from the amyloid motifs might underlie the aggregation mechanism of tau in vitro and potentially in vivo. Study from our lab further characterized the importance of the local structure around amyloid motifs in regulating tau aggregation and proposed a model where the exposure of the amyloid motifs from flanking motifs drives tau aggregation (Chen, 2018).

The recent breakthroughs in Cryo-Electron Microscopy (Cryo-EM) have provided unprecedented insights into tau filament formation in disease. Using Cryo-EM, the Scheres group from MRC has determined structures of tau filaments derived from human brains of

the Alzheimer's Disease (AD) (Fitzpatrick, 2017; Falcon 2018), Pick's Disease (PiD) (Falcon, 2018), Corticalbasal Degeneration (CBD) (Zhang, 2020), Progressive Supranuclear Palsy (PSP), Chronic Traumatic Encephalopathy (CTE) (Falcon, 2019), Globular Glial Tauopathy (GGT) (Shi, 2021) and Argyrophilic Grain Disease (AGD) (Shi, 2021). It was found that different tau filament fold was adopted in different diseases and the particular fold was conserved in individuals of the same disease (Scheres, 2020). This provides strong molecular basis for the design of potential diagnostics and therapeutics to distinguish and treat different tauopathies given the stage where tau filaments present at detectable concentrations is not too late for any intervention. Heparin-induced in vitro tau filament on the other hand, showed distinct shapes than filaments from disease (Zhang, 2019). This further raises the skepticism on the translatability and significance of using heparin induced tau fibrils to study disease.

Looking at the atomic resolution, the amyloid motif ^{306}VQIVYK^{311} was involved in the core of the tau fibrils consistently across all tauopathies studied by the Scheres group. In the monomer of fibril structures, ^{306}VQIVYK^{311} was found in close interaction with other hydrophobic patches, for example, ^{375}KLTF^{378} in the AD and CTE structure, or ^{336}QVEVK^{340} in the PiD and CBD structure. In addition, several other similar hydrophobic patches were also observed in the fibril structure of each disease. The presence of these hydrophobic patches potentially contributes to the stability of fibril structures and might explain the formation of different shapes of tau filament. Apart from the high-resolution solved fibril core, there were unknown densities at the periphery of the fibril core from each disease. Some densities named as the "fuzzy coat" come from the more disordered flanking sequence.

Some fit well with post-translational modifications such as acetylation, ubiquitination and phosphorylation. Others are undetermined co-factors that formed part of the interior of amyloid filaments, such as those in CTE and CBD structures. This evidence suggests that the formation of distinct fibril shape might be resulted from a combination of isoform composition, post-translational modifications and co-factors.

Effect of phosphorylation on tau

Post-translational modifications (PTMs) including phosphorylation, acetylation, ubiquitination have long been identified to present on tau amyloid fibrils. Among these, phosphorylation has been studied most extensively as hyperphosphorylated tau is strongly correlated with AD. Tau fibrils derived from post-mortem AD brains are found in a hyperphosphorylated form (Grundke-Iqbal, 1986; Kopke 1993). Tau isolated from normal adult human brain contains 2-3 moles of phosphate per mole of tau, whereas AD-associated tau displayed 3- to 4-fold higher phosphorylation level (Kenessey, 1993; Kopke 1993), which means on an AD-associated tau molecule, about 6 – 12 sites are phosphorylated.

Under physiological conditions, phosphorylation activity regulates the interaction of tau with microtubules. The RD together with the Proline-rich domain is the main component of tau that binds microtubule, specially at the hydrophobic pocket between the tubulin heterodimers (Kadavath, 2015) and the binding of tau is hypothesized to stabilize microtubule as well as to promote polymerization of tubulins. Abnormal hyperphosphorylation has been demonstrated to inhibit tau from binding to microtubules (Alonso, 2001). As the ability of tau to bind microtubule is hypothesized to be essential for

normal neuronal function and the instability of microtubule can potentially cause disease, hyperphosphorylation of tau is linked as a potential cause of AD. However, whether hyperphosphorylation is the cause or correlation or even by-product has not been well demonstrated.

There are 39 phosphorylation sites (37 Ser/Thr and 2 Tyr) identified in tau from the post-mortem AD brain through mass-spectrometry (Hanger, 2007). With on average 6-12 sites present in each tau molecule from AD brain, there are diverse combinations of phosphorylation sites on AD-associated tau. Studies of abnormal phosphorylation effect on aggregation propensity have showed various conclusions due to difference in positions that are phosphorylated as well as methods used to introduce phosphorylation (Hasse, 2004; Liu, 2008).

The hypothesis that abnormal phosphorylation drives aggregation can be attributed to a theory that phosphorylation at the flanking regions of RD potentially neutralizes the basic charges around the aggregation-prone region in the RD, which will reduce the repulsion in the intermolecular interaction and thus promote tau self-assembly (Alonso, 2001). This proposed aggregation mechanism is reminiscent of the poly-anion induced aggregation in that they both neutralize the charges flanking the aggregation core. However, there is still a knowledge gap in the detailed structural mechanism of how abnormal phosphorylation neutralizes the flanking regions of the aggregation core and hence promotes the aggregation.

With the structures of tau disease fibrils solved, a closer look at the potential mechanism of phosphorylation driving aggregation was made possible. Nonetheless, no phosphorylation was observed within the fibril core of any of the tau fibrils, despite the

possibility of them decorating on the periphery. This poses the question whether hyperphosphorylation actually happens before or after the formation of fibrils. Due to its intrinsic disordered nature, tau is a substrate to an array of kinases and phosphatases and harbors 40 phosphorylatable sites. It is not difficult to imagine the phosphorylation pattern can be very heterogenous and the process can be very dynamic. Although we cannot exclude that a pool of tau phosphorylated at one or several particular sites can act as seeds and drive aggregation, the likelihood is low and to identify this particular pattern is essentially impossible with currently available techniques. Together, it is still unclear whether hyperphosphorylation is a cause, a correlation or a consequence of tau aggregation in disease and the determination of it will not be trivial.

Disease-Associated Mutations in Tau

While the effect of phosphorylation on tau can be non-trivial to study, disease-associated mutations and their structural and functional impact can provide a straightforward model for understanding how tau is involved in disease. To date, 62 disease-associated mutations in tau have been identified, which cause inherited forms of tauopathy. Several functional impacts from mutations have been characterized which can be grouped into decreasing the ability of microtubule binding and assembly, altering splice isoform ratio, and enhancing tau aggregation kinetics. However, the biophysical mechanism of how these mutations for example enhance aggregation kinetics is unclear.

Many of the missense mutations linked to tauopathy in humans occur within tau RD and cluster near the inter-repeat interfaces, for example P301L and P301S at the interface of

R2 and R3. These regions are interestingly also where amyloid motif or its homologs locates, such as ^{306}VQIVYK^{311} at the interface of R2 and R3. In Chapter Three, my colleagues and I characterized the biophysical mechanism of tau aggregation driven by P301L/S mutations using a combination of computational simulations, crosslinking mass spectrometry, in vitro and in cell aggregation assays. It is proposed that the aggregation driven by amyloid motif is regulated by the flanking sequence and the exposure of amyloid motif leads to aggregation of tau.

Although a great number of mutations promote tau aggregation, most of them still require an inducer, most commonly heparin, to initiate aggregation. Cryo-EM structures of heparin-induced tau fibrils revealed conformational heterogeneity in contrast to tauopathy conformations which adopted defined homogeneous conformations in each disease. This poses a skepticism in the value of using the simple recombinant system to shed light on the mechanism of tau aggregation in disease. In Chapter Three, I show a manuscript in preparation where a mutation that causes spontaneous tau aggregation without inducer, S320F, was leveraged to understand the mechanism of tau aggregation under a more physiologically relevant condition.

CHAPTER TWO

CROSS-LINKING MASS SPECTROMETRY ANALYSIS OF METASTABLE COMPACT STRUCTURES IN INTRINSICALLY DISORDERED PROTEINS

Abstract

Protein assembly into beta-sheet rich amyloids is a common phenomenon in neurodegenerative diseases including Alzheimer's Disease (AD) and Parkinson's Disease (PD). The proteins implicated in amyloid deposition are often intrinsically disordered proteins (IDPs) and are characterized by not folding into a defined globular conformation. The amyloidogenic properties of IDPs are determined by the presence of short sequence elements, referred to as amyloid motifs, that drive ordered aggregation(Goldschmidt, Teng, Riek, & Eisenberg, 2010; Thompson et al., 2006). The microtubule associated protein tau adopts amyloid assemblies in over 20 different diseases commonly referred to as tauopathies. However, native tau is extremely aggregation resistant despite encoding at least three amyloid motifs(Chen et al., 2019). Recent cryogenic electron microscopy (cryo-EM) structures of tau amyloid fibrils isolated from patient brains showed the involvement of amyloid motifs in the fibril core(Falcon et al., 2018; Fitzpatrick et al., 2017; Zhang et al., 2020). How does tau change from an aggregation-resistant state to an aggregation-prone state? Consistent with the fibril structures, we hypothesize that tau protein must change conformation to expose the amyloid motifs that allow self-association into beta-sheet-rich aggregates. This would suggest that the amyloid motifs are likely buried in native tau to

prevent self-assembly. We developed an approach that couples cross-linking mass spectrometry (XL-MS) with temperature denaturation to probe the loss of contacts as a proxy to measure protein unfolding with sequence resolution. Using this approach, we demonstrated that disease-associated mutations in tau located near an amyloid motif disrupt the protective local structure, promote amyloid motif exposure, and thus lead to aggregation(Chen et al., 2019). In this chapter, we describe the detailed protocol for this approach. We anticipate that our protocol can be generalized to other IDPs and will help discover critical structural elements to better understand important biological questions including protein aggregation.

Introduction

Intrinsically disordered proteins (IDPs) are prevalent in the human proteome, many of which are involved in important biological functions, such as signal transduction, transcriptional, and translational processes, as well as diseases including cancer and neurodegenerative diseases(Dyson & Wright, 2005). Alpha-synuclein and the microtubule associated protein tau encode intrinsically disordered regions which are important for their biological function but also can adopt insoluble ordered beta-sheet rich structures associated with PD, and AD, respectively. It is not clear how these proteins perform their biological functions while avoiding self-assembly. Although disordered proteins do not adopt a compact tertiary structure, they rarely exhibit pure random coil character under physiological conditions. Instead, they feature residual local structures. Conventional structural and biophysical methods are limited in capturing weak and transient structures. Solution nuclear

magnetic resonance (NMR) is the most powerful method but transient interactions in IDPs remain difficult to detect due to weak sensitivity. Therefore, it requires high sample concentrations which are typically not compatible with aggregation-prone proteins. Typical NMR experiments thus require data acquisition in non-physiological conditions such as low pH or low temperatures and high concentrations to facilitate data acquisition over long time periods.

We have developed a method that combines crosslinking mass-spectrometry (XL-MS) with heat denaturation to allow capture of transient local secondary structures as well as global contacts to infer the stability of these interactions. Chemical crosslinkers can trap interactions between reactive amino acids and when coupled with mass spectrometry can identify points of contact and their frequency. XL-MS has typically been used to probe the structure of large hetero-multimeric complexes to infer topologies and to assist structural determination in cryo-EM or X-ray crystallography. More recently it has been applied to study the effects of mutations the dynamics of IDPs(Chen et al., 2019). Temperature-based unfolding of proteins has classically been studied in the context of circular dichroism (CD) where loss of global secondary structure signal changes as a function of temperature can be used to derive ΔG of unfolding(Nölting et al., 1997). These CD experiments, however, yield very little information about what specific contacts are lost. Thus, coupling XL-MS with heat denaturation helps better understand the energetic landscape of the different contacts in IDPs not easily captured with other methods.

As our model system to study IDP dynamics, we employed the repeat domain fragment of tau (herein tauRD). The tauRD composed of four imperfect repeats encodes the minimal

fragment of tau that binds to microtubules. This fragment also encompasses the sequence composing the core of tau fibrils isolated from tauopathy patient brains(Falcon et al., 2018; Fitzpatrick et al., 2017; Zhang et al., 2020). Each repeat terminates in a P-G-G-G beta-turn stabilizing motif and precedes short sequence motifs essential to drive aggregation. These short sequences are termed "amyloid motifs" and highlighted by the ^{306}VQIVYK^{311} sequence, which is a central driver of tau aggregation (**Fig. 1a**). Surprisingly, native tau is aggregation-resistant and we proposed that beta-hairpins can engage amino acids preceding the amyloid motifs thereby regulating their exposure and thus aggregation propensity. We recognized that the disease-associated mutations also localize near the inter-repeat interfaces in proximity to the amyloid motifs(Chen et al., 2019). These pathogenic mutations cause early onset tauopathies and promote tau aggregation *in vitro* and in cells(Chen et al., 2019). In our beta-hairpin model, we proposed that pathogenic mutations would disrupt the local structure and expose the amyloid motifs. This model was tested and supported by a series of experiments and simulations including the development of the XL-MS assay that probed the unfolding of WT and mutant forms of tau(Chen et al., 2019).

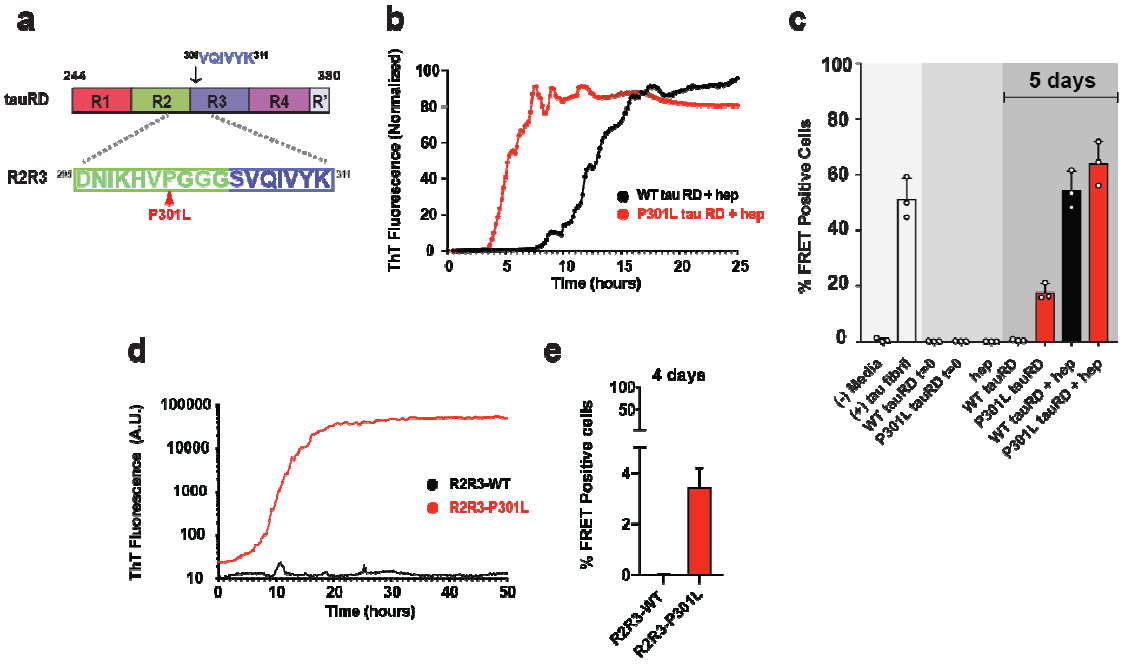


Figure 1. Mutation at the inter-repeat interface drives aggregation in vitro and in cells. **(a)** tauRD (repeat 1 = red; repeat 2 = green; repeat 3 = blue; repeat 4 = purple). Amyloidogenic sequence ^{306}VQIVYK^{311} locates at the R2R3 interface. **(b)** WT tauRD and mutant P301L and tauRD at a 4.4 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 and P301L tauRD in the absence of heparin yielded no detectible ThT signal change over the course of the experiment. **(c)** After 120 h of in vitro incubation, proteins from previous ThT experiments were transduced into tau biosensor cells via lipofectamine. FRET signal from each condition (tauRD-CFP/tauRD-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. **(d)** 200 µM of R2R3-WT and R2R3-P301L peptide were allowed to aggregate in the presence of ThT at room temperature. **(e)** After 96 h of in vitro incubation, peptides from previous ThT experiments were transduced into tau biosensor cells via lipofectamine and FRET signal was measured as described in **(c)**.

In our study, we compared the structural dynamics between wildtype (WT) and P301L tauRD. P301L is a Frontotemporal dementia-associated mutation and occurs directly upstream of ^{306}VQIVYK^{311} within the P-G-G-G motif (**Fig. 1a**). As expected, we observed that P301L tauRD ($t_{1/2}$ = 5.2 ± 0.1 h) aggregated more rapidly when induced with heparin compared WT tauRD ($t_{1/2}$ = 12.5 ± 0.2 h) in a Thioflavin T aggregation assay (ThT) (**Fig. 1b**).

Next, we employed tau biosensor HEK293 cells that stably express tauRD (P301S) fused to cyan or yellow fluorescent proteins(Holmes et al., 2014). These cells sensitively report a fluorescence resonance energy transfer (FRET) signal (tauRD-CFP/tauRD-YFP) only when aggregated in response to tau amyloid seeds(Furman, Holmes, & Diamond, 2015). The tau biosensor cells responded to heparin-treated WT tauRD and showed an increase in seeding activity for the heparin-treated P301L mutant (**Fig. 1c**). The P301L tauRD alone after 5 days also showed a hint of seeding activity (**Fig. 1c**). Additionally, informed by modeling studies, we proposed that a minimal 17 amino acid fragment encoding the P-G-G-G turn and the ^{306}VQIVYK^{311} element (herein termed R2R3) would be sufficient to form beta-hairpin-like conformations and thus control aggregation(Chen et al., 2019) (**Fig. 1a**). Indeed, the R2R3-WT peptide did not aggregate over the course of the experiment while the R2R3-P301L mutant aggregated rapidly (**Fig. 1d**). Consistent with these findings, transduction of these samples into tau biosensors cells revealed seeding for the R2R3-P301L but none for R2R3-WT (**Fig. 1e**). We proposed that the mutation disrupts the beta-turn and therefore the stability of local structure, which results in an exposure of the amyloid motif to drive aggregation. We used XL-MS to probe change in structure comparing WT and P301L tauRD focusing on local structural rearrangements in proximity to the 301 site but also global rearrangements. Pre-equilibrated WT and P301L tauRD samples were reacted with disuccinimidyl suberate (DSS) crosslinker for 1 min at 37 °C, 50 °C or 75 °C and immediately quenched. Samples were confirmed to be monomeric following crosslinking using SDS-PAGE (**Fig. 2**). Cross-linked samples were processed as five replicates, analyzed by mass spectrometry, and the resulting spectra were searched against tauRD and a decoy sequence using xQuest(Rinner et

al., 2008) to identify intramolecular crosslink pairs with low false discovery rates (FDR). We reported a consensus set of crosslinks that are identified across all five independent replicates with average frequencies of detection for each pair as a proxy for stability of structure. We summarized the number of unique consensus crosslinks into 3 sectors: N-terminal short-range contacts, C-terminal short-range contacts, and N- to C-terminal long-range contacts (**Fig. 3a**). The individual crosslinked contacts were illustrated in the contact map color-coded by the frequency of each crosslink and compared to an exhaustive set of theoretical lysine-lysine pairs (**Fig. 3b**). The experimentally observed crosslinks sampled a subset of theoretical pairs indicated discrete mode of contacts in tauRD. As expected, recombinant WT tauRD showed a gradual loss of long-range contacts with increases in temperature, accompanied by a slight loss in the N-terminal short-range contacts, while C-terminal contacts remained constant (**Fig. 3b-d**). In contrast, P301L tauRD showed a more drastic loss in both the long-range and N-terminal short-range contacts (**Fig. 3e-g**), which sits upstream of P301L mutation, suggesting that the mutation destabilizes both global and local structures and makes tauRD more susceptible to heat. Particularly, a stepwise loss of crosslinks at the interface of repeat 2 and 3, encompassing the amyloid motif ^{306}VQIVYK^{311} and where P301L is located, was observed in both WT and P301L in the heat denaturation and was more remarkable in the P301L mutant (**Fig. 3b-g**). This indicates that P301L destabilizes and facilitates unfolding of local structure, which potentially lowers the threshold of tau to enter into an aggregation-prone conformation (**Fig. 4**).

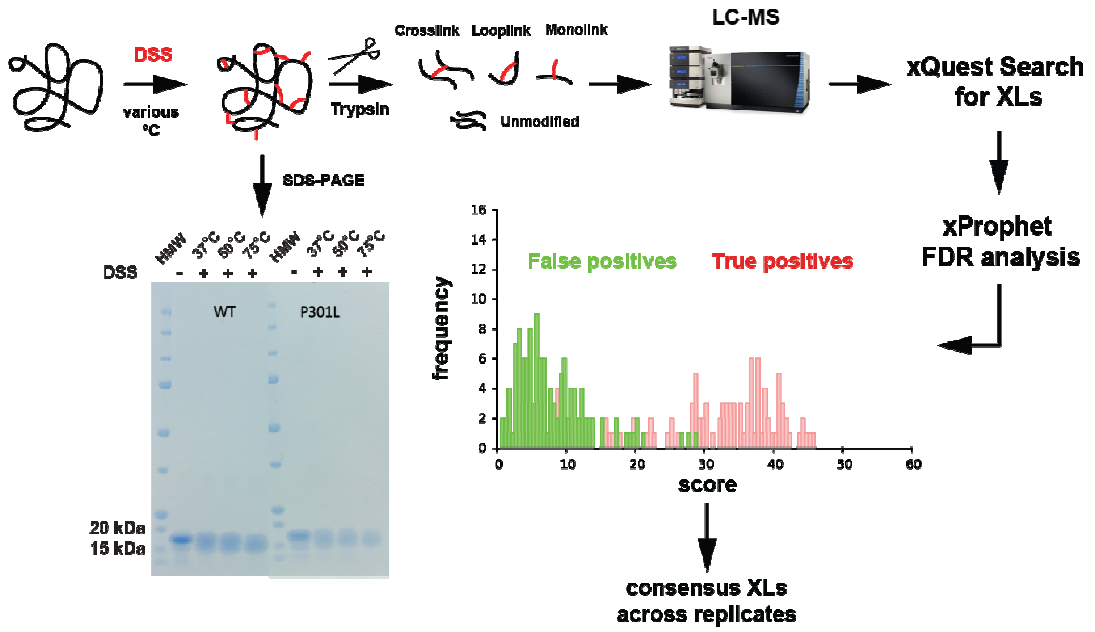


Figure 2. Schematic of XL-MS procedure and data analysis. Protein samples are incubated with crosslinker (i.e. DSS) at the desired temperatures. Non-crosslinked and crosslinked samples are resolved on an SDS-PAGE gel to confirm the reaction has occurred (as indicated by a smearing of the band compared to the control samples; see Note 2) and that the reaction conditions do not yield oligomeric forms. Digest the crosslinked samples with a choice of protease for example trypsin, during which step peptides with crosslink, looplink and monolink modifications as well as non-modified peptides will be formed. Samples are then analyzed by mass spectrometry followed by a xQuest search to identify crosslinked peptides and FDR analysis by xProphet. Only true positive crosslinked contacts with a high xQuest score are used for subsequent analysis and for a more stringent analysis, only consensus crosslinked contacts across replicates are used.

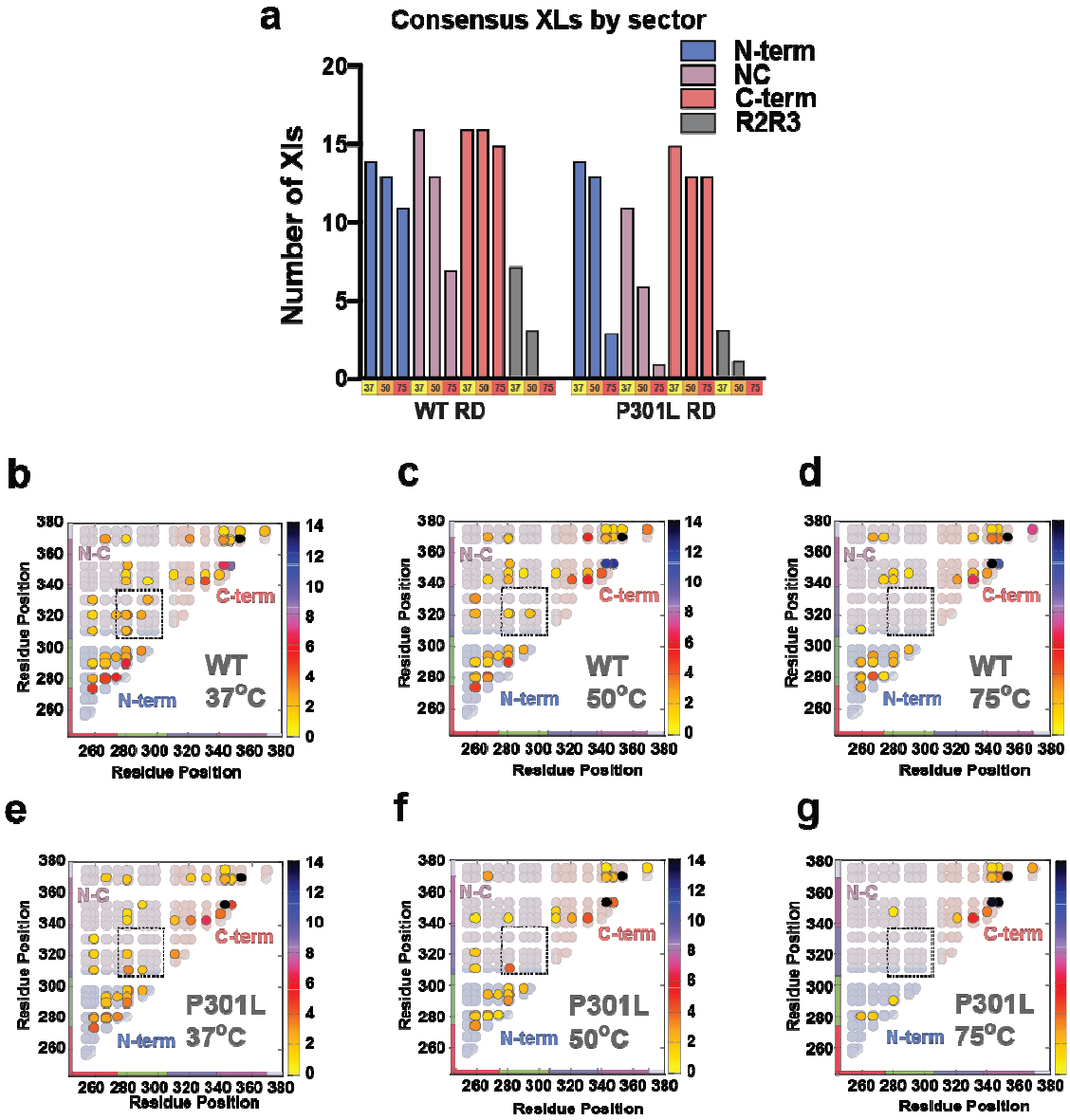


Figure 3. Contacts of tauRD reviewed by XL-MS in heat denaturation. (a) Total consensus cross-links summarized by temperature and location in WT and P301L tauRD: within N-terminus (blue; residues 243–310; N-term), within C-terminus (orange; residues 311–380; C-term), span N- and C-terminus (magenta; 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). **(b-d)** Consensus cross-links (circles) of WT tauRD 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. 1a The dashed boxes define inter-repeat cross-links observed

between repeat 2 and repeat 3. **(e-g)** Same as **(b-d)** above, except with tauRD that contains a P301L mutation.

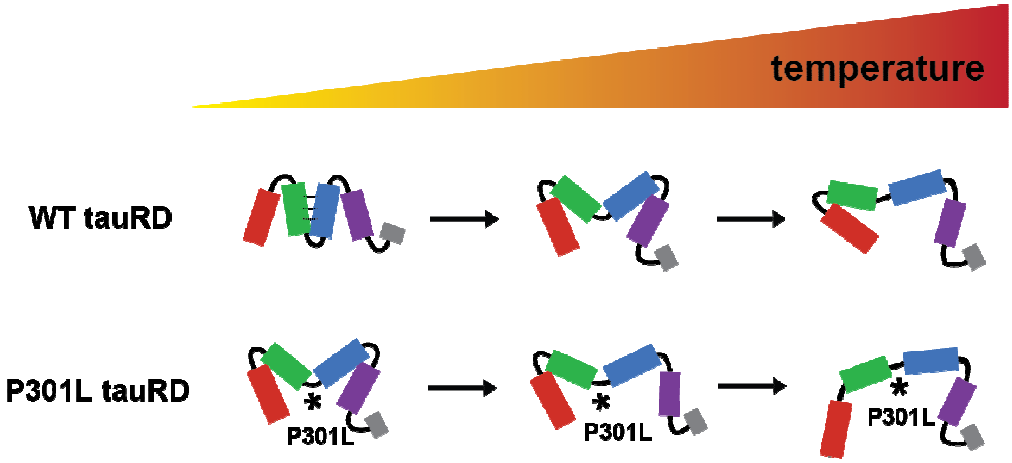


Figure 4. Models of tauRD structural change with mutation and heat. Based on the stepwise contact loss from Fig. 3, as temperature increases, tauRD is predicted to lose R2-R3 contacts most dramatically followed by contacts of R3-R4. P301L mutation (denoted by an asterisk) destabilizes the global and local structures and exacerbates the unfolding as temperature increases. The tauRD cartoon is colored as in Fig. 1.

An array of crosslinkers are available allowing optimization of crosslinking reactions for your protein of interest. Common crosslinkers available include DSS, ADH, and DMTMM, each targeting different amino acids. DSS was used to detect structures in our tauRD samples because tauRD is rich in lysine residues. If your protein possesses relatively more basic residues, we suggest starting off with DSS, while if your protein has more acidic residues, starting with ADH in combination with DMTMM. DSS and ADH insert an alkyl linker between amino acids that are in close contact. DMTMM acts to activate ADH reaction and at the same time also captures salt bridges between basic and acidic residues. DMTMM can be used alone as a crosslinker.

Materials

1. Crosslinker choices
    - Disuccinimidyl suberate (DSS-H12/D12, Creative Molecules), lysine-reactive crosslinker
    - Adipic acid Dihydrazide (ADH-H8/D8, Creative Molecules), crosslinker of acidic residues
    - 4-4-methylmorpholinium chloride (DMTMM, Sigma Aldrich), crosslinker of lysine with acidic residues
2. Dimethyl formamide, anhydrous (DMF)
3. Phosphate-buffered saline (PBS)
4. Dithiothreitol (DTT)
5. Ammonium bicarbonate, BioUltra grade
6. Iodoacetamide, >99%
7. Tris (2-carboxyethyl) phosphine hydrochloride
8. Urea
9. Trypsin, sequencing grade
10. Formic acid, puriss. p.a.
11. Acetonitrile, LC-MS grade
12. Thermomixer for 1.5-ml tubes
13. Lyophilizer
14. SPE cartridges with reversed-phase octadecyl silica (C_{18}) material
15. Vacuum manifold for SPE cartridges

16. Eksigent 1D-NanoLC-Ultra HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system
17. New Objective PicoFrit columns (11 cm × 0.075 mm I.D.) containing Magic C_{18} material (3 µm particle size, 200 Å pore size)
18. msconvert (proteowizard.sourceforge.net)
19. xQuest/xProphet, v. 2.1.1 (http://proteomics.ethz.ch/cgi-bin/xquest2_cgi/index.cgi) installed on a single computer workstation or on a computer cluster

Methods

XL-MS sample preparation

1. Prepare protein sample for crosslinking. Dilute 100 µg of starting material to optimal concentration, 0.5 - 1.0 mg/mL using 1 × PBS with 3 mM DTT (see Note 1). Equilibrate the samples for each temperature condition for 1 hr before cross-linking.
    - For DSS
        - Dissolve 1mg DSS in 108uL DMF to 25mM concentration to make a stock solution. Initiate the chemical crosslinking by adding DSS stock to a final concentration of 1 mM. Incubate reactions at 37 °C, 50 °C, or 75 °C for 1 min with 350 RPM shaking in an Eppendorf thermomixer.
    - For ADH enabled by DMTMM(Leitner, Joachimiak, et al., 2014)
        - Dissolve ADH in H_2O to an 83mg/ml stock concentration, and dissolve DMTMM in H_2O to a 120mg/ml stock concentration. Scale the volume according to the size of your experiment. Initiate the

crosslinking reaction by adding ADH and DMTMM to final concentrations of 8.3mg/ml and 12mg/ml respectively. Incubate reactions at temperature desired for 15min with 750RPM shaking in an Eppendorf thermomixer.

- For DMTMM alone
  - Initiate the crosslinking reaction by adding DMTMM to final concentrations of 12mg/ml. Incubate reactions at temperature desired for 15min with 750RPM shaking in an Eppendorf thermomixer.

2. DSS, ADH and DMTMM reactions can be quenched using 100mM ammonium bicarbonate followed by additional incubation at 37 °C for 30 min in an Eppendorf thermomixer. Take an aliquot out from each sample for SDS-PAGE analysis.
3. Flash freeze in liquid nitrogen and evaporate to dryness in a lyophilizer. Store samples in -80^{o}C until the next step.
4. Resuspend samples in 50ul 8 M urea with 2.5 mM TCEP and incubate for 30min at 37 °C.
5. Alkylate with 5 mM iodoacetamide for 30 min at RT and protected from light (see Note 2).
6. Dilute samples to 1 M urea with 50 mM ammonium bicarbonate and digest with trypsin at an enzyme-to-substrate ratio of 1ug:50ug O/N at 37 °C.
7. Next day stop proteolysis by adding formic acid to 2% (v/v).

8. Purify peptides from trypsin with solid-phase extraction, Sep-Pak tC18 cartridges according to standard protocols. Evaporate samples to dryness and reconstitute in water/acetonitrile/formic acid (95:5:0.1, v/v/v) to a final concentration of ~ 0.5 µg/µL.
9. Quantify concentration of eluted peptides on a small volume spectrophotometer (see Note 4).

Mass spectrometry

10. Inject 2 µL of each sample for duplicate LC-MS/MS analyses on an Eksigent 1D-NanoLC-Ultra HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system.
11. Peptides are separated on self-packed New Objective PicoFrit columns (11 cm × 0.075 mm I.D.) containing Magic C_{18} 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).
12. The mass spectrometer is 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).
13. Fragment ions are detected at low resolution in the linear ion trap. Dynamic exclusion is enabled (repeat count 1, exclusion duration 30 s).

Analysis of mass spectrometry data

14. Convert Thermo.raw files to the open.mzXML format using msconvert (proteowizard.sourceforge.net) and analyze using xQuest(Rinner et al., 2008) modified based on your protein and the crosslinker used.
15. Extract spectral pairs with a precursor mass difference of 12.075321 Da and search against the protein sequence (query) and reversed sequence (decoy) of your protein.
    - Example sequence format:

>tau

MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTED
GSEEPGSETSDAKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTP
SLEDEAAGHVTQARMVSKSKDGTGSDDKKAKGADGKTKIATPRGAAPPGQKGQAN
ATRIPAKTPPAPKTPPSSGEPPKSGDRSGYSSPGSPGTPGSRSRTPSLPTPPTREPKKVA
VVRTPPKSPSSAKSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLS
NVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKS
EKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPV
VSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSASLAKQGL

>tau_decoy

LGQKALSASVEDALTALQPSDVMDISGTSSVNSLHRPSTDGSVVPSKYVIEAGHDTK
AKANERFTLKHTEIKKNGGGPVHTINDLSGIKSQVRDKFDLKESKVEVQGGGPKHHI
NGLSGCKSTVKSLDVPKYVIQVSGGGPVHKINDKSGCKSQVNSLDLKKNIIQVKGGG
PQHKLNETSGIKSKVNKLDPMPVPATQLRSKASSPSKPPTRVVAVKKPERTPPTPLSP
TRSRSGPTGPSGPSSYGSRDGSKPPEGSSPPTKPAPPTKAPIRTANAQGKQGPPAAGRP

TAIKTKGDAGKAKKDDSGTGDKSKSVMRAQTVHGAAEDELSPTDGIGAEEATTGEP
IETHPQAAAQKGPAGEDVLPATVDEATPTSKADSTESGPEESGDETPTQLPSEKLGA
DTDGEQDQHMTYGGQDKRDGLGYTGAHDEMVEFEQRPEAM

16. xQuest settings for DSS are as follows: Maximum number of missed cleavages (excluding the cross-linking site) = 2, peptide length = 4–50 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.
    - In file Xquest.def, adjust “xkinkerID”, “AArequired”, “Xlinkermw” and “monolinkmw” based on your choice of crosslinker.
    - For ADH, “xkinkerID” will be “ADH”, “AArequired” will be “D,E”, “Xlinkermw” will be “138.09055” and “monolinkmw” will be “156.10111”; (see Note 5)
    - For DMTMM, “xkinkerID” will be “DMTMM”, “AArequired” will be “K,D,E”, “Xlinkermw” will be “-18.010595”, and “monolinkmw” will be “0”. (see Note 5)
17. Perform post-search manual validation and filtering 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. Cutoff score is flexible and can be determined so that you get a reasonably low FDR.

18. In addition, at least four assigned fragment ions (or at least three contiguous fragments) are required on each of the two peptides in a cross-link.
19. Use xProphet to determine FDRs for the identified cross-links(Walzthoeni et al., 2012). For example, our FDRs were estimated to be 1.3–10%.
20. For more detailed instruction on XL-MS data analysis, refer to Leitner et al, Nature Protocols, 2013(Leitner, Walzthoeni, & Aebersold, 2014).

Illustration of mass spectrometry results

21. Determine consensus XLs across replicates for each condition. For our case, only cross-links present in five of the five data sets were used to generate a consensus data set.
22. Information such as cross-linked residue positions and frequency (nseen) can be visualized in many ways. Here we provide a customized gnuplot script to visualize the consensus crosslink pairs and their frequency.

Gnuplot script:

```
set pm3d map
set style fill solid border lt -1
set style circle radius 3
plot 'XXXX' using 1:2:3 with circles lc palette
#substitute XXXX with the name of your file containing the sequence positions and the average frequency of each XL pairs
set pointsize 1
```

```
set size square 1,1
set terminal eps
set cbrange [0:15]
set palette defined (0 "yellow", 3 "orange", 6 "red", 9 "purple", 12 "blue", 15 "black")
#set coloring numbers according to the range of your average frequency of the XLs
set xrange [243:380]
set yrange [243:380]
#set xrange and yrange based on your protein sequence
set cblabel "Mean Frequency" font 'Verdana,13'
set output "XXXX.eps"
set xlabel "Residue Position"
set ylabel "Residue Position"
unset key
replot
```

Notes

1. You might not get sufficient number of crosslinks to interpret local structural information if your protein concentration is too low. The optimal protein concentration for a crosslinking reaction is ~0.5-1 mg/ml.
2. You might observe oligomeric bands in crosslinked samples as visualized by SDS-PAGE if incubation time is too long or the crosslinker concentration is too high. If that's the case, we suggest optimizing time and crosslinker concentrations to

determine the conditions where only single band of your protein size is observed, and the band appears smeary indicating a successful intramolecular crosslinking reaction. If conditions cannot be optimized to prevent formation of oligomeric species, the crosslinked monomers can be purified by gel extraction or chromatography.

3. Cover the iodoacetamide reactions in foil to protect from light.
4. Following Solid Phase Extraction (SPE) isolation of peptides, the yields can be determined on a small volume spectrophotometer. To accurately measure peptide concentrations, calculate the E0.1% (1 mg/mL) value at 205 nm: extinction coefficient at 205nm [$M^{-1}cm^{-1}$] divided by molecular weight [g/mol]. Update the value on your small volume spectrophotometer per peptide. Sample calculator: spin.niddk.nih.gov/clore/Software/A205.html.
5. For DMTMM enabled ADH crosslinking reaction, DMTMM is required to activate ADH. However, DMTMM can also act as a crosslinker itself. In this case, you can apply xQuest searches on both ADH and DMTMM.

Acknowledgement

This chapter was replicated in full from my first author manuscript accepted by Springer Nature book *Protein Aggregation*, DOI : 10.1007/978-1-0716-2597-2

CHAPTER THREE

TAU LOCAL STRUCTURE SHIELDS AMYLOID MOTIF AND CONTROLS AGGREGATION PROPENSITY

Introduction

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 is 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 methods (Cleveland, 1977). In contrast, the tau repeat domain (tau RD), which spans residues 243 to 365, is predicted to be more structured (Eliezer, 2005), forms the core of amyloid fibrils (Fitzpatrick, 2017), and is the minimal region to propagate tau prion strains (Sanders, 2014). Tau RD contains an amyloid motif (^{306}VQIVYK^{311}) (**Figure 1 a**) that is central to conversion between the soluble and insoluble states, as it mediates self-assembly, drives amyloid formation *in vitro* (Sawaya, 2007) and promotes pathology *in vivo* (Von Bergen, 2001). Nuclear Magnetic Resonance (NMR) experiments on tau indicate that in solution the ^{306}VQIVYK^{311} motif adopts a β-strand conformation (Eliezer, 2005; Mukrasch, 2005). Recent cryo-EM studies of tau patient-derived fibrils have shown that ^{306}VQIVYK^{311}

mediates important contacts in these structures (Fitzpatrick, 2017; Falcon, 2018). Despite these structural studies, it's 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* (Zhang, 2017; Ismail, 2018; Kuret, 2005). Solution NMR experiments mapped the tau-heparin binding site to repeat 2 just prior to the ^{306}VQIVYK^{311} motif, but how this binding event modulates tau aggregation remains unclear (Zhao, 2017). Double electron-electron resonance (DEER) experiments indicated an expansion of this region upon heparin binding (Zhang, 2017). Cryo-EM structures also suggested an extended conformation of tau when bound to tubulin (Kellogg, 2018). Other work mapping the recruitment of molecular chaperones to tau indicated that many chaperones, including Hsp40, Hsp70 and Hsp90, localize around ^{306}VQIVYK^{311} (Mok, 2018). 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 aggregation (Baughman, 2018). 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 amyloid (Mirbaha, 2018). M_s itself adopts multiple stable structures that encode different tau prion strains (Sharma, 2018), 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 (Elbaum-Garfinkle, S. & Rhoades, E., 2012; Chirita, 2005), but our study was the first to biochemically isolate and characterize this species (Mirbaha,

2018). Different forms of M_s have been purified from recombinant protein, and tauopathy brain lysates (Mirbaha, 2018; Sharma, 2018). Using multiple low-resolution structural methods, we have mapped critical structural changes that differentiate M_i from M_s to near the ^{306}VQIVYK^{311} motif and indicated that the repeat 2 and 3 region in tau is extended in M_s, which exposes the ^{306}VQIVYK^{311} motif (Mirbaha, 2018). In contrast, intramolecular disulfide bridge between two native cysteines that flank ^{306}VQIVYK^{311} in tau RD is predicted to form a local structure that is incompatible with the formation of amyloid (Walker, 2012). Thus, conformational changes surrounding the ^{306}VQIVYK^{311} amyloid motif appear critical to modulate aggregation propensity. A fragment of tau RD in complex with microtubules hinted that ^{306}VQIVYK^{311} forms local contacts with upstream flanking sequence (Kadavath, 2015). This was recently supported by predicted models guided by experimental restraints from crosslinking mass spectrometry (Mirbaha, 2018) and is consistent with independent NMR data (Mukrasch, 2007).

Based on our prior work (Mirbaha, 2018) we hypothesized that tau adopts a β-hairpin that shields the ^{306}VQIVYK^{311} 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 ^{306}VQIVYK^{311} [24] (**Figure 1 a, 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 (Holmes, 2014; Yoshiyama, 2007). 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 propensity (Fischer, 2007). 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 observed (Karamanos, 2015).

Name	**Amino Acid Sequence**
Tau RD AlzForum Mutations[1]	R1: 244 QTAPVPMPDLKN-VkSKiGSTENlKHQPggGK 274 R2: 275 VQIInKKLDlsN-VQSKCGSKDnIKHvpGgGs 305 R3: 306 VQIVYKPVDlsk-vTSKCGSLGNIHHKpGGgq 336 R4: 337 vEVKSEkLDFKDRVQsKIGsLDNiTHvpGgGN 368

Table 1: List of AlzForum Disease-Associated Mutations. [1]sites of mutation are shown in lowercase.

As with disease-associated mutations, alternative splicing also changes the sequence N-terminal to ^{306}VQIVYK^{311}. Tau is expressed in the adult brain primarily as two major splice isoforms: 3-repeat and 4-repeat (Goedert, 1988). The truncated 3-repeat isoform lacks the second of four imperfectly repeated segments in tau RD. Expression of the 4-repeat isoform correlates with the deposition of aggregated tau tangles in many tauopathies (Williams, 2006) and non-coding mutations that increase preferential splicing or expression of the 4-repeat isoform cause dominantly inherited tauopathies (Williams, 2006; Schoch, 2016; Hutton, 1998). 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 RD (Stöhr, 2017). These have described a relationship between the length of a repeat fragment, its propensity to spontaneously aggregate, and its seeding capacity in cells (Stöhr, 2017). However, inter-repeat interactions may also influence aggregation given that both alternative splicing and many disease-associated mutations cluster around the repeat interfaces (**Figure 1 a**). Our prior work suggested that wild-type tau aggregates less efficiently because the flanking sequence shields ^{306}VQIVYK^{311} (Mirbaha, 2018). We hypothesize that the intrinsically disordered tau protein evolved to minimize aggregation by adopting local structure that shields the ^{306}VQIVYK^{311} 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 ^{306}VQIVYK^{311} within tau. Our data supports a model where disease-associated mutations, alternative splicing, or other factors can destabilize this local structure and expose ^{306}VQIVYK^{311} leading to enhanced self-assembly.

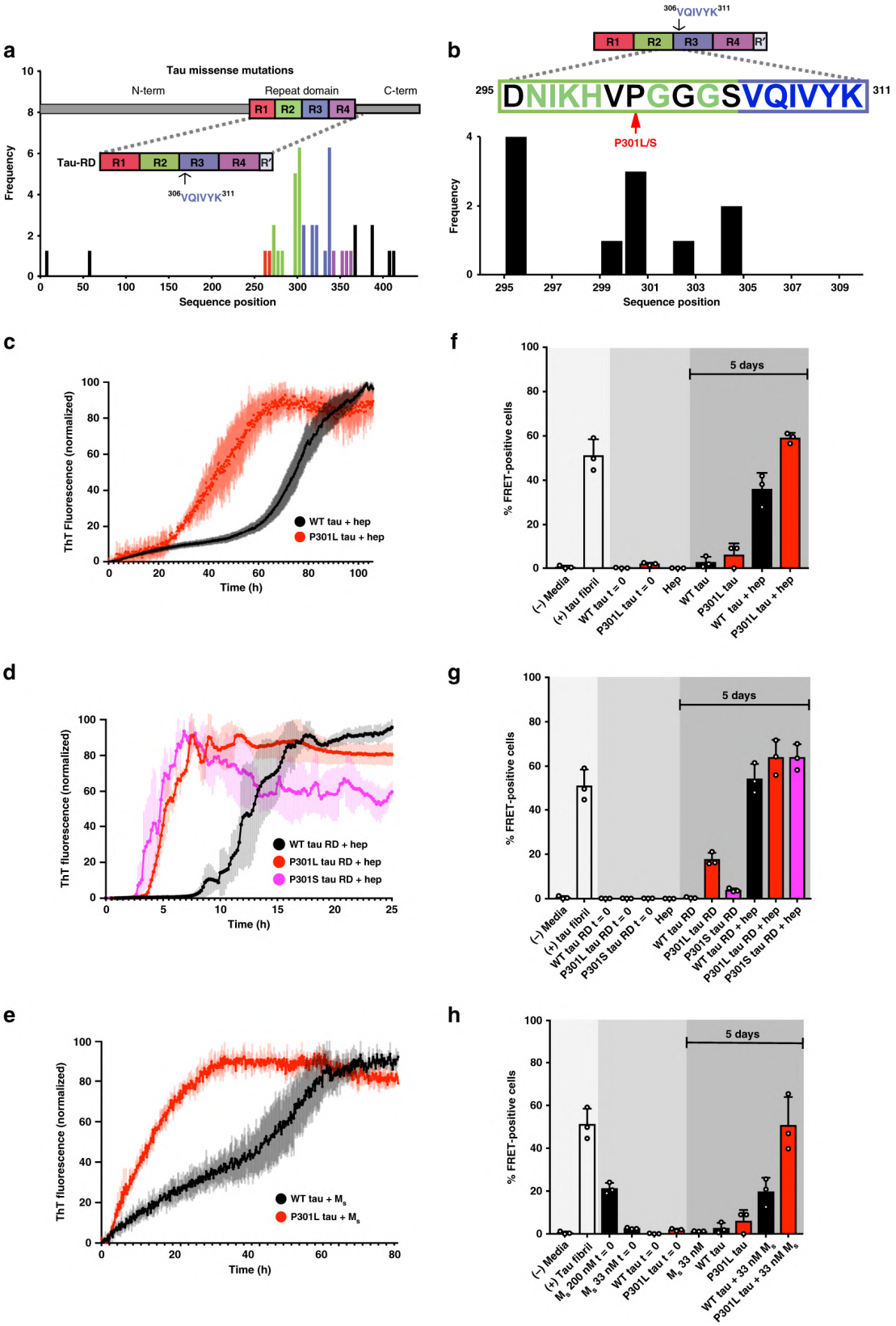

a
Tau missense mutations
N-term
Repeat domain
C-term
R1
R2
R3
R4
Tau-RD
R'
306VQIVYK311
Frequency
Sequence position
b
306VQIVYK311
295
DNIKHVPGGGSVQIVYK
311
P301L/S
Frequency
Sequence position
c
ThT Fluorescence (normalized)
Time (h)
WT tau + hep
P301L tau + hep
d
ThT fluorescence (normalized)
Time (h)
WT tau RD + hep
P301L tau RD + hep
P301S tau RD + hep
e
ThT fluorescence (normalized)
Time (h)
WT tau + Ms
P301L tau + Ms
f
% FRET-positive cells
5 days
(–) Media
(+) tau fibril
WT tau t = 0
P301L tau t = 0
Hep
WT tau
P301L tau
WT tau + hep
P301L tau + hep
g
% FRET-positive cells
5 days
(–) Media
(+) tau fibril
WT tau RD t = 0
P301L tau RD t = 0
P301S tau RD t = 0
Hep
WT tau RD
P301L tau RD
P301S tau RD
WT tau RD + hep
P301L tau RD + hep
P301S tau RD + hep
h
% FRET-positive cells
5 days
(–) Media
(+) Tau fibril
Ms 200 nM t = 0
Ms 33 nM t = 0
WT tau t = 0
P301L tau t = 0
Ms 33 nM
WT tau
P301L tau
WT tau + 33 nM Ms
P301L tau + 33 nM Ms

Figure 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}VQIVYK^{311} is shown in the inset cartoon. b. Detailed mutation frequencies found near the ^{306}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 2-fold ratio to background signal) over the course of the experiment. 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 2-fold ratio to background signal) over the course of the experiment. e. WT FL tau and mutant P301L tau at a 4.4 μM concentration were mixed with substoichiometric M_s tau seed (33nM) 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 2-fold ratio to background signal) over the course of the experiment. 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 hours 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 3 biological triplicates of at least 10,000 cells per condition. Error bars represent a 95% CI of each condition.

Results

P301 mutations promote aggregation in vitro and in cells

Missense mutations that change proline 301 to leucine or serine cause dominantly inherited tauopathy (Rizzu, 1999) and are associated with neurodegeneration in model systems (Yoshiyama, 2007; Mocanu, 2009), 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) (**Table 2**). 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 ± 0.5 hrs) aggregated more rapidly compared to WT tau ($t_{1/2}$ = 75 ± 0.3 hrs) (**Figure 1 c**). 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 ± 0.1 hrs; P301S tau RD, $t_{1/2}$ = 3.9 ± 0.1 hrs) than WT tau RD (WT tau RD, $t_{1/2}$ = 12.5 ± 0.2 hrs) (**Figure 1 d**). Consistently, we found that mutations at position 301 (from proline to either leucine or serine) increased aggregation rates by ~2-fold compared to WT in both FL tau and tau RD constructs. Thus, *in vitro,* tau RD recapitulates key aspects of aggregation observed in FL tau.

The inert conformation of monomeric tau (M_i) requires cofactors, such as heparin, to spontaneously aggregate *in vitro*, while the seed-competent monomer (M_s), derived from recombinant protein or Alzheimer's patient brain material, readily self-assembles to form amyloid (Mirbaha, 2018). 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 reactions (Mirbaha, 2018). In this study, we evaluated the aggregation propensity of the P301L mutant compared to 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 ± 0.6 hrs) than the WT protein (WT tau, $t_{1/2}$ = 40 ± 1.1 hrs) (**Figure 1 e).** P301L tau aggregated faster than WT tau with a 4-fold 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 to 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 (Mirbaha, 2018). Mutations at the P301 may exacerbate aggregation by unfolding the region surrounding the amyloid motif ^{306}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 proteins (Holmes, 2014). These cells sensitively report a 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-synuclein (Furman, 2015). 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 to WT fibrils (**Figure 1 f**). 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 (**Figure 1 g**). Lastly, the seeding activity for the M_s-induced FL tau fibrils showed a 2-fold higher activity for P301L compared to WT (**Figure 1 h**). WT FL tau and tau RD control samples (no heparin or M_s) did not produce seeding activity in cells while P301 mutants, both FL and tau

RD, showed hints of seeding activity despite not yielding positive ThT signal *in vitro*, perhaps due to the formation of oligomers not captured by ThT. As expected, 33nM M_s control exhibited seeding activity at the onset and did not change after 5 days, but overall signal was low due to the low concentrations used in the aggregation experiments. Interestingly, WT tau induced with 33nM M_s seeded at similar levels to concentrated control (200nM) M_s samples highlighting efficient conversion of WT tau into seed-competent forms (**Figure 1 h**). Thus, P301 mutations promote aggregation *in vitro* and in cells across different constructs. Importantly, these effects are conserved between FL tau and tau RD.

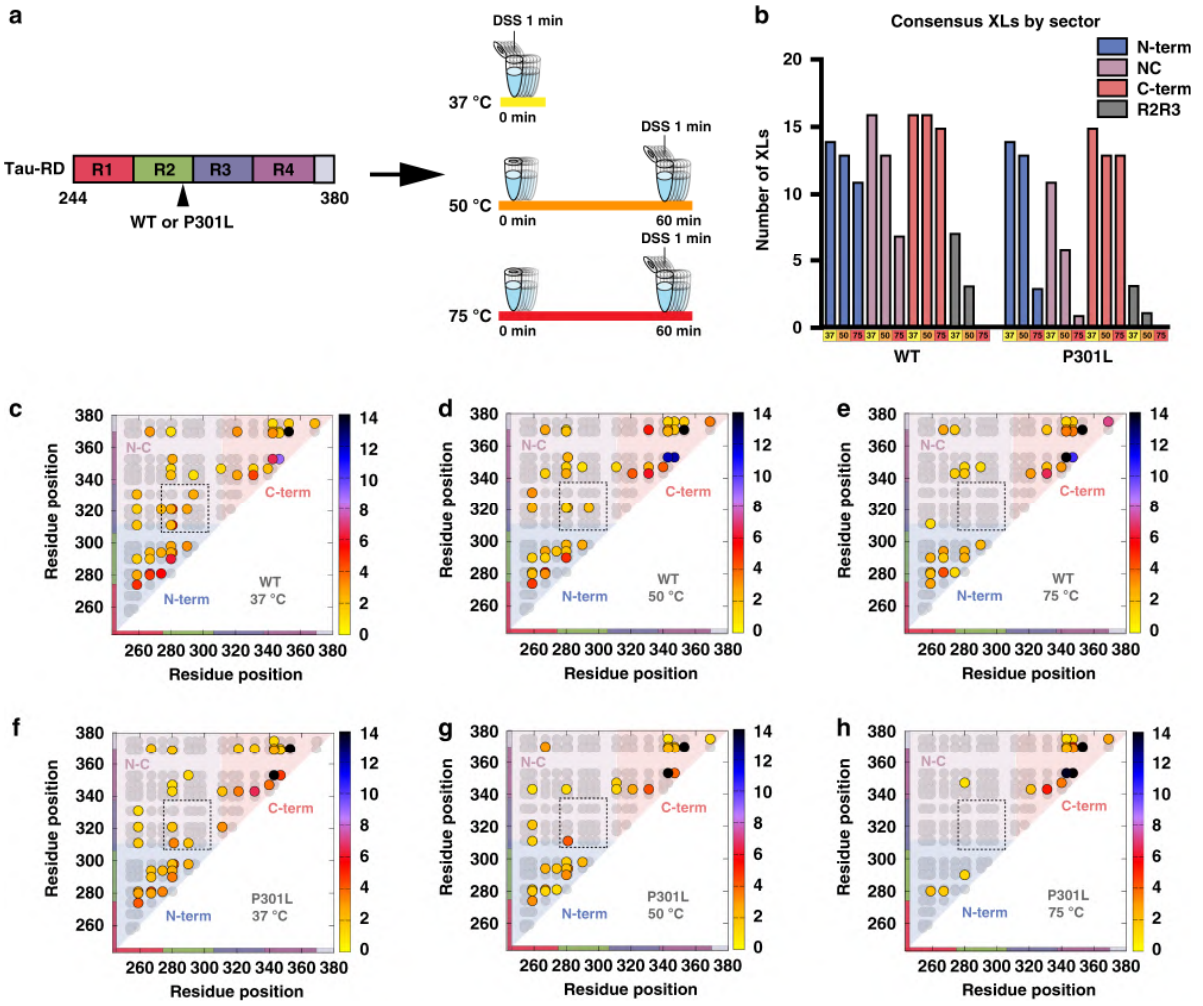


Figure 2. Tau RD encodes global and local structure. a. Cartoon schematic of tau RD used for XL-MS studies colored according to repeat domain. Recombinant WT and P301L tau RD were heated at 37ºC, 50ºC or 75ºC for one hour, then chemically crosslinked using DSS. After cross-linking, trypsin fragmentation, and LC-MS/MS analysis were performed. Each sample was carried out in 5 technical replicates. **b**. Total consensus crosslinks 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) (grey; between residues 275-305 and 306-336). **c-e**. Consensus crosslinks (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. Crosslink 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 Figure 1. The dashed boxes define inter-repeat crosslinks observed between repeat 2 and repeat 3. **f-h**. Same as f-h above, except with tau RD that contains a P301L mutation.

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 (Mirbaha, 2018; Leitner, 2012; Joachimiak, 2014). To preclude the formation of intermolecular crosslinks between monomers, low concentration samples of WT, P301L and P301S tau RD (**Table 2**) were incubated at different temperatures, reacted with disuccinimidyl suberate (DSS; a primary amine crosslinker) for 1 minute and quenched (**Figure 2 a**). The crosslinked protein monomers were confirmed by SDS-PAGE (**Supplementary Figure 2 a**). Cross-linked samples were trypsin digested, analyzed by mass spectrometry and the spectra were searched using the Xquest (Rinner, 2008) to identify intramolecular protein contact pairs (**Methods**). In each dataset, the cross-links reported represent consensus data across five independent samples with a low false discovery rate (**Methods, Supplementary Figure 3**).

XL-MS of recombinant WT tau RD acquired at 37ºC 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) (**Figure 2 b, c**). Importantly, the experimentally observed crosslinks represent only a small subset of all theoretical lysine pairs suggesting that tau RD samples have discrete modes of contacts (**Figure 2 c, grey circles**). Heating the samples to 50ºC or even 75ºC decreased the number of N-C long-range and N-term short-range contacts identified (**Figure 2 b, d, e and Supplementary Figure 2 b).** The data acquired for WT tau RD in physiological conditions is 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 37ºC, the crosslinks found in P301L tau RD (**Figure 2 f**) were similar in pattern to WT, except for fewer N-C terminal long-range contacts **(Figure 2 b and Supplementary Figure 2 b, c)**. However, samples incubated at 50ºC and 75ºC revealed a significant reduction in both long-range and short-range contacts in P301L tau RD compared to WT **(Figure 2 b)**. The loss of short-range contacts, both the total number of crosslinks and the abundance of each crosslink, were detected particularly within the N-terminal sector, which sits upstream of P301 (**Figure 2 b, f-h and Supplementary Figure 2 b, 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 crosslinks 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}VQIVYK^{311} (**Figure 2 b, grey bars; Figure 2 c-h, inset box**). Thus, while the number of crosslinks identified was comparable between WT and P301L tau RD at 37ºC, P301L tau RD retained approximately half as many crosslinks as WT when heated. Similar crosslinking profiles were observed for the P301S tau RD sample (**Supplementary Figure 2 d**). Thus, the lack of thermostability in P301 mutated tau RD as compared to 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 5,000 models using two parallel strategies in ROSETTA: *ab initio* which employed fragment libraries derived from experimental structures (Ovchinnikov, 2018) and CS-ROSETTA which leveraged available chemical shifts for tau RD to produce fragment libraries (Lange, 2012). Both approaches led to a diversity of models consistent with experimentally determined radii of gyration (Mylonas, 2008) **(Supplementary Figure 4 a-d)**. 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}PGGG^{274}, ^{301}PGGG^{304}, ^{333}PGGG^{336} and ^{366}PGGG^{369} sequences (**Supplementary Figure 4 e, Supplementary Figure 5 a-d and 6**). Previously published solution NMR data have shown that the PGGG sequences in tau can adopt type II

β-turns (Mukrasch, 2005), and the ^{301}PGGG^{304} sequence preceding ^{306}VQIVYK^{311} is compatible with the formation of a β-hairpin. We illustrated the R2R3 ^{306}VQIVYK^{311}-containing fragment derived from low energy expanded models produced by each method (**Supplementary Figure 4 c, d**). The ^{306}VQIVYK^{311}-containing interface has the highest frequency of disease-associated mutations, particularly P301L and P301S (**Figure 1 a**). Other potential amyloid-forming regions, such as ^{275}VQIINK^{280}, which is capable of aggregation (**Supplementary Figure 6**), is also preceded by ^{271}PGGG^{274} and predicted to form a β-hairpin (**Supplementary Figure 4 e and Supplementary Figure 5**), however, it is absent in recent cryo-EM structures of tau aggregates (Fitzpatrick, 2017; Falcon. 2018). Mapping known missense mutations onto the *ab initio* β-hairpin structure at R2R3 interface (**Supplementary Figure 4 f**), we hypothesized that this cluster of disease-associated mutations could destabilize the β-hairpin secondary structure, thus exposing the amyloid motif ^{306}VQIVYK^{311} and enabling aggregation. This model is compatible with recent cryo-EM findings that indicate a disengagement of the ^{306}VQIVYK^{311} N-terminal flanking sequence in a fibril structure (Fitzpatrick, 2017). Thus, we focused our studies on the R2R3 motif of tau that contains ^{306}VQIVYK^{311}.

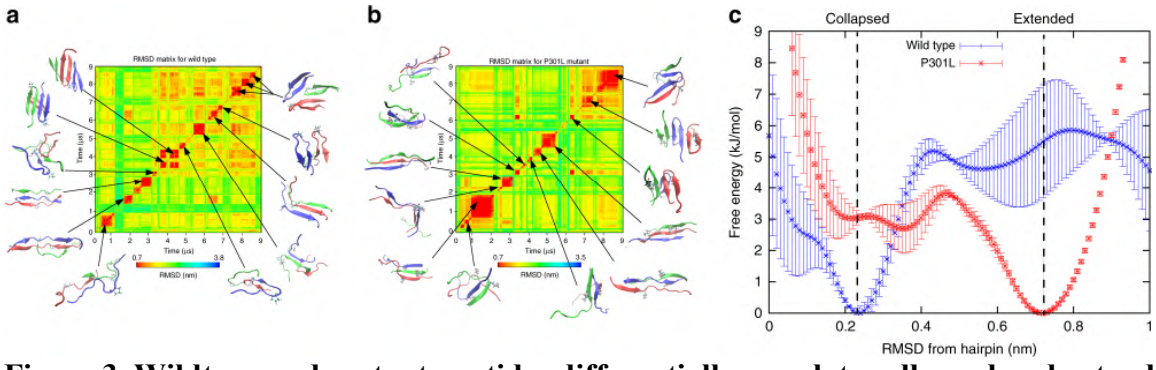


Figure 3. Wildtype and mutant peptides differentially populate collapsed and extended conformations. a. Conformations obtained for a trimer of WT peptide fragment (R2R3-WT)

with the sequence ^{295}DNIKHVPGGGSVQIVYK^{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 above, 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.

P301L promotes extended forms of tau

In silico modelling corroborated recent biochemical findings (Mirbaha, 2018) and suggested a minimal sequence necessary to form a collapsed structure around ^{306}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}DNIKHVPGGGSVQIVYK^{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 approach (Lin, 2017). The trimer conformations obtained in simulations are depicted on a root-mean-square deviation (RMSD) matrix for both R2R3-WT (**Figure 3 a**) and the R2R3-P301L mutant peptide fragments (**Figure 3 b**). For R2R3-WT peptide fragment, we observed a dominant population of trimeric conformations composed of hairpins, while the P301L disease-associated mutation stabilized an extended fibrillar form. The energy basin for R2R3-WT peptide fragment was predicted to be 5 to 6 kJ/mol lower in a collapsed state than an extended state, whereas the R2R3-P301L peptide fragment was 3 kJ/mol lower in an extended state than a collapsed state (**Figure 3 c**). Additionally, the free energy surface suggested an energy barrier of approximately 5 kJ/mol to convert R2R3-WT

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

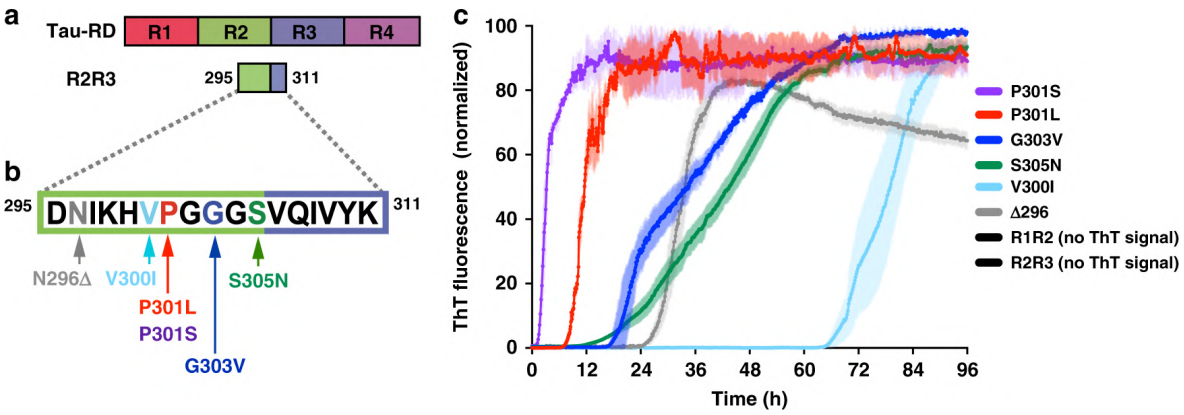


Figure 4. Tauopathy mutations drive aggregation propensity. **a**. Schematic of tau RD and the derived peptides representing the minimal structural element around ^{306}VQIVYK^{311}. **b**. WT and mutant peptides were disaggregated, resuspended to 200 µM, and allowed to aggregate in the presence of ThT at room temperature. WT R2R3 and R1R2 fragment peptides yielded no detectible ThT signal change (less than 2-fold ratio to background signal) over the course of the experiment. 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.

Tau amyloid formation is governed by flanking residues

In tau RD, ^{306}VQIVYK^{311} is necessary for amyloid formation (Sawaya, 2007; von Bergen, 2001). In solution, ^{306}VQIVYK^{311} hexapeptide aggregates spontaneously and rapidly as measured by ThT fluorescence intensity whereas the upstream N-terminal sequence ^{295}DNIKHV^{300} does not aggregate (**Supplementary Figure 6**). To experimentally test the prediction of a local hairpin structure encompassing ^{306}VQIVYK^{311}, we employed a mutagenesis study on synthetic peptide systems that recapitulate the minimal hairpin sequence (**Table 3**). Consistent with the prediction from MD simulation, R2R3-WT peptide

fragment did not aggregate readily, with no ThT detected within 96 hrs (**Figure 4 a, c**). By contrast, single disease-associated mutations (**Figure 4 b**) substituted into the R2R3 peptide fragment were sufficient to promote spontaneous amyloid formation: R2R3-P301S ($t_{1/2}$ = 4.1 ± 1.3 hours), R2R3-P301L ($t_{1/2}$ = 7.2 ± 0.2 hours), R2R3-N296Δ ($t_{1/2}$ = 31.9 ± 0.2 hours), R2R3-G303V ($t_{1/2}$ = 32.1 ± 0.7 hours), R2R3-S305N ($t_{1/2}$ = 41.2 ± 0.2 hours), and R2R3-V300I ($t_{1/2}$ = 77.8 ± 1.3 hours**) (Figure 4 c**). 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 (**Figure 5 b-h**).

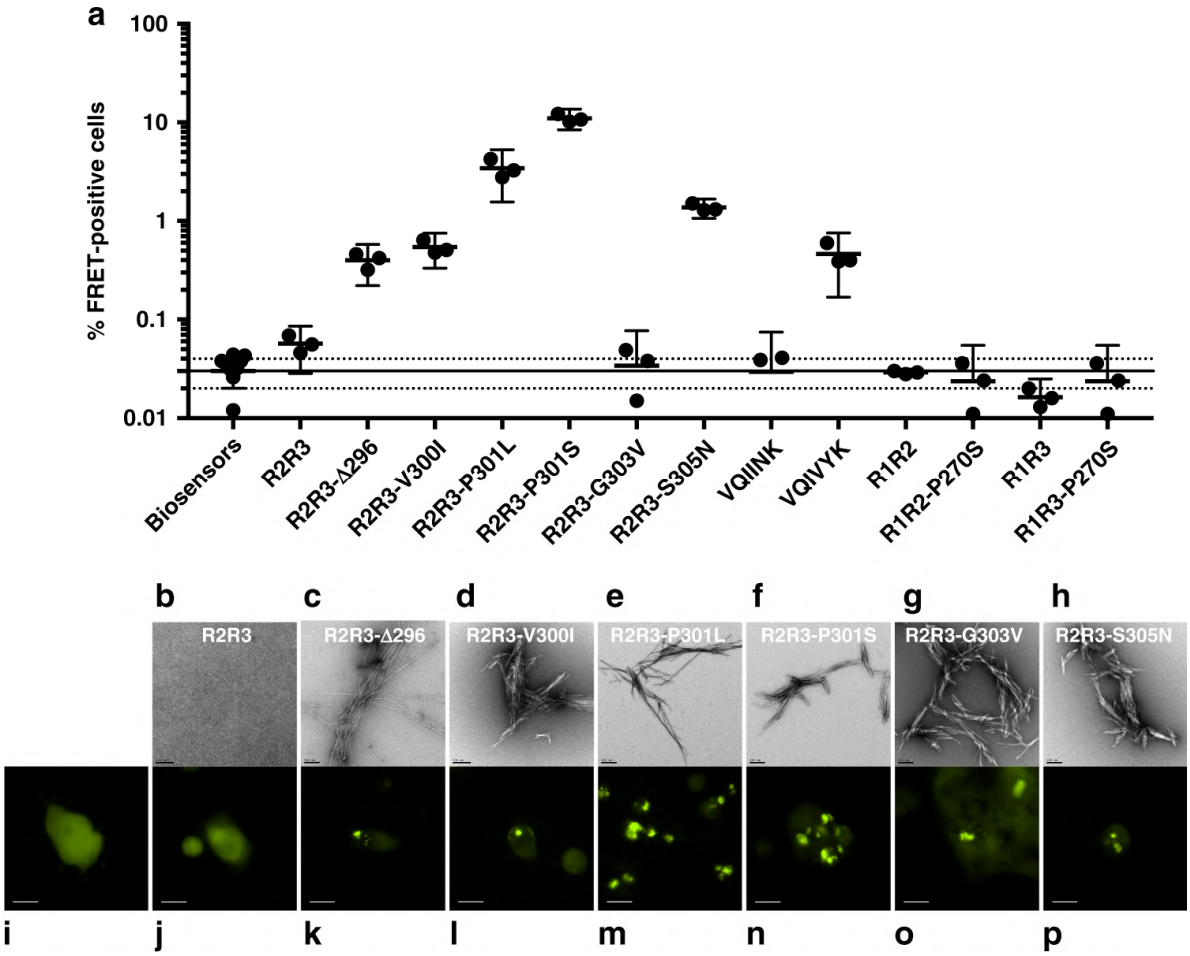


Figure 5. Peptides form amyloid structures and seed *in vivo*. a. After 96 hours of *in vitro* incubation, peptides from previous ThT experiments (**Figure 4 c**) were transduced into tau biosensor cells *via* lipofectamine (**Methods**). FRET signal from each condition (tau RD-CFP/tau RD-YFP) was measured by flow cytometry on 3 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 (**Figure 4 c**). 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 μm distance in each image.

To test the structural compatibility of peptide aggregates formed by *in vitro* tau models, we again employed tau biosensor cells (Holmes, 2014). 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*) (**Figure 5 a**). 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 (**Figure 5 i-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 (Sanders, 2014). Thus, 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.

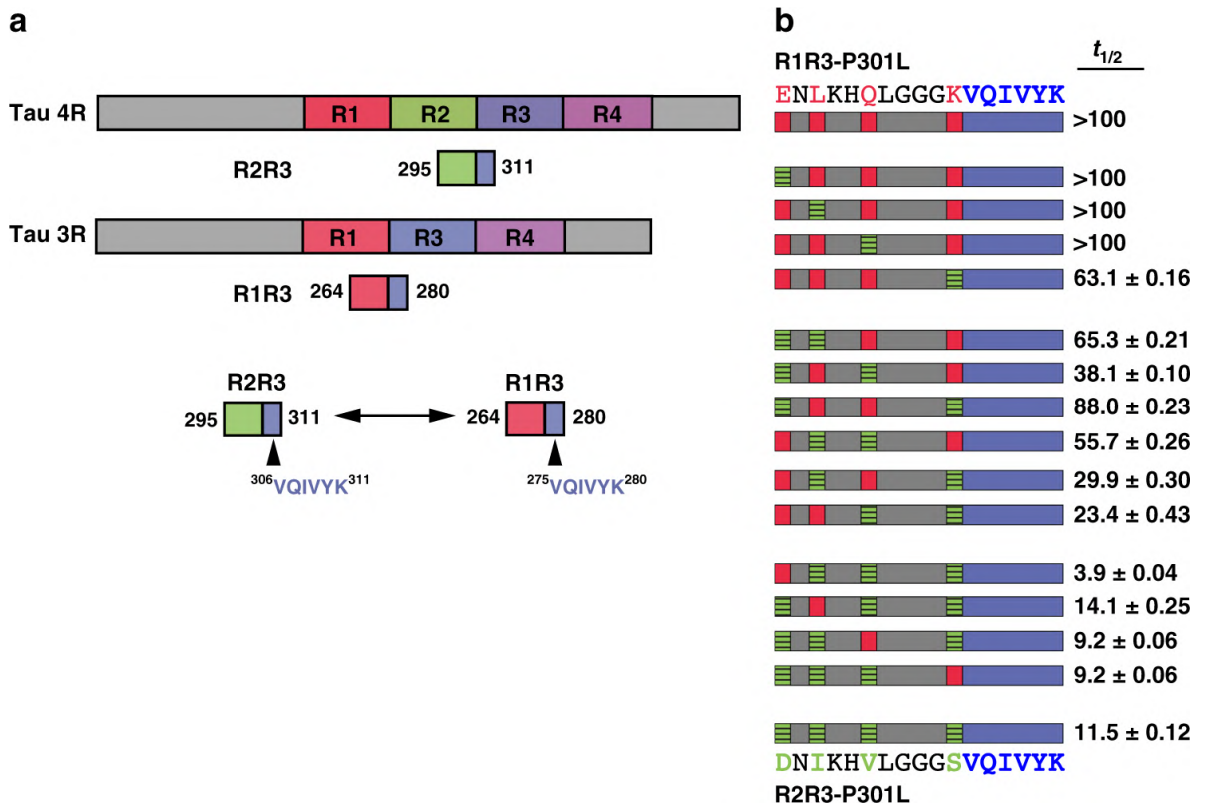


Figure 6. Alternative splicing modulates aggregation propensity. a. Cartoon schematic for tau 4R and 3R splice isoforms illustrate the difference in primary amino acid sequence leading into the amyloidogenic ^{306}VQIVYK^{311} motif. **b**. A full combinatorial panel of R2R3-P301L and R1R3-P301L chimeras were aggregated *in vitro*. ^{306}VQIVYK^{311} 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.

Tau splice variants reveal different aggregation propensity

Tau is expressed in the adult brain as 6 major splice isoform types that include either 3 or 4 repeated segments within RD (**Figure 6 a**). 3R tau lacks the second of four imperfect repeats. 4R tau correlates strongly with aggregation in most tauopathies (Williams, 2006) and mutations that increase splicing of the 4R isoform cause dominantly inherited tauopathies (Williams, 2006; Schoch, 2016; Hutton, 1998). We examined whether this splice isoform

affects the propensity of ^{306}VQIVYK^{311}-mediated aggregation due to the different composition of upstream flanking sequence. We constructed a series of peptide fragments to encompass the R1R3 interface (**Table 3**). This wild type peptide fragment R1R3 mimicking a 3R splice isoform did not spontaneously aggregate. Surprisingly, an R1R3 peptide fragment with a corresponding P301L mutation (R1R3-P270L) also did not aggregate (**Figure 6**). We hypothesized that R1 leading sequence stabilizes the amyloid motif ^{306}VQIVYK^{311} resulting in the aggregation resistance in the presence of disease-associated mutations.

The R1 leading sequence ^{264}ENLKHQPGGGK^{273} differs from R2 ^{295}DNIKHVPGGGS^{304} 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 (**Figure 6 b**) and measured their aggregation kinetics (**Supplementary Figure 7)**. We identified a general trend where the R2R3-P301L peptide fragment aggregates in hours with zero or one R1 substitutions. With two R1 substitutions, R2R3-P301L peptide aggregation was delayed roughly an order of magnitude to tens of hours. With three R1 substitutions, 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 (**Figure 6 b and Supplementary Figure 7**). Thus, all four amino acids contributed to the ability of the R1 leading sequence to delay ^{306}VQIVYK^{311}-mediated spontaneous aggregation in a 3R splice isoform. This may explain the differential aggregation propensities of tau isoforms in human pathology.

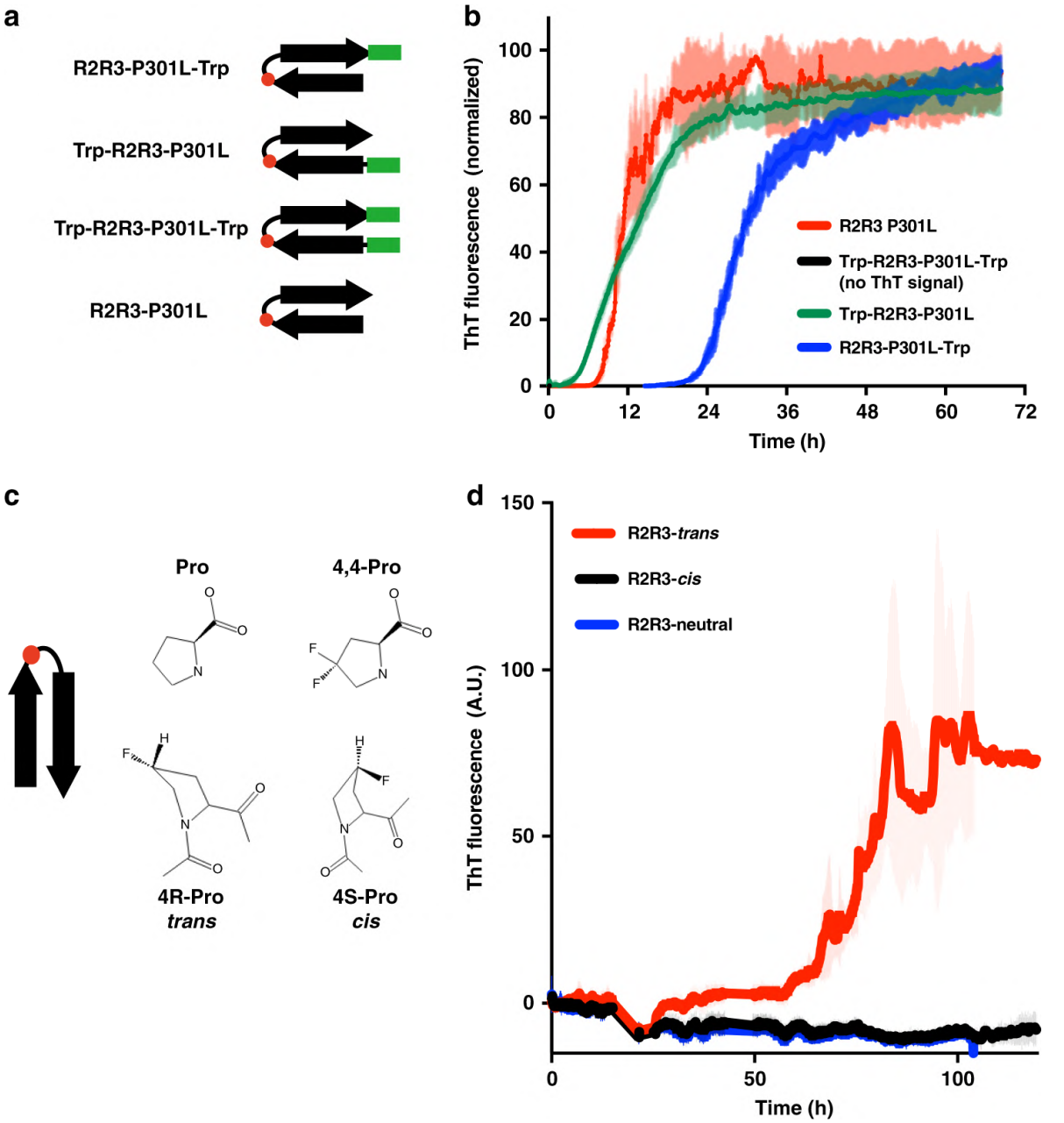


Figure 7. Enhancing β-hairpin structure rescues spontaneous aggregation phenotypes. a. Cartoon schematic representation of the tryptophan zipper motif (green bar) and controls used to stabilize a β-hairpin structure in an R2R3-P301L peptide fragment (**Table 1**). **b**. Aggregation reactions of the tryptophan zipper peptide and controls measured by ThT fluorescence. The Trp-R2R3-P301L-Trp fragment peptide yielded no detectible ThT signal change (less than 2-fold ratio to background signal) over the course of the experiment 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 6 independent experiments with standard deviation shown.

Stabilizing β-hairpin blocks P301L-mediated aggregation

Our model predicted shielding of the ^{306}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, **Table 3 and Figure 7 a**), which stabilizes a β-hairpin structure approximately -2.5 to -7 kJ/mol[46]. Consistent with our model, Trp-R2R3-P301L-Trp peptide fragment does not spontaneously aggregate *in vitro* (**Figure 7 b**).

To ensure that this effect wasn't 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 (**Figure 7 a**). Both half-sequence controls spontaneously aggregated, implying that a tryptophan at either position is insufficient to block aggregation (**Figure 7 b**). 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, **Table 1**) retarded R2R3-P301L peptide fragment aggregation over an order of magnitude ($t_{1/2}$ = 7 hours to $t_{1/2}$ > 70 hours, **Supplementary Figure 8**).

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 ± 1% of tau biosensor cells, but only 0.36 ± 0.12% of the tryptophan zipper stabilized biosensor cells (**Supplementary Figure 9**).

Proline 301 cis-trans isomerization modulates aggregation

Many proteins in the cell utilize proline isomerization as a molecular switch, such as heat shock protein activation (Vogel, 2006) or cell cycle regulation (Gustafson, 2017). In some proteins, proline isomerization directly induces or mitigates aggregation into amyloid (Pastorino, 2006; Torbeev, 2013; Lim, 2008). Proline isomerization events in tau have been proposed to play a role in aggregation and disease (Pastorino, 2006)[49], but P301 isomerization has not been linked to tau aggregation and pathology. With the fact that serine or leucine substitutions at P301 proximal to ^{306}VQIVYK^{311} 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 (**Table 3, Figure 7 c**). Only R2R3-Trans peptide spontaneously aggregated (**Figure 7 d**), 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-repeat 3 interface), which encompasses the amyloidogenic ^{306}VQIVYK^{311} 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 form (Mirbaha, 2018). Thus, the proposed model rationalizes the fundamental molecular mechanisms of aggregation for P301L and at least 5 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 (Ismail, 2018; Kuret, 2005; Zhang, 2017). 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 (Zhao, 2017; Mirbaha, 2018; Zhang, 2017). This process, however,

requires stoichiometric amounts of polyanion and is not a physiological condition since 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 pathology (Mirbaha, 2018). 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 ^{306}VQIVYK^{311} motif is preferentially exposed in M_s or P301L mutant in contrast to normal tau where it is relatively shielded. Thus, the dramatic sensitivity of P301L to seeds can be explained by an increased exposure of the aggregation-prone ^{306}VQIVYK^{311} 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 (Eliezer, 2005; Mukrasch, 2005; Mukrasch, 2007; Mukrasch, 2009). 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 (Meng, 2013; Ding, 2005). 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 ^{306}VQIVYK^{311} regulates aggregation of tau and that the P301L mutation increases susceptibility to conformational changes that expose the ^{306}VQIVYK^{311} amyloid motif. While these differences are subtle, we observe that P301L-mediated structural rearrangements only manifest under moderate stress conditions (i.e. heat, seed). Hence, as compared to 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 ^{306}VQIVYK^{311}. 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-turns (Mukrasch, 2005) and that the P301L mutation increases local β-strand propensity (Fischer, 2007). 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 ^{306}VQIVYK^{311} did not aggregate spontaneously; however, single point substitutions of 6 disease-associated mutations immediately N-terminal to ^{306}VQIVYK^{311} consistently induced spontaneous aggregation. Given that destabilization of local structure around ^{306}VQIVYK^{311} 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) (Mukrash, 2005; Mukrasch, 2007), 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 disproportionally 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 (Holmes, 2014; Yoshiyama, 2007). 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 RD (Stöhr, 2017). 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 seed (Stöhr, 2017). 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 strains (Sanders, 2014). 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 brain (Goedert, 1998). However, the 3R:4R ratio of aggregate deposits is disproportionally shifted towards 4R in most tauopathies (Williams, 2006). Mutations in tau that affect alternative splicing and generate excess 4R isoforms correlate with some genetic tauopathies (Schoch, 2016; Hutton, 1998). The N-

terminal flanking sequences leading into ^{306}VQIVYK^{311} 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 hours vs $t_{1/2}$ > 100 hours, 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 ^{306}VQIVYK^{311} 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 ^{300}HXPGGG^{304} sequence suggests it's possible to utilize inter-repeat regions to select between pathogenic and non-pathogenic conformations of tau (Kontsekova, 2014).

Studying the missense mutations in tau has generated valuable disease models (Yoshiyama, 2007; Mocanu, 2008); however, the majority of human tauopathies have no observed genetic mutation in tau (Rizzu, 1999). 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 amyloidosis (Eichner, 2009). Other proline residues outside of the tau repeat domain have also been proposed to undergo proline isomerization (Pastorino, 2006). 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 (**Figure 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 nearly 75% of aggregation nucleating regions in the human proteome use two or more "gatekeeper" residues, with proline being the most common single gatekeeping residue (Reumers, 2009). 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 ^{306}VQIVYK^{311} 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.

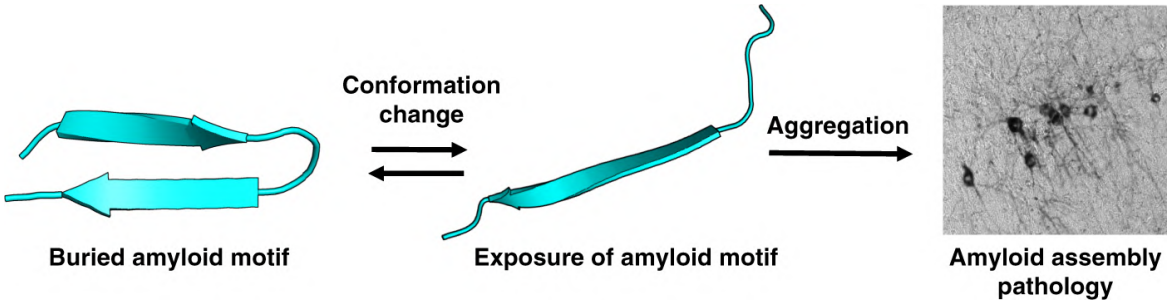
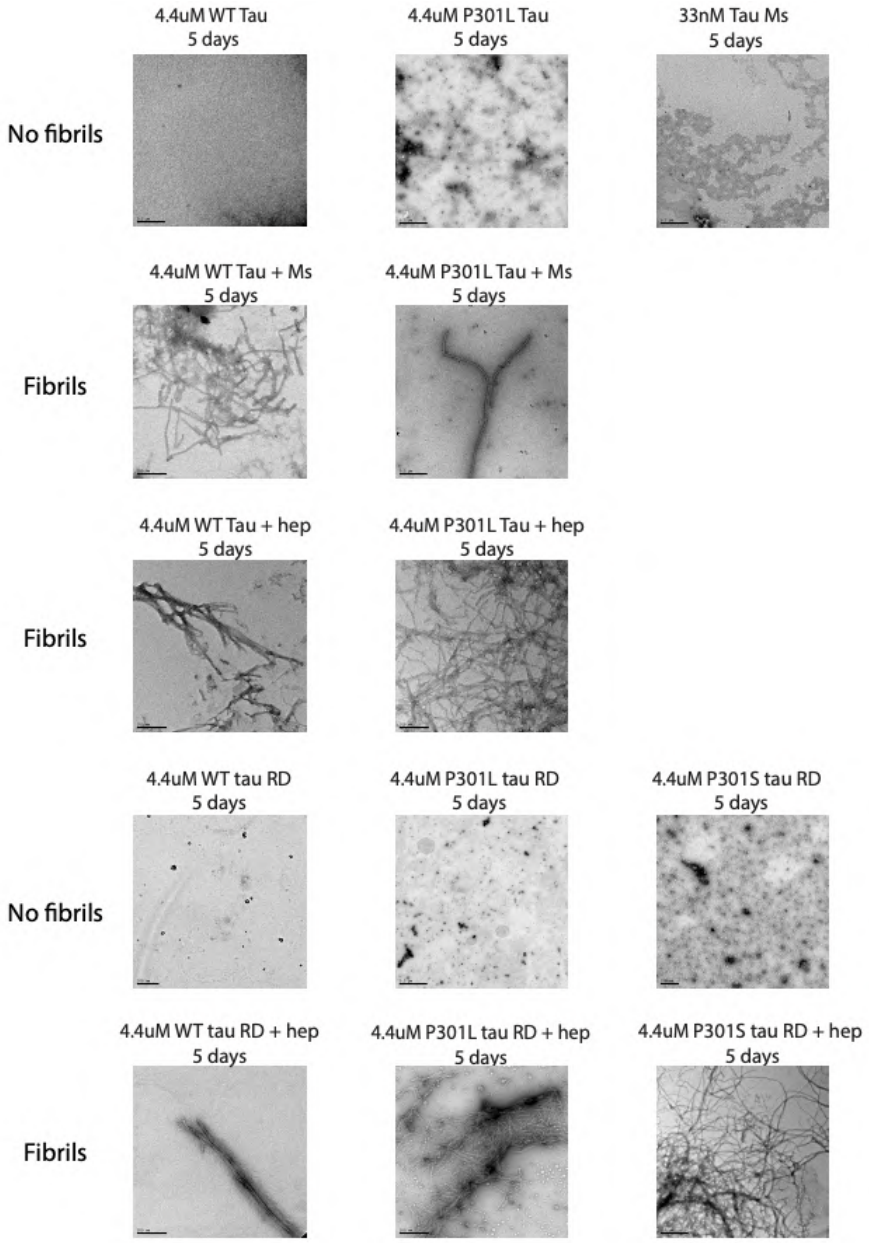
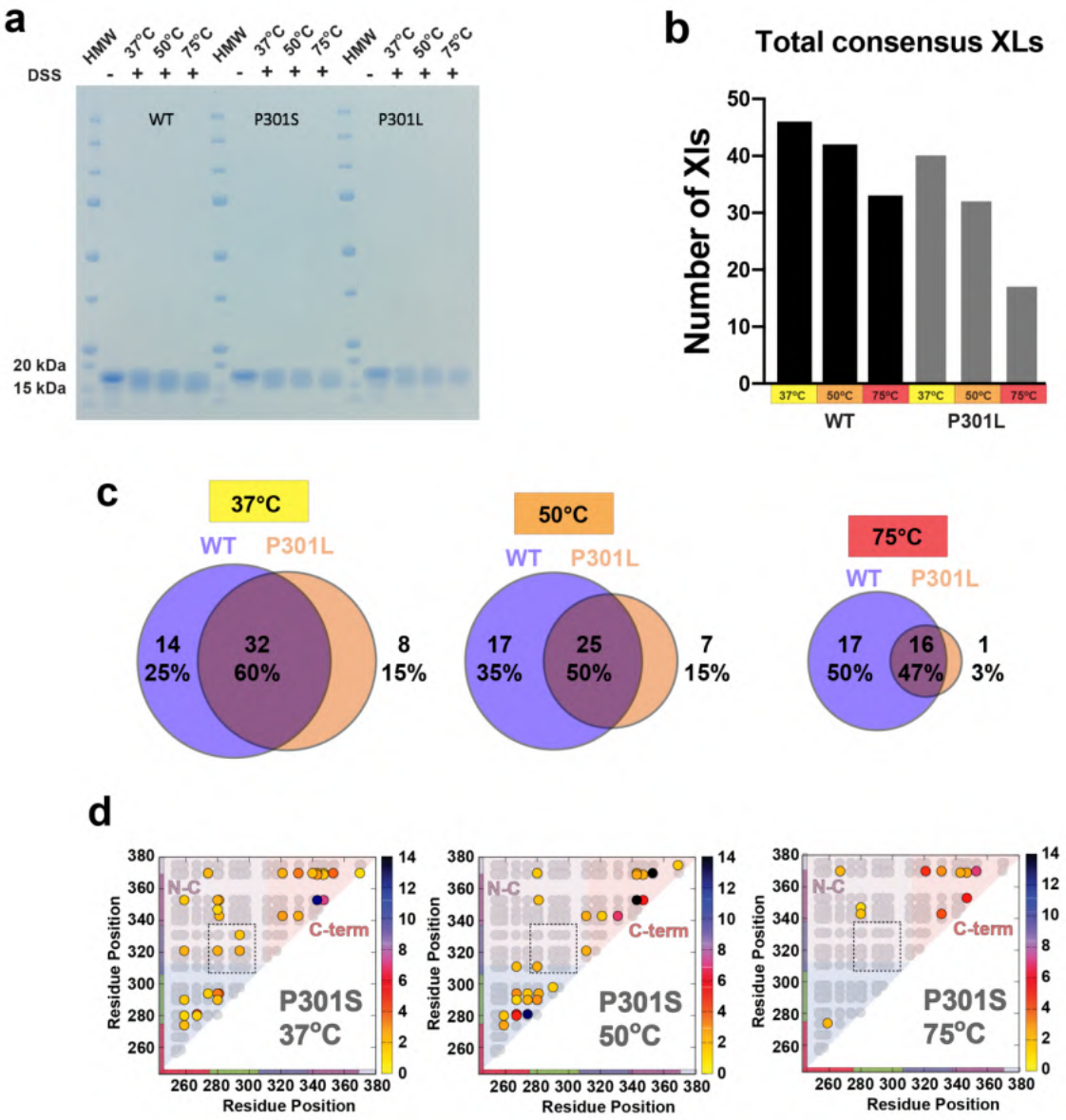


Figure 8. Molecular model of tau amyloid domain structural rearrangement and subsequent aggregation. Naïve tau monomer (left) exists with a propensity to form a relatively collapsed structure, which buries the amyloid domain ^{306}VQIVYK^{311}. In the presence of disease-associated mutations, proline isomerization events, or certain splice isoforms, the equilibrium is shifted to disfavor local compact structure. This exposes the aggregation prone ^{306}VQIVYK^{311} amyloid motif and enhances aggregation propensity, leading to subsequent tau pathology.

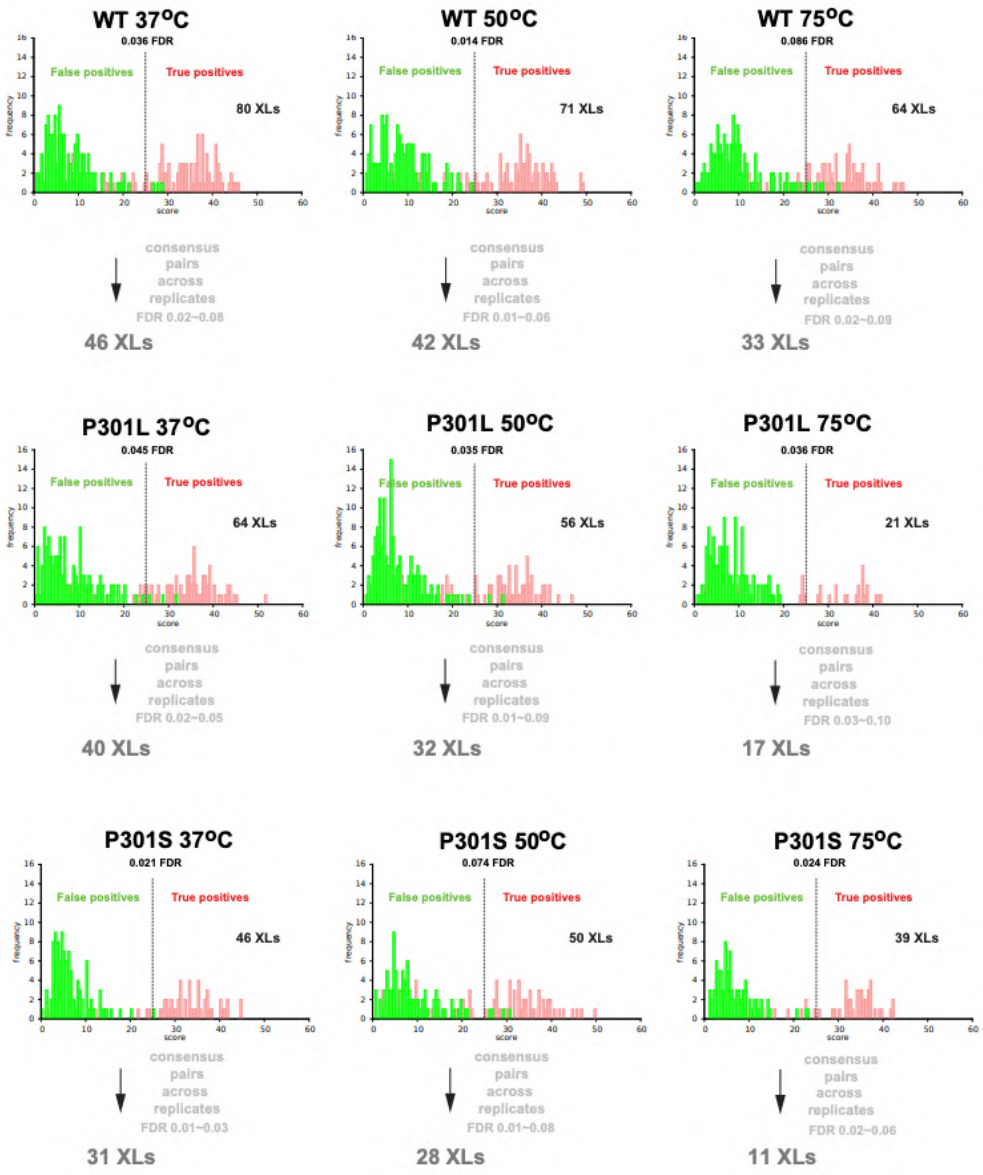


Supplementary Figure 1. TEM of tau and tau RD aggregates. Transmission electron microscopy images of each protein sample from previous ThT experiments after 5 days (Figure 1 c-e). The black bar represents 200 nm or 500 nm distance in each image.

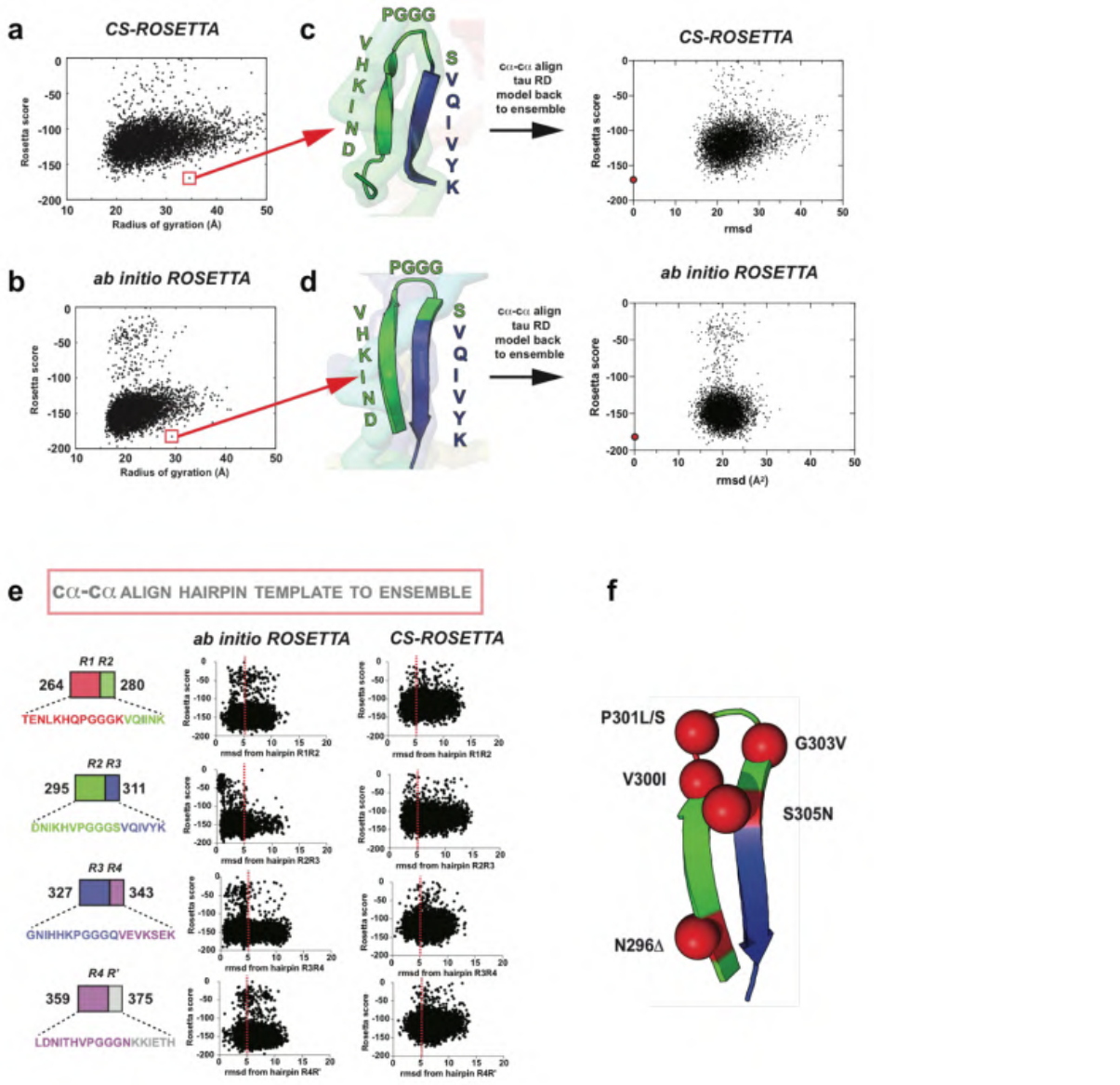


Supplementary Figure 2. Comparison of consensus crosslinks between WT, P301L and P301S tau RD. (a) SDS-PAGE/coomassie of control and DSS reacted samples, indicating the protein remains a monomer following cross-linking (indicative of intra- and not inter-crosslinks) for the WT and mutants at each temperature. (b) Total consensus crosslinks (crosslinks found in all 5 of 5 replicates) in WT (black) and P301L (dark gray) tau RD at 37ºC (yellow), 50ºC (orange) and 75ºC (red). (c) Venn diagrams showing overlap between WT (purple) and P301L (orange) tau RD consensus crosslinks acquired at 37ºC, 50 ºC and 75ºC. (d) P301S tau RD samples were incubated at 37ºC, 50ºC or 75ºC for one hour. After cross-linking, trypsin fragmentation, and LC-MS/MS analysis, consensus cross-link patterns (circles) are shown as contact maps color coded by average frequency across replicates. The theoretical contacts are shown in the background as gray circles. Short range crosslinks

within the N-term (blue), C-term (red) and contacts across N- to C-term (purple) are shown as sectors. The x- and y-axis are colored according to repeat number as in Figure 1. The dashed boxes define inter-repeat crosslinks observed between repeat 2 and repeat 3.

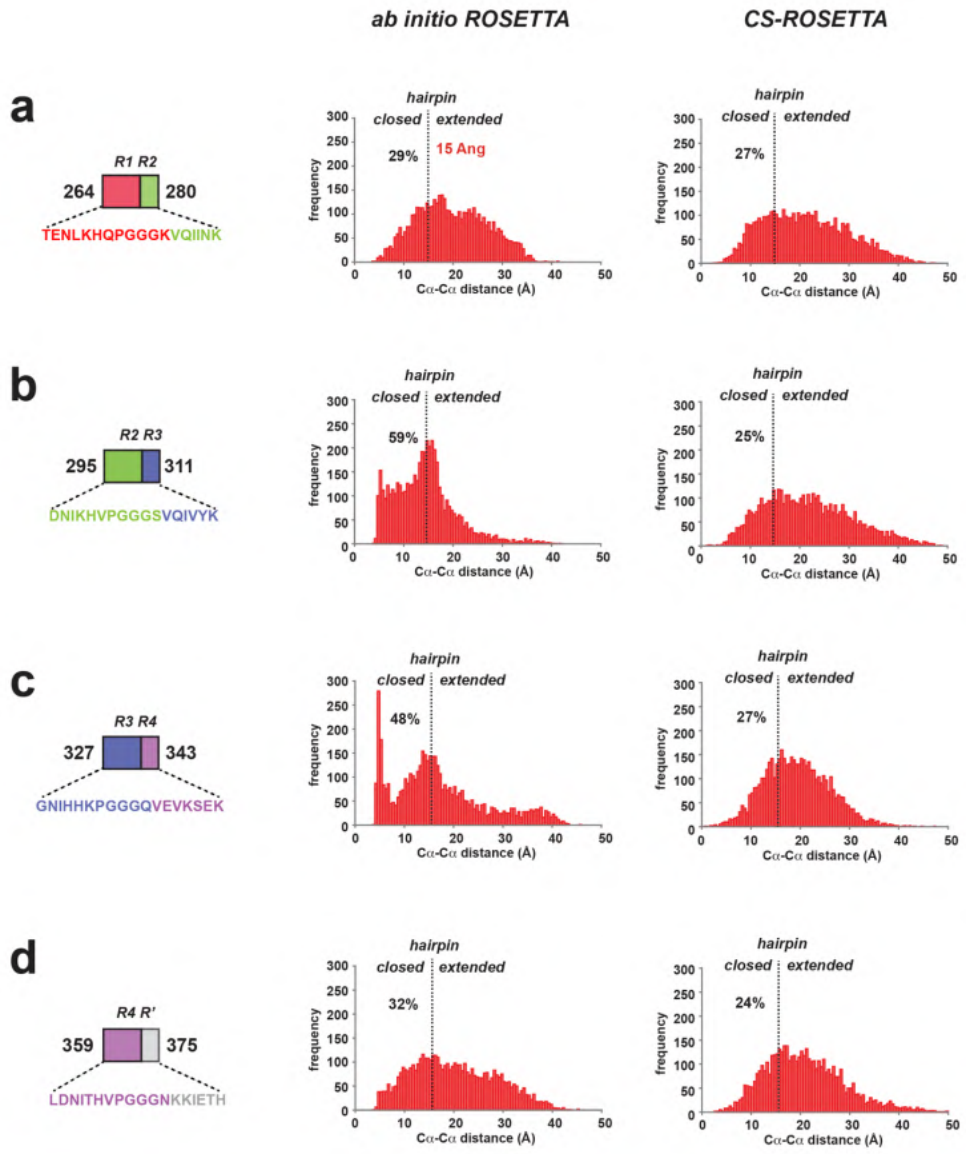


Supplementary Figure 3. Quantification of false discovery rate in crosslink identification. Representative true positive (red) and false positive (green) distribution plots separated by score-id to calculate false discovery rates (FDRs) for each XL-MS dataset. A cutoff with score=25 (vertical dashed lines) is applied to estimate FDR (Methods). FDR is a qualitative measure calculated as ([Number of false positives with score ≥ 25] / [total number of crosslinks with score ≥ 25]) or (1 / [total number of crosslinks with score ≥ 25]) if there is no false positive above score 25.

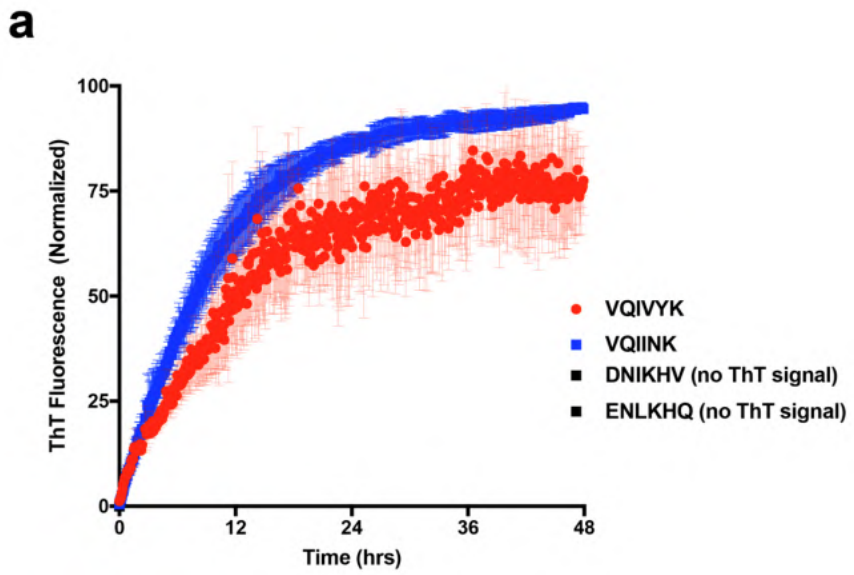


Supplementary Figure 4. ROSETTA-based simulations of tau RD reveal possible hairpin structures. The structural ensembles are shown as a distribution of total energy of each model and radius of gyration for (a) tau RD CS-ROSETTA and (b) tau RD ab initio. (c) Secondary structure models of the repeat2-repeat3 interface (295DNIKHVPGGGSVQIVYK311) predicted from CS-ROSETTA (d) or ab initio shown in cartoon and colored by repeat domain. The amyloid motif 306VQIVYK311 is in blue, the leading repeat-2 sequence is green. Alignment of the structural ensemble to the template model highlights the broad diversity of structures produced by the simulation. The data are shown as Rosetta score versus cα-cα root mean square deviation (rmsd) plots. (e) Pairwise alignment of template hairpin for R1R2, R2R3, R3R4 and R4R' region to each member in the ensemble is shown as a function of Rosetta score revealing all hairpin containing models.

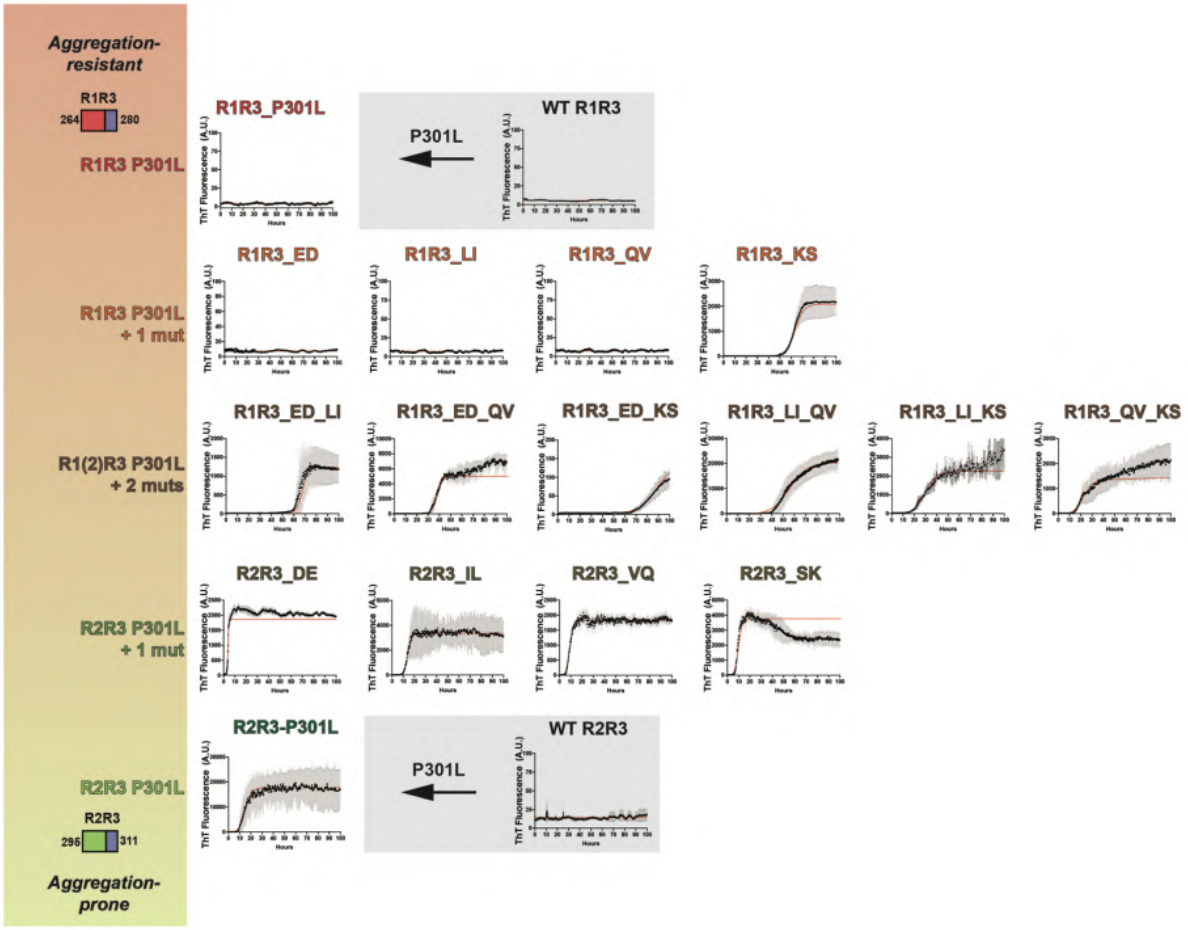
A 5 Å2 rmsd was used as a threshold to define similarity to a template hairpin. The regions and sequences used in the alignment are shown on the left and colored according to the repeat domain as in Figure 1. (f) Positions of disease-associated mutations (red spheres) mapped onto the 306VQIVYK311-containing ab initio β-hairpin structure.



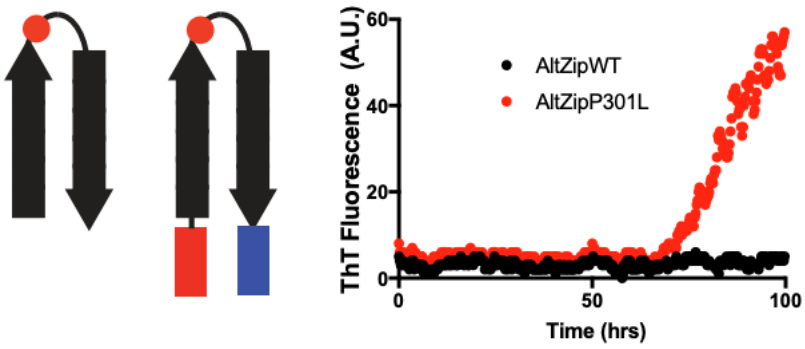
Supplementary Figure 5. Energetics and hairpin content in ROSETTA structural ensembles. (a-d) Ensemble-wide distance distribution of end-to-end distances between residues 264-280, 295-311, 327-347 and 359-375 across all inter-repeat hairpins from ab initio and CS-ROSETTA models. A 15 Å cα-cα distance was used as a threshold to define a collapsed hairpin-like structure.



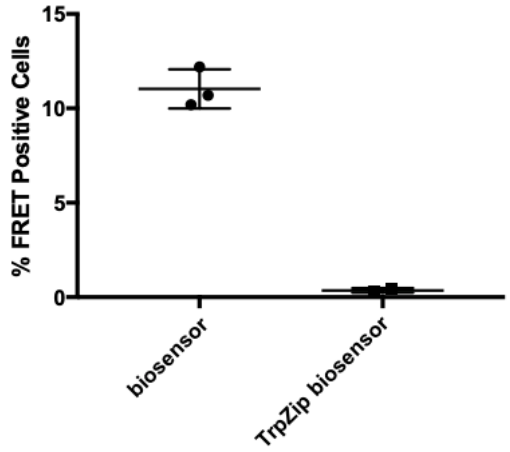
Supplementary Figure 6. An amyloidogenic hexapeptide motif is sufficient to aggregate spontaneously. The aggregation of hexapeptides representing recognized amyloidogenic regions of tau (275VQIINK280 and 306VQIVYK311) or the N-terminal flanking sequences (264ENLKHQ269 and 295DNIKHV300) were studied using ThT fluorescence. The signals are shown as the average of triplicates with standard deviation. The 264ENLKHQ269 and 295DNIKHV300 peptides yielded no detectible ThT signal change (less than 2-fold ratio to background signal) over the course of the experiment. ThT signals are shown as average of triplicates with standard deviation and are normalized to the maximum for each condition.



Supplementary Figure 7. Aggregation kinetics of chimeric peptides derived from R1R3 and R2R3 sequences. The ThT curves for 16 peptides containing single and double mutations are shown as an average from technical triplicates with standard deviations. The mutants are colored according to repeat composition R1R3-P301L (red), R1R3-P301L single mutants (+1 mut, orange), R1(2)R3-P301L double mutants (+2 muts, brown), R2R3-P301L single mutants (+1 muts, dark green) and R2R3-P301L (green). The R1R3-WT and R2R3-WT peptide controls are shown in grey. The data were fit to a non-linear regression model (red line) to derive t1/2 values shown in Figure 6.



Supplementary Figure 8. A salt bridge stabilizes and rescues P301L mutant spontaneous aggregation. An alternative method to stabilize the ends of predicted β-hairpin was used in addition to a tryptophan zipper. Two additional glutamic acid and lysine were placed at the flanking N- and C-termini respectively (Supplementary Table 2). Doing so delayed aggregation of R2R3-P301L (t1/2 = 7 hours; Figure 4) by over an order of magnitude (t1/2 > 70 hours).



Supplementary Figure 9. Tau biosensor cells with enhanced local structure are resistant to being seeded by exogenous aggregates. Tau biosensor cells were constructed with a tryptophan zipper as what was used in the trp-R2R3-P301L-trp peptide fragment, but in the context of tau RD-CFP/YFP (Methods). When transduced with exogenous tau aggregate material, the TrpZip encoded biosensors showed a significant reduction in capacity for seeding. % FRET positive cells were measured using Flow Cytometry in triplicates with at least 10,000 cells per condition. FRET signals are shown as average of triplicates with standard deviation.

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 1**. Each plasmid was transformed into BL21-Gold (DE3) cells. Cells were grown in 1x Terrific Broth media to OD600 1.4 and induced with 1 mM IPTG for 3 hours at 37ºC. The cells were harvested and lysed in 50 mM Tris, 500 mM NaCl, 1 mM β-Mercaptoethanol, 20 mM imidazole, 1 mM 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, 1mM β-Mercaptoethanol (pH 6.0) by PD-10 column GE. Exchanged fractions were applied to a HiTrap SP HP (GE) and eluted with a 50mM-1M 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 1x PBS (136.5 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). pet28b-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 1**. Tau RD wildtype and mutants were expressed the same way as full-length tau. The cells were harvested and lysed in 1X BRB-80 (80mM K-PIPES, 1mM $MgSO_4$, 1mM EGTA, pH 6.8), 0.1% β-ME, 1mM 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 500RPM for 20 min. The supernatant after centrifugation was filtered using 0.22um filter and loaded on HiTrap SP HP (GE) and eluted with a 50mM-1M 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 1x PBS (136.5 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Aliquots were all stored at -80ºC in 1x PBS. Tau seeding monomer (M_s) was produced as previously described[16]. Specifically, 16uM WT tau was incubated with heparin (Amsbio) at a 1:1 ratio for 1hr at 37^{o}C in 1xPBS. The reaction was resolved on a Superdex 200 Increase 10/300 GL (GE) equilibrated in 1x PBS. The M_s peak eluted at 12mls, the concentration of the fraction was measured, the sample aliquoted and flash frozen in liquid nitrogen.

ThT fluorescence aggregation assays

Wildtype or mutant FL tau and tau RD protein was diluted in 1x PBS with 5µM β-Mercaptoethanol and boiled at 95ºC for 5 min. A final concentration of 4.4µM heparin (Amsbio) or 33nM M_s seed was added to 4.4µM tau or tau RD protein in a 60µL volume mixed with 25 µM ThT and aliquoted into a 96-well clear bottom plate. Peptides were disaggregated as previously described (O'Nuallain, 2006). Briefly, peptides were resuspended in a 1:1 mixture (v/v) of TFA (Pierce) incubated at room temperature (RT) for 1 hour. 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 2x PBS to a 200uM concentration to adjust the peptide to buffered reaction conditions. 25 μM ThT was added to 200uL of 200 μM peptide in a 96-well clear bottom plate. All conditions were done in triplicates (except for the R2R3-IEZip experiment) at room temperature. ThT kinetic scans were run every 5 minutes 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 greater than 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 seconds, washed with distilled water, and then negatively stained with 2% uranyl acetate for 1 min. Images were acquired on a Tecnai G^2 spirit transmission electron microscope (FEI, Hillsboro, OR), serial number: D1067, equipped with a LaB_6 source at 120kV 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, 10μL of 4.4μM aggregated protein material was mixed with 1.25μL lipofectamine and 8.75μL Opti-MEM,

incubated at room temperature 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 (1x HBSS, 1% FBS, 1 mM EDTA, 1x DPBS) and analyzed by flow cytometry. For peptide experiments, 10 μg of aggregated peptide material was added to 0.5 uL lipofectamine and Opti-MEM to a total volume of 10 μL, incubated at room temperature 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 described (Holmes, 2014). Briefly, the FM5-YFP and FM5-CFP vectors were digested with NdeI (NEB) and ApoI (NEB). The P301L TrpZip tau RD fragment was ordered as a geneblock (IDT) (see **Supplementary Table 1**). Gibson assembly (NEB) was used to insert the fragment into the plasmid. To produce biosenors, HEK 293T 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 TrpZip 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. Briefly, 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 and 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).

XLMS sample preparation and mass spectrometry

Preparation of tau RD was crosslinked at a total protein concentration of 1.0 mg/mL using 100 µg of starting material. The crosslinking buffer was 1X PBS with 3mM 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 hour before crosslinking. The crosslinking reaction was initiated by adding disuccinimidyl suberate (DSS) stock solution (25 mM DSS-d_0 and $-d_{12}$, 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 37ºC 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 5mM iodoacetamide (30min, room

temperature, 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 37ºC 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 approximately 0.5 µg/µl. 2µL 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 (11cm x 0.075mm I.D.) containing Magic C_{18} material (Michrom, 3µm particle size, 200Å pore size) at a flow rate of 300nL/min using the following gradient. 0-5min = 5 %B, 5-95min = 5-35 %B, 95-97min = 35-95 %B and 97-107min = 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%, 30ms activation). Fragment ions were detected at low resolution in the linear ion trap. Dynamic exclusion was enabled (repeat count 1, exclusion duration 30sec).

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 xQuest

(Walzthoeni, 2012). 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 crosslinking 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, MS^1 tolerance = 10 ppm, MS^2 tolerance = 0.2 Da for common ions and 0.3 Da for crosslink 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 +3.8ppm, %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 crosslink. False discovery rates (FDR's) for the identified crosslinks were estimated using xprophet (Walzthoeni, 2012) and estimated to be 1.3 - 10% (**Supplementary Figure 3**). At each temperature, the 5 replicate datasets were compared and only crosslinks present in 5 of the 5 datasets were used to generate a consensus dataset. Crosslink data with information of crosslinked residue positions and nseen (frequency) was 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 (bmrbId=19253) were used in CS-Rosetta to generate fragment libraries for subsequent model refinement (Shen, 2009). 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 protein (Shen, 2009). 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 tool (Ovchinnikov, 2018). The Rosetta energy function was used to assemble and iteratively refine 5000 structural models using each set of fragments (Ovchinnikov, 2018; Lange, 2012; Raman, 2010). Cα–based root mean square deviations were computed in Rosetta for tau RD and hairpin fragments. Clustering analysis of the tau RD ensemble showed similar results yielding median root mean square deviations of 19.5 Ang^2 and 19.75 Ang^2 for *ab initio* and CS-ROSETTA simulations, respectively. Ensemble wide calculation of Cα-Cα end to end distances between residues 264-280, 295-311, 327-343 and 359-375 were carried out using a python script. 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 Metadynamics (Barducci, 2008) 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 landscapes (Lin, 2017). The initial dodecahedron simulation box was constructed from a trimer of a randomly unfolded structure of 295-311 by adding 7587 SPCE explicit waters and 3 neutralizing Cl ions (one for each monomer). The AMBER99sb-ildn force-field (Lindorff-Larsen, 2010) 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 Berendsen (Berendsen, 1984) and the last 10 with Parrinello-Rahman (Parrinello, 1981) barostats) equilibrations were performed. The subsequent production level trajectories are based on 5 fs time steps using hydrogen-only virtual sites (Feenstra, 1999). Production level trajectories were obtained for an NPT ensemble with Parrinello-Rahman barostat, and periodic boundary conditions with Particle Mesh Ewald (PME) (Darden, 1993) summation for long-range electrostatics. The tuned well-tempered metadynamics parameters are 10, 1.4 kJ/mole, 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 package (Tribello, 2014) as an external patch to Gromacs-5.0.4 (Abraham, 2015). A total of 18 μs trajectories were generated, 9 μs for wild type and 9 μs for the P301L mutant, over a total of 6 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 3 independent experiments, whose values are plotted. Error bars represent a 95% CI.

Code availability

All tau RD simulations were carried out with CS-ROSETTA and ROSETTA (available at https://www.rosettacommons.org/). All peptide molecular dynamics 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)

Supplemental Tables

Supplemental Table 1. List of Protein Sequences

Name	Amino Acid Sequence	
WT tau RD	R1: 244	QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK 274
	R2: 275	VQIINKKLDLSN-VQSKCGSKDNIKHVPGGGS 305
	R3: 306	VQIVYKPVDLSK-VTSKCGSLGNIHHKPGGGQ 336
	R4: 337	VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN 368
	R′: 369	KKIETHKLTFRE 380
P301L tau RD	R1: 244	QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK 274
	R2: 275	VQIINKKLDLSN-VQSKCGSKDNIKHVLGGGS 305
	R3: 306	VQIVYKPVDLSK-VTSKCGSLGNIHHKPGGGQ 336
	R4: 337	VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN 368
	R′: 369	KKIETHKLTFRE 380
P301S tau RD	R1: 244	QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK 274
	R2: 275	VQIINKKLDLSN-VQSKCGSKDNIKHVSGGGS 305

	R3: $_{306}$ VQIVYKPVDLSK-VTSKCGSLGNIHHKPGGGQ $_{336}$
	R4: $_{337}$ VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN $_{368}$
	R′: $_{369}$ KKIETHKLTFRE $_{380}$
WT tau	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQT
	PTEDGSEEPGSETSDAKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEG
	TTAEEAGIGDTPSLEDEAAGHVTQARMVSKSKDGTGSDDKKAKGADGKTK
	IATPRGAAPPGQKGQANATRIPAKTPPAPKTPPSSGEPPKSGDRSGYSSP
	GSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAKSRLQTAPVPM
	PDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV
	PGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRV
	QSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVS
	GDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSASLAKQGLLEHHHHHH
P301L tau	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQT
	PTEDGSEEPGSETSDAKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEG
	TTAEEAGIGDTPSLEDEAAGHVTQARMVSKSKDGTGSDDKKAKGADGKTK
	IATPRGAAPPGQKGQANATRIPAKTPPAPKTPPSSGEPPKSGDRSGYSSP
	GSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAKSRLQTAPVPM
	PDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV
	LGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRV
	QSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVS
	GDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSASLAKQGLLEHHHHHH

Supplemental Table 2. List of Peptide Sequences

[1]Fluorinated proline analogs were used to preferentially populate *cis, trans,* or lowered E_a barrier conformers at the positions indicated as p.

Name	Amino Acid Sequence
VQIINK	----------VQIINK
VQIVYK	----------VQIVYK
R1R2 (264-280)	------ENLKHQPGGGKVQIINK
R2R3 (295-311)	------DNIKHVPGGGSVQIVYK
R2R3-N296Δ	------D-IKHVPGSSSVQIVYK
R2R3-V300I	------DNIKHIPGGGSVQIVYK
R2R3-P301L	------DNIKHVLGGGSVQIVYK
R2R3-P301S	------DNIKHVSGGGSVQIVYK
R2R3-G303V	------DNIKHVPGVGSVQIVYK
R2R3-S305N	------DNIKHVPGGGNVQIVYK
R1R3	------ENLKHQPGGGSVQIVYK
R1R3-P270L	------ENLKHQLGGGSVQIVYK
R1R3-P270S	------ENLKHQSGGGSVQIVYK
R1R3-G272V	------ENLKHQPGVGSVQIVYK
R1R3-G273R	------ENLKHQPGGRSVQIVYK
Trp-R2R3-P301L-Trp	WTGKSKDNIKHVLGGGSVQIVYKEGGW
Trp-R2R3-P301L	WTGKSKDNIKHVLGGGSVQIVYKPVDL
R2R3-P301L-Trp	SKCGSKDNIKHVLGGGSVQIVYKEGGW
R2R3-IEZip	----DDDNIKHVPGGGSVQIVYKKK
R2R3-IEZip-P301L	----DDDNIKHVLGGGSVQIVYKKK
R2R3-Cis[1]	------DNIKHVpGGGSVQIVYK
R2R3-Trans[1]	------DNIKHVpGGGSVQIVYK

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, crosslink 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 molecular dynamics simulations. K.D., D.C. and L.A.J. wrote the manuscript, and all authors contributed to its improvement.

Acknowledgement

This chapter was replicated in full from my co-first author publication (Chen *et al. Nature Communications* 2019)

CHAPTER FOUR

FTD-TAU S320F MUTATION IN TAU STABILIZES LOCAL STRUCTURE AND ALLOSTERICALLY PROMOTES AMYLOID MOTIF-DEPENDENT AGGREGATION

Introduction

The microtubule-associated protein tau deposits as beta-sheet-rich structures in over 25 neurodegenerative diseases commonly referred to as tauopathies that include Alzheimer's (AD), Corticobasal degeneration (CBD), Picks (PiD) and Chronic traumatic encephalopathy (CTE). These diseases are most often sporadic and are not linked to mutations in the microtubule-associated protein tau gene (*MAPT*). Over the last 30 years, more than 50 disease-associated mutations have been identified in *MAPT* linking this gene to Frontotemporal dementia with tau (FTD-tau). While the mechanistic basis of FTD-tau mutants has yet to be characterized, several functional impacts on tau have been described that include decreased capacity to bind and stabilize microtubules, disruption of splicing regulation leading to changes in isoform production, and directly enhancing tau aggregation kinetics. To date there are no available cryo-EM structures of tau fibrils isolated from patients harboring *MAPT* pathogenic mutations despite these mutations being commonly used in cell and animal models to study tau disfunction and aggregation. The most commonly employed *MAPT* mutation in models of tau aggregation and disease encode a tau mutant in which proline at position 301 is substituted to serine or leucine. This P301S/L mutation is thought to decrease microtubule binding but also to enhance aggregation kinetics

(zweckstetter and Chen et al). A relatively new mutation in *MAPT* changes serine 320 to phenylalanine that causes familial frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) was identified in 2002. Across a panel of FTDP-17 *MAPT* mutations, S320F is the only one that can spontaneously aggregate when expressed in the mammalian cells (Strang, 2018). Under normal conditions tau is largely disordered (Mylonas, 2008) and is rather resistant to aggregation. Aggregation studies with recombinant tau are made possible with inducers such as heparin, RNA or other polyanions while in cell or animal models, an exogenous seed is required. Cryo-EM structures of recombinant polyanion induced tau fibrils are different from ex vivo patient conformations questioning the physiological relevance of these artificially induced states. Most disease-associated mutations on tau enhance aggregation propensity but still require addition of inducers or seeds aggregate. The S320F mutation in tau may circumvent this problem as it was shown to aggregate spontaneously in cell (Strang, 2018).

The S320F mutation was first reported by the van Swieten group in 2002. The proband presented presenile onset of dementia and other symptoms similar to Pick's Disease at age 38 and passed away at age 53 (Rosso, 2002), a relatively young age compared to an average age at symptom onset of 49 years among FTD associated with *MAPT* mutations (Ghetti, 2014). Immunohistochemistry of tau inclusions indicated presence of "Pick-like" bodies and "more diffused" staining. Both 3-repeat and 4-repeat tau were found in the soluble as well as the insoluble tau fractions by immunochemistry (Rosso, 2002). This evidence indicates that S320F tauopathy is distinct from the Pick's Disease.

Most of the *MAPT* mutations fall into the repeat domain region of tau (tau RD), which has been demonstrated to be crucial for microtubule binding and composes the core of tau fibrils formed in AD (Fitzpatrick, 2017; Falcon 2018), PiD (Falcon, 2018), CBD (Zhang, 2020), progressive supranuclear palsy (PSP), CTE (Falcon, 2019), globular glial tauopathy (GGT) (Shi, 2021) and argyrophilic grain disease (AGD) (Shi, 2021). The largely disordered nature of monomeric tau makes it refractory to high-resolution structural biology methods. However, evidence such as nuclear magnetic resonance (NMR) has shown the repeat domain region of tau (tau RD) to be more structured as compared to the rest (Eliezer, D. et al., 2005). The repeat domain contains four imperfect repeats each separated by a PGGG motif, which is prone to form a β-turn. Immediately following the PGGG at the end of repeat 2, it lies a well-characterized amyloid motif, ^{306}VQIVYK^{311} (Sawaya, M. R. et al. 2007; von Bergen, M. et al. 2001). A number of mutations fall into the inter-repeat interfaces. We have previously characterized that mutations at inter-repeat interfaces, such as P301L/S, drive tau aggregation by destabilizing local structures encompassing the amyloid motif, ^{306}VQIVYK^{311}, and the more frequent exposure of amyloid motif leads to enhanced self-assembly (Chen, 2019). S320F is however located in the middle of repeat 3 and does not fall into the inter-repeat interface suggesting it may follow an alternative structural mechanism that underlies its unique capacity to spontaneously self-assemble.

Here we employed a series of *in silico*, *in vitro*, and cellular assays to delineate the mechanism of how S320F may drive spontaneous aggregation of tau. Our data support a model where sequence motifs preceding PGGG in repeat 2 and repeat 3 provide synergistic protection on the amyloid motif ^{306}VQIVYK^{311} and the disengagement of this protection

driven allosterically by S320F mutation exposes ^{306}VQIVYK^{311} and causes enhanced self-assembly.

Results

S320F aggregates readily in vitro in the absence of an inducer and spontaneously aggregates in cells

Prior studies on S320F tau have shown that it can form seeds in vitro that can be detected in cell-based models of tau aggregation (Strang, 2018). To understand the mechanism of S320F tau aggregation we sought to reproduce this initial observation using 4R tau repeat domain (tauRD; aa 243 – 380), 3R tauRD (aa 243 – 380 missing 275 – 305) and full-length (FL) 2N4R tau under spontaneous aggregation conditions in the absence of polyanionic inducers such as heparin. We produced recombinant FL 2N4R tau, tauRD and 3R tauRD and compared aggregation propensity of WT, S320F and a previously characterized FTD-tau P301S mutant (Fig. 1a). We evaluated spontaneous aggregation of our purified proteins in a Thioflavin T fluorescence aggregation assay at a series of concentrations. We find that tauRD S320F was able to spontaneously aggregate rapidly without any inducer at 62.5 μM (Fig 1x, $t_{1/2max}$ = 1.30 ± 0.07 hrs). TauRD P301S was also able to aggregate spontaneously at this concentration but signal did not increase until after 72 hrs. TauRD WT at 62.5 uM remained flat within the 96 hr experiment (Fig. 1b). At a lower concentration, 25 μM, tauRD S320F again aggregated spontaneously with a $t_{1/2max}$ of 1.49 ± 0.03 hrs but to a lower fluorescence magnitude while both tauRD WT and P301S remained flat. Fibrils were detected by TEM in the samples of tauRD S320F at 25 uM and 62.5 uM as

well as tauRD P301S at 62.5 uM but not in the P301S at 25 uM or any concentrations of tauRD WT (Fig. 1c). Repeating the experiment at 10uM, tauRD S320F assembled with slightly slower kinetics relative to the higher concentrations ($t_{1/2max}$ = 3.09 ± 0.32 hrs) whereas tauRD WT and P301S remained flat (Supplementary Fig. 1a). In addition, we tested FL 2N4R S320F tau and found it to aggregate spontaneously at 10 uM while the FL 2N4R WT tau counterpart stayed flat. As both 3R tau and 4R tau isoforms were showed to be present in the sarkosyl insoluble fractions of the FTD-tau S320F patient brain (Rosso, 2002), we directly tested a concentration titration of 3R tauRD and tauRD in a ThT fluorescence assay (Supplementary Fig. 1c). Consistent with prior data (Chen, 2019), we find that the 3R tauRD to be less aggregation prone than 4R tauRD where only at the highest concentration, 100 uM, did 3R tauRD spontaneously aggregate. Presence and absence of fibrils for tauRD at 10uM, 2N4R tau and 3R tau RD series were confirmed by TEM (Supplementary Fig. 1d).

To test the structural compatibility of fibrils formed by *in vitro* tau models cellular models of tau aggregation, we employed tau biosensor HEK293T cells that stably express tauRD P301S as fusions to cyan and yellow fluorescent proteins (CFP/YFP) in two separate constructs (Holmes B. B., 2014; Sanders, D. W. et al. 2014). These cells report a fluorescence resonance energy transfer (FRET) signal between the YFP and CFP tauRD constructs only when treated with exogenous tau seeds. Reported in % FRET positive cells, tauRD S320F at 25 uM after 2 hr incubation at 37 °C *in vitro* was able to induce tauRD seeding in cells and the seeding activity was elevated if the cells were treated with the 4 hr *in vitro* incubated S320F tauRD (Supplementary Fig. 1e). WT tauRD at any incubation

timepoints did not induce seeding consistent with in vitro ThT and TEM quantification of samples (Supplementary Fig. 1e).

We finally tested the spontaneous aggregation behavior of S320F tauRD directly in cells. TauRD WT or with S320F or P301S mutations were expressed as a C-terminal fusion to mEos3.2 in HEK293T cells (Fig. 1d). mEos3.2 is a FRET compatible photoconvertible fluorescent protein which emits green fluorescence and after irradiation with UV, will emit red fluorescence that can be excited by the emission fo the green fluoraphore (Fig. 1d). Cells expressing the three tauRD- mEos3.2 constructs were partially green to red photoconverted to an optimal ratio of green and red fluorescence suitable for FRET quantification. Each cell line was monitored for spontaneous aggregation as determined by FRET. Cells were fixed and analyzed by flow cytometry at two time points, days 4 and 7 to ensure sufficient number of cells to collect sufficient cell population statistics. We found that tauRD S320F showed an increase in FRET as a function of tauRD-mEos3.2 expression and photoconversion, indicated by levels of the photoconverted red mEos3.2 fluorescence signal (Fig. 1e) by gating different levels of red fluorescence intensity in each experiment (Supplementary Fig. 1g).. Neither WT nor P301S tauRD-mEos3.2 constructs showed positive FRET even at high expression and photoconversion levels (Fig. 1e). Similar data was obtained for cells at Day 7, although the level of FRET was lower (Supplementary Fig. 1f) likely due to toxicity of the formed aggregates. Representative images of cells expressing S320F, P301S or WT tauRD-mEos3.2 showed puncta for S320F but P301S or WT remained diffuse (Fig. 1f). Our data support that the S320F mutation in tau drives spontaneous tau aggregation in vitro and in cells.

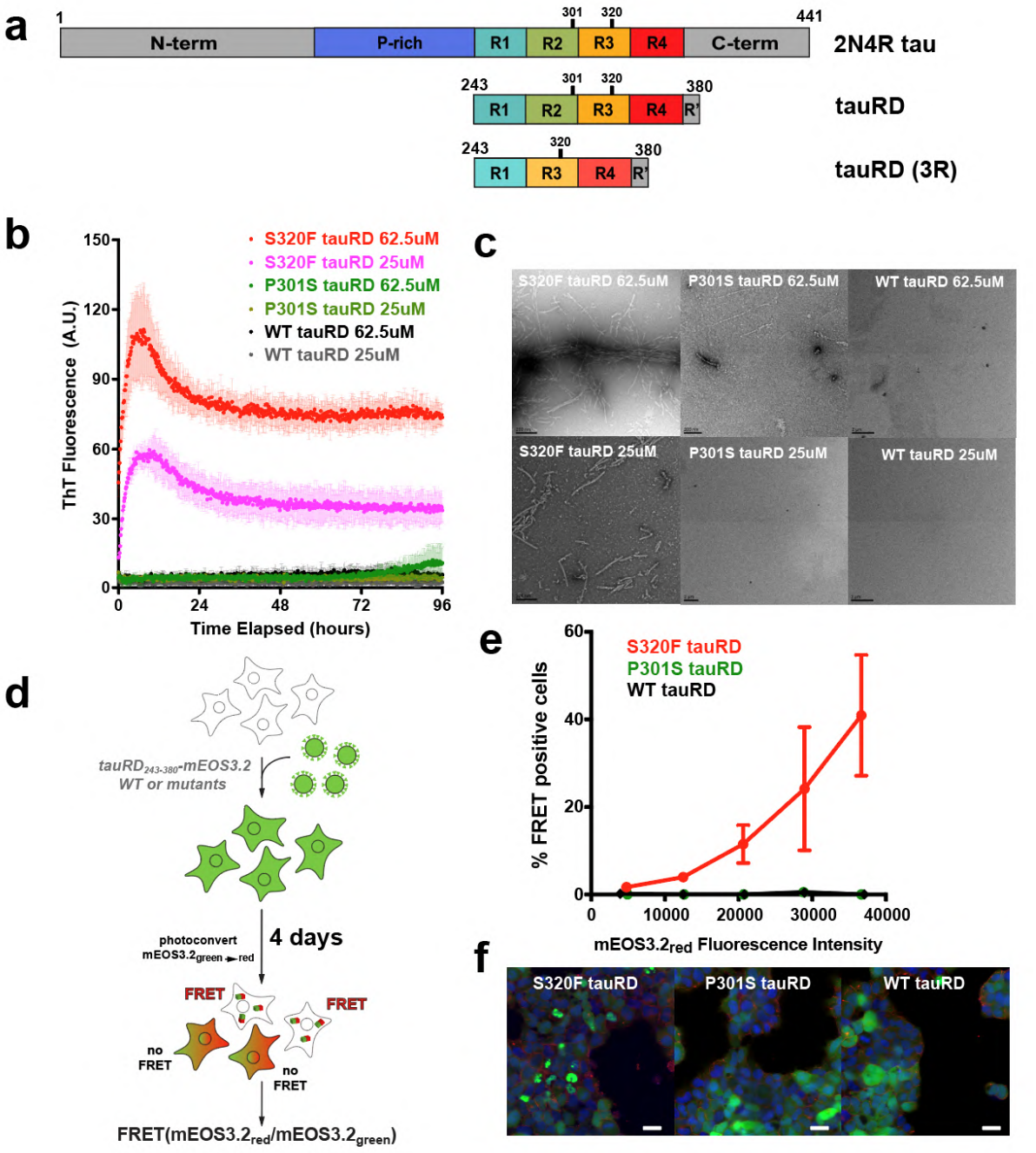


Fig. 1: S320F causes spontaneous tau aggregation in vitro and in cells.
a schematic of 2N4R tau and tauRD where N-term = gray, proline-rich (P-rich) = purple, repeat 1 = cyan; repeat 2 = green; repeat 3 = yellow; repeat 4 = red, R' = gray, C-term = gray. Relative positions of 301 and 320 are indicated by black bars. **b** ThT fluorescence assay on tauRD S320F (red and magenta), P301S (green and olive) and WT (black and gray) at 62.5 uM and 25 µM, 37 °C. **c** Representative Transmission Electron Microscopy images of the

endpoint ThT assay on tauRD S320F, P301S and WT at corresponding concentrations. The black bar represents 200 nm distance in 3 images including S320F tauRD 62.5 uM, S320F tauRD 25 uM, and P301S tauRD 62.5 uM, or 2 um in P301S tauRD 25 uM, WT tauRD 62.5 uM and WT 25 uM. **d** Schematic of cell model assay to monitor aggregation activity on mEos3.2 fused WT or mutant tauRD. HEK293T cells expressing S320F, P301S or WT tauRD-mEos3.2 were fixed at Day 4. After photoconversion with UV, a portion of $mEos3.2_{green}$ will be converted to $mEos3.2_{red}$ and FRET can then be detected. **e** FRET (tauRD-CFP/tauRD-mCherry) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition of S320F (red), P301S (green)), or WT (black). Comparison was conducted at multiple fluorescent intensity levels of $mEos3.2_{red}$. Error bars represent a 95% CI of each condition. **f** Representative images of S320F, P301S, and WT tauRD-mEos3.2 expressed HEK293T cells prior to photoconversion. mEos3.2 (green), Hoechst33342 (blue, nuclei stain), and Wheat Germ Agglutinin (red, cell membrane stains) fluorescence signals are shown in green and blue, respectively. Scale bar, 15 μm, is shown in white.

S320F engages with local sequence and controls VQIVYK-based aggregation

In our previous study, we identified that minimal aggregation regulatory elements in tau based on short peptide fragments encoding inter repeat fragments (i.e. R2R3, ^{295}DNIKHVPGGGSVQIVYK^{311}) that normally adopt a beta-turn stabilized by the P-G-G-G motif. Derived from this work we proposed a model that beta-turn stability is inhibitory to aggregation by engaging with the ^{306}VQIVYK^{311} motif while destabilization of this structure via mutations (i.e. P301S) promotes assembly (Chen, 2019). Since S320F is within R3 and downstream of the ^{306}VQIVYK^{311} motif, we hypothesized that the S320F mutation may allosterically disrupt the aggregation preventing beta-turn conformation. To determine the regulatory elements that play essential roles in regulating S320F aggregation, we designed a series of peptide fragments encoding WT or S320F (Fig. 2a). The peptides were synthesized and their aggregation behavior was evaluated in a ThT fluorescence assay (Fig. 2b). All peptide aggregation assays were performed at 200 uM in PBS at 37^{o}C with shaking. The data

were fit to derive $t_{1/2max}$ values thus facilitating interpretation of aggregation kinetics across the different sequences (Fig. 2a and Supplementary Fig. 2)

We first tested a minimal fragment spanning 316-330 and excluding the ^{306}VQIVYK^{311} motif to test whether the S320F itself introduces an amyloid motif into the sequence that could directly drive aggregation. Neither $WT_{316-330}$ nor $S320F_{316-330}$ were able to aggregate spontaneously at 200 μM within the time frame of the experiment (Fig. 2b, $t_{1/2max} > 72$ hrs and Supplementary Fig. 2a). This suggests that the S320F mutation itself did not introduce a novel amyloid motif. In the next fragment spanning 306-324, we shifted the sequence window upstream to include the well-established ^{306}VQIVYK^{311} amyloid motif (Sawaya, M. R., 2007, von Bergen, M., 2001). The $S320F_{306-324}$ and $WT_{306-324}$ fragments aggregated rapidly preventing accurate capturing of the lag or growth phase ($t_{1/2max} = 0$, Fig. 2b, Supplementary Fig. 2b). The difference in the aggregation behavior of $S320F_{306-324}$ from $S320F_{316-330}$ indicates that the aggregation mechanism of S320F requires the amyloid motif ^{306}VQIVYK^{311}. Aggregation data on these fragments point to an intramolecular role of S320F in facilitating spontaneous aggregation and that the phenylalanine likely allosterically influences ^{306}VQIVYK^{311} based assembly.

From our previous study, we concluded that the local sequence upstream of the amyloid motif, ^{295}DNIKHVPGGGS^{305} shields ^{306}VQIVYK^{311} from aggregation (Chen et al). We were curious what effect S320F mutation will have on the upstream sequence. We tested the aggregation behavior of fragments spanning residues 295-324, $S320F_{295-324}$ and $WT_{295-324}$. As expected, the aggregation of $S320F_{295-324}$ was delayed ($t_{1/2max} = 5.4 \pm 1.0$ hrs) compared to $S320F_{306-324}$ (Fig. 2b, Supplementary Fig. 2c). Surprisingly, $WT_{295-324}$ was able to aggregate

although with slower kinetics ($t_{1/2max}$ = 23.3 ± 4.8 hrs) than the S320F counterpart (Fig. 2b and Supplementary Fig. 2c). This raises a new hypothesis that additional downstream sequence might disturb the protection of ^{306}VQIVYK^{311} from ^{295}DNIKHVPGGGS^{305}. The hydrophobic-rich property of the sequence ^{312}PVDLSKVTSKCGS^{324} might compete the interaction between ^{295}DNIKHV^{300} and ^{306}VQIVYK^{311}. The addition of another hydrophobic residue S320F in $S320F_{295-324}$ might further compete this interaction which could explain the faster aggregation kinetics of $S320F_{295-324}$ than its WT counterpart.

We then shifted the window of fragment sequence downstream of 320 and found that the aggregation of $WT_{306-330}$ was further delayed ($t_{1/2max}$ = 45.2 ± 13.3 hrs) than $WT_{295-324}$ (Fig. 2b, Supplementary Fig. 2d). This suggests that the downstream sequence ^{25}LGNIHH^{330} might additionally contribute to protection of the amyloid motif beyond ^{295}DNIKHV^{300}. As expected, $S320F_{306-330}$ was again able to aggregate more readily ($t_{1/2max}$ = 7.0 ± 1.8 hrs) than its WT counterpart (Fig. 2b, Supplementary Fig. 2d). However, the larger difference in aggregation kinetics between S320F and WT in the context of 306-330 than in 295-324 implies that the downstream sequence of 320 might have a stronger role with 320F than the upstream sequence in promoting aggregation.

If the protection of ^{306}VQIVYK^{311} from either ^{295}DNIKHV^{300} or ^{325}LGNIHH^{330} can be perturbed by introducing S320F mutation, we were curious if both ^{295}DNIKHV^{300} and ^{325}LGNIHH^{330} are present, so we designed and tested fragment 295-330. $WT_{295-330}$ as expected was even further delayed in aggregation ($t_{1/2max}$ = 55.3 ± 4.5 hrs), suggesting that both ends contribute to the inhibition of aggregation driven by ^{306}VQIVYK^{311}(Fig. 2b, Supplementary Fig. 2e). On the other hand, with the putative double protection from both

ends, $S320F_{295-330}$ was able to aggregate with even faster kinetics ($t_{1/2max}$ = 0.8 ± 0.1 hrs) than $S320F_{306-330}$ or $S320F_{295-324}$ (Fig. 2b, Supplementary Fig. 2e), implying that the S320F mutation may negate the ^{295}DNIKHV^{300} and ^{325}LGNIHH^{330} -based protection of ^{306}VQIVYK^{311}. Presence and absence of fibrils were shown by TEM and were consistent with the ThT assays.

Across the pairs of different fragments, largest difference was observed between WT and S320F in fragment 295-330. Moreover, the kinetics of $S320F_{295-330}$ seems to recapitulate the behavior of S320F tauRD and 2N4R tau (Fig. 1b, Supplementary Fig. 1b), where no lag phase and an instant growth phase were observed. Altogether, we decided to use 295-330 fragment for Molecular Dynamics (MD) simulation to further assess on the residue level the effect of S320F. Together our peptide fragments imply a structural mechanism where ^{306}VQIVYK^{311} is essential in S320F aggregation and the sequence motifs ^{295}DNIKHV^{300} in repeat 2 and ^{325}LGNIHH^{330} in repeat 3 might protect the amyloid motif ^{306}VQIVYK^{311} in a synergistic manner. The introduction of S320F might perturb the protections on ^{306}VQIVYK^{311} possibly through interaction with the downstream sequence of 320. The increased exposure of ^{306}VQIVYK^{311} is likely again, similar to the consequence of P301S mutation, the culprit for S320F aggregation.

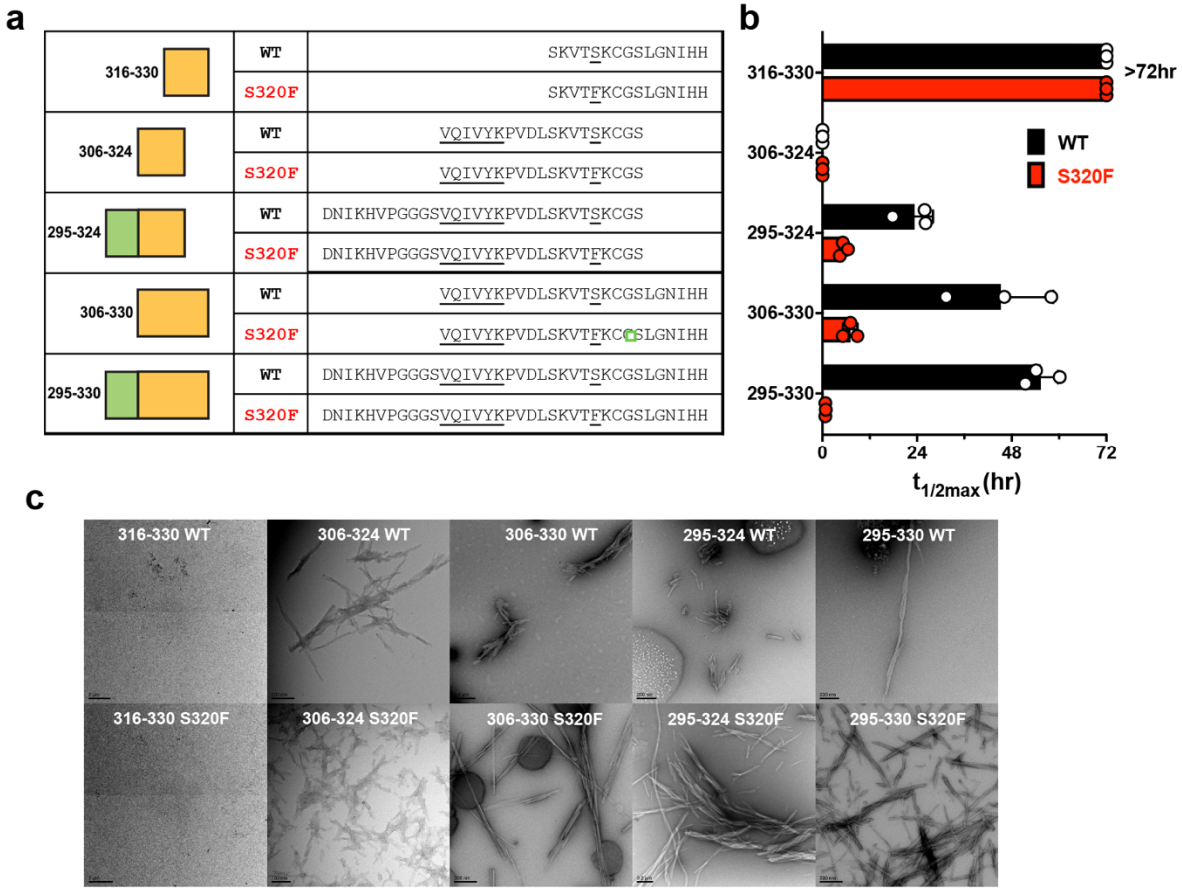


Fig. 2: VQIVYK is essential in S320F facilitated aggregation and the surrounding sequence regulates aggregation. a Design of sequence fragments encompassing 320 positions in WT or S320F context. The 320 position and ^{306}VQIVYK^{311} are underlined. **b** ThT fluorescence signal at 72 hr for each sequence fragments of WT (black) and S320F (red). Error bars represent the standard deviations. **c** Representative TEM images of each peptide from the ThT assays at 72 hr endpoint.

S320F allosterically disrupts the shielding of the amyloid motif

In order to gain molecular insights into the structural mechanism of S320F aggregation behavior, we employed a Molecular Dynamics (MD) approach on two tau peptides that comprised the sufficient regulatory elements encoding the entire repeat 3 and

the C-terminal portion of R2 spanning residues 295-330: $WT_{295-330}$ and $S320F_{295-330}$ fragments (Fig. 3a and Supplementary Table 1). Simulations were started from the pre-minimized structures and were performed as five 3 μs independent trajectories. Cumulative contact maps derived from merged ensembles identified features that may distinguish the WT from the S320F fragment behavior. Consistent with our prior work, the simulations on the WT fragment showed that the ^{306}VQIVYK^{311} motif was engaged in stabilizing interactions with ^{295}DNIKHV^{300} through the P-G-G-G beta-turn stabilizing motif (Fig. 3b). These local N-terminal structural elements however, were on average shorter in $S320F_{295-330}$ than in $WT_{295-330}$ (Fig. 3c), where K311 and several residues C-terminal were more likely included in the ensemble of WT (Fig. 3b), although the middle of the N-terminal local structure was tighter in S320F on average (Fig. 3c). The C-terminal ^{325}LGNIHH^{330} was found also on average in contact with both ^{295}DNIKHV^{300} and ^{306}VQIVYK^{311} as well as several downstream residues in $WT_{295-330}$ (Fig. 3b). However, in $S320F_{295-330}$ these two contacts were weakened and instead two of more local contacts emerged in the C-terminus. The more obvious contacts linked F320 to the C-terminal tail sequence (i.e. ^{325}LGNIHH^{330}) but also sampling long-range contacts to ^{306}VQIVYK^{311} (Fig. 3c). This rearrangement of contacts suggest that the S320F mutation might allosterically control the aggregation driven by ^{306}VQIVYK^{311}. Specifically, S320F might stabilize local structures in the C-terminus which can disturb the protection of ^{325}LGNIHH^{330} on ^{306}VQIVYK^{311} and the reduction of which seemed to disrupt the local N-terminal structure. This implies that ^{295}DNIKHV^{300} and ^{315}LSKVTF^{320} might work synergistically to protect ^{306}VQIVYK^{311}, which could explain why $S320F_{295-330}$ or in the longer context, tauRD, was able to aggregate even with the

previously characterized protector ^{295}DNIKHV^{300} (Chen, 2019). The middle structures of the top five clusters within 0.6 nm RMSD supported our extrapolations on the WT and S320F conformations (Fig. 3d for $WT_{295-330}$ and 3e for $S320F_{295-330}$, Supplementary Fig. 3a). Specifically, these more expanded N-terminal local structures were observed in most of the $WT_{295-330}$ clusters with C-terminus folding back and in close contact with the N-terminal local structure (Fig. 3d). Most structures became more modular for $S320F_{295-330}$, where C-terminus formed its own local structure separate from the N-terminal local structure which became shorter (Fig. 3e).

To determine which residues within 325-330 might be particularly important in the interaction with S320F, we plotted the Cα distance distribution of all residues to 320 (WT, Fig. 3f and S320F, Fig. 3g). Compared to WT which showed more evenly distributed distance to S320 of the downstream C-terminal residues (Fig. 3f), S320F mutant had two residues, L315 and I328, that are far in the sequence position to 320 and each presented a distinct population (>15%) that are within 0.2 nm to F320. Since I328 is farther away in the sequence position, we thought it was the more interesting candidate to test. We hypothesized that hydrophobic contacts between F320 and I328 may play are role in stabilizing alternate interactions that could expose the ^{306}VQIVYK^{311} motif.

To test the role of the F320-I328 interaction, we repeated the MD simulations on the 295-330 fragment encoding S320F and I328S (i.e. $S320F\_I328S_{295-330}$). The cumulative contact map from the merged replicates shows that we regained WT-like contacts between the C-terminal ^{325}LGNIHH^{330} and to ^{306}VQIVYK^{311} and ^{295}DNIKHV^{300} (Fig. 3h). The median structures from the top five clusters from the $S320F\_I328S_{295-330}$ ensemble supported

our observation from the merged contact map (Fig. 3i). The Cα distance distribution of contacts between I328 and F320 in the S320F_I328S$_{295-330}$ fragment were also consistent with the stabilizing hydrophobic interaction between F320 and I328, as the frequency of close distance between the 320 and 328 residues was reduced (Fig. 3j). Fig. 3k summarizes the above extrapolated changes in local structures induced by S320F and the partial repair effect of I328S. In parallel, Cα distance between 295-311, 320-328, and 308-328 for the middle structures of top five clusters were extracted and compared between WT, S320F and S320F/I328S conditions (Supplementary Fig. 3b and c). The top clusters followed the same trend where 295-311 distance was smaller in WT and S320F_I328S than in S320F, 320-328 distance was larger in WT and S320F_I328S, and 308-328 (proxy of VQIVYK to I328) distance was smaller in WT and S320F_I328S (Supplementary Fig. 3c).

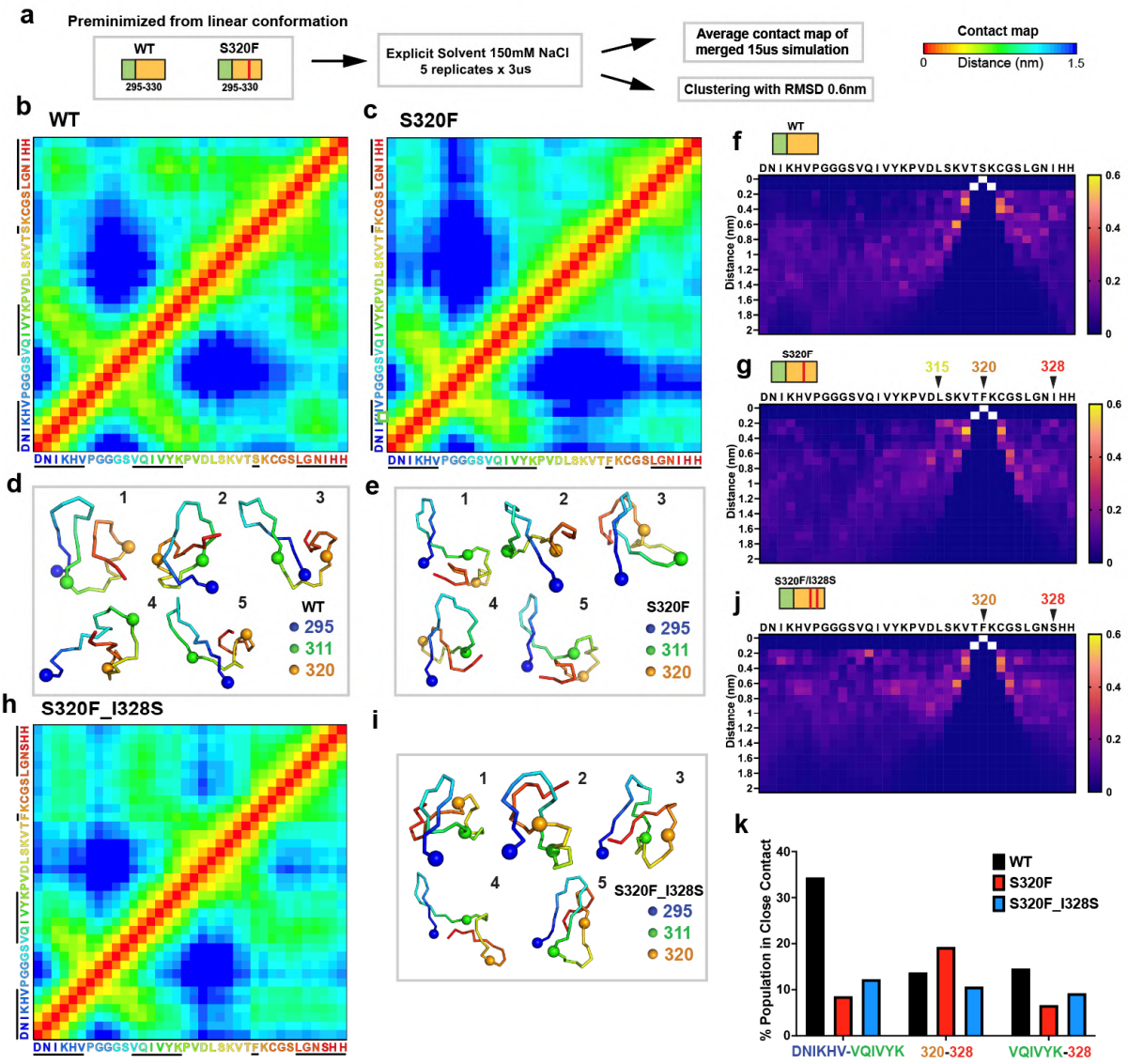


Fig. 3: MD simulations reveal different conformations for WT and S320F 295-330 fragment and I328 appears to be involved in the aggregation mechanism of S320F. a Simplified schematic of the MD simulation procedure. The color bar of the contact map indicates the distance between pairs of residues in the range of 0 – 1.5 nm. Average of five replicate simulations from independent trajectories of 3 μs is shown in a contact map for **b** WT, and **c** S320F. The middle structure of each of the top five clusters that are within 0.6 nm RMSD for simulation on **d** WT, and **e** S320F. Position 295, 311 and 320 are in blue-, green- and cantaloupe-colored spheres. Sequences are colored with the rainbow spectrum and match the residue coloring in **b** and **c**. Cα distance distribution of each residue to the S320 position in **f** WT and **g** S320F replicate average. The color scheme indicates the % population (scale 0-1) of the residue at a particular distance to 320. Arrows point to the positions of interest. h Average of five replicate simulations from independent trajectories of 3 μs for S320F/I328S

is shown in a contact map. **i** Top five clusters of 0.6 nm RMSD for simulation on S320F/I328S. **j** Cα distance distribution of each residue to the S320 position in S320F/I328S. **k** % population in close contact for three sets of contacts of interest in WT (black), S320F (red), and S320F/I328S (blue).

I328S suppresses the effect of S320F-based tau aggregation in a peptide system

Connecting the MD results with the fragment aggregation assay, our data supports that I328 interacts with S320F through nonpolar interactions and hypothesize that mutating I328 to a polar residue such as serine could disrupt this interaction. We tested this hypothesis directly *in vitro* in a ThT aggregation assay leveraging the fragment spanning 295-330 as well as tauRD and in cellular models of tau aggregation. We first performed aggregation experiments on S320F_I328S$_{295\text{-}330}$,S320F$_{295\text{-}330}$ and WT$_{295\text{-}330}$ tau fragments. We find that combining S320F with I328S slows down the aggregation kinetics over 10-fold yielding a $t_{1/2max}$ 11.75 ± 0.18 hrs for S320F/I328S$_{295\text{-}330}$, compared to a $t_{1/2max}$ 0.79 ± 0.27 hrs for S320F$_{295\text{-}330}$ alone while also reducing the ThT fluorescence amplitude (supplementary Fig. 4x). As a comparison, the S320F_I328S$_{295\text{-}330}$ mutant still aggregated faster than the WT control which yielded a $t_{1/2max}$ = 49.6 ± 2.6 hrs (Fig. 4a). From the simulations, we also identified L315 as another potential nonpolar residue that is proximal to S320F that could stabilize these putative local structures. Using our 295-330 peptide system, we tested whether L315S could also suppress the aggregation effect of the S320F mutation and as a control we tested L325S which is nonpolar but did not appear to be in close contact with 320F (Fig. 3c). We found that L315S_S320F$_{295\text{-}330}$ aggregated more slowly compared to S320F$_{295\text{-}330}$ but similarly to S320F_I328S$_{295\text{-}330}$, while S320F_L325S$_{295\text{-}330}$ had little effect (Supplementary Fig. 4a). These experiments support that nonpolar contacts with S320F observed in the MD

simulations play an important role stabilizing alternate local structures that allosterically expose ^{306}VQIVYK^{311} for aggregation.

Following up with I328S mutation, we hypothesized that the aggregation suppression role of I328S was specific to S320F due to the disruption of local interaction and not simply the removal of the hydrophobicity of this residue. Thus, we tested the effect of I328S on another disease-associated mutant, P301S, which is far from I328, is not predicted to associated with I328 and most importantly directly drives aggregation by exposing^{306}VQIVYK^{311} (Chen et al 2019). P301S_I328S$_{295\text{-}330}$ was found to aggregate to a comparable ThT magnitude with P301S$_{295\text{-}330}$ despite being moderately slower ($t_{1/2max}$ = 6.2 ± 0.3 hr for P301S$_{295\text{-}330}$ and $t_{1/2max}$ = 11.6 ± 0.5 hr for P301S/I328S$_{295\text{-}330}$, Supplementary Fig. 4b). Therefore, although hydrophobicity at 328 can still have a role in aggregation, the repair effect of I328S is likely specific to the S320F mutant and likely mediated by the disruption of the interaction between F320 and I328. Since hydrophobicity can be important for aggregation, next we tested whether substituting serine at a different position with phenylalanine will produce the same effect, so we designed S324F$_{295\text{-}330.}$ S324F accelerated the aggregation of 295-330 with $t_{1/2max}$ 33.9 ± 1.0 hrs compared to WT of $t_{1/2max}$ 49.6 ± 2.6 hrs, but much slower than the S320F counterpart of $t_{1/2max}$ 0.79 ± 0.27 hrs (supplementary Fig. 4c, conditions done in the same experiment as supplementary Fig. 4a). Although once again the hydrophobicity at the end of repeat 3 was demonstrated to be important for aggregation, this emphasized the crucial role of the 320 position in facilitating aggregation. The presence and absence of fibrils were confirmed by TEM at the endpoints of ThT experiments and were consistent with the ThT assay (Supplementary Fig. 4d).

I328S suppresses the effect of S320F-based aggregation using tauRD in vitro and in cells

We next wanted to confirm the suppression of S320F aggregation effect by the I328S mutation can be translated to tauRD. We produced tauRD S320F/I328S and tested the effect of the I328S mutation to suppress the aggregation of effects from S320F in a ThT fluorescence aggregation assay comparing against tauRD S320F. TauRD S320F again aggregated spontaneously with a $t_{1/2max}$ of 1.9 ± 0.17 hrs, while tauRD S320F/I328S showed a significantly reduced and delayed ThT signal while the WT tauRD remained flat (Fig. 4a). TEM imaging at the endpoints of the ThT experiments again confirmed formation of fibrils in tauRD S320F (Fig. 4b). A small amount of fibrils were also detected in the endpoint sample of tauRD S320F/I328S but none were found in the WT tauRD sample (Fig. 4b). In parallel, the spontaneous aggregation behavior of tauRD S320F and the repair effect from I328S were tested in the HEK293T cell model of tauRD aggregation leveraging the mEos3.2 system (Fig 1d). Using lentivirus expression we produced cell lines expressing tauRD S320F and S320F_I328S as well as controls WT and P301S tauRD as fusions to mEos3.2. The cells were photoconverted and the cells analyzed by flow cytometery. To interpret FRET in our cell populations, we employed a gating strategy to compare FRET in cells with similar mEos3.2 expression levels. As observed before, tauRD S320F-mEos3.2 aggregated spontaneously while the tauRD S320F/I328S displayed a greater than 50% drop in FRET. The cells expressing tauRD P301S and WT showed close to zero FRET. Representative images of live HEK293T cells expressing tauRD S320F, tauRD S320F/I328S and tauRD P301S were shown in Fig. 4d. Our data support that the I328S mutation can suppress aggregation activity of the S320F mutant in vitro and in cells. Importantly, the peptide

fragment experiments can be translated to tauRD suggestin that we can capture these local interactions in fragments but also tauRD.

To gain insight into the possible conformational changes between WT, S320F and S320F_I328S tauRD, we employed a crosslinking mass-spectrometry (XL-MS). In a previous study (Chen, 2019) we used XL-MS to capture changes in tauRD conformation caused by the P301S mutation and discovered that the disruption of the P-G-G-G alters the stability of the turn locally and globally. Here we extended this method to probe the differences in conformation between tauRD WT, S320F and the reversion mutant S320F_I328S. To detect possible changes in the conformation we used disuccinimidyl suberate (DSS) which reacts with primary amines (i.e. lysines). Samples of tauRD WT, S320F and S320F_I328S were reacted with DSS, quenched and the monomer species isolated from an SDS-PAGE gel. The samples were in-gel digested and processed through our XL-MS pipeline to identify crosslinked species that include crosslinks (i.e. two peptides linked by DSS) and looplinks (i.e. single peptide dually linked by DSS). In each dataset, the DSS modifications are reported as consensus contacts across five independent samples with a low false discovery rate (FDR) using a high score cut-off (>27) (Source Data 1). For the WT, S320F and S320F_I328S tauRD constructs we observe 47, 38 and 40 consensus crosslinks and looplinks, respectively (Fig. 4h). We find 33, 31 and 35 overlapping contacts between the WT-S320F, S320F-S320F_I328S and WT-S320F_I328S data sets, respectively. This translates to roughly to 63%, 66% and 67.5% overlap in contacts, respectively and indicates that the WT mutant has a more similar pattern to S320F_I328S than S320F. Comparison of the crosslink patterns uncovers that the S320F_I328S recovers more long range contacts and

is more similar to WT (Fig. 4f-h, box). Additionally, we find specific contacts within R3 in XL-MS (Fig. 4f-h, arrows) that report on similar contacts observed in MD that varied between $WT_{295-330}$, $S320F_{295-330}$ and $S320F\_I328S_{295-330}$ (Fig. 3k and Supplementary Fig. 3c). Specifically, the contacts between K311-K317 and K321-K331 report as a proxy on the spacing between K311-320 and 320-328, respectively. We find that K321-K311 is observed 12±1.2 in WT, drops to 2.7±0.6 in S320F and increases towards WT to 3.7±1.1 in S320F_I328S (Fig. 4f-h,i). For the other contacts, we find that the K311-K317 contact in WT is observed 28±8 times while in S320F it increases to 35±5 and in S320F_I328S it drops back down to 29±4 (Fig. 4f-h,i). Other contacts, including K331-K343 and K343-K353 also change but are further away from S320F and thus may have secondary effects to the ones observed closer to position 320. Because tau is intrinsically disordered and the interactions it samples are transient it is difficult to interpret these changes except that the WT and S320F_I328S contacts involving local elements are more similar and distinct from S320F and correlate with the aggregation behavior. Together, our data support that S320F contacts I328 to allosterically expose ^{306}VQIVYK^{311} and drive amyloid assembly and that this can be reversed by incorporation of I328S to reduce aggregation in vitro and in cells using tauRD as a model. And finally, we use XL-MS to probe changes in conformation and identify key contacts that track between WT and S320F_I328S and are distinct from those observed in S320F.

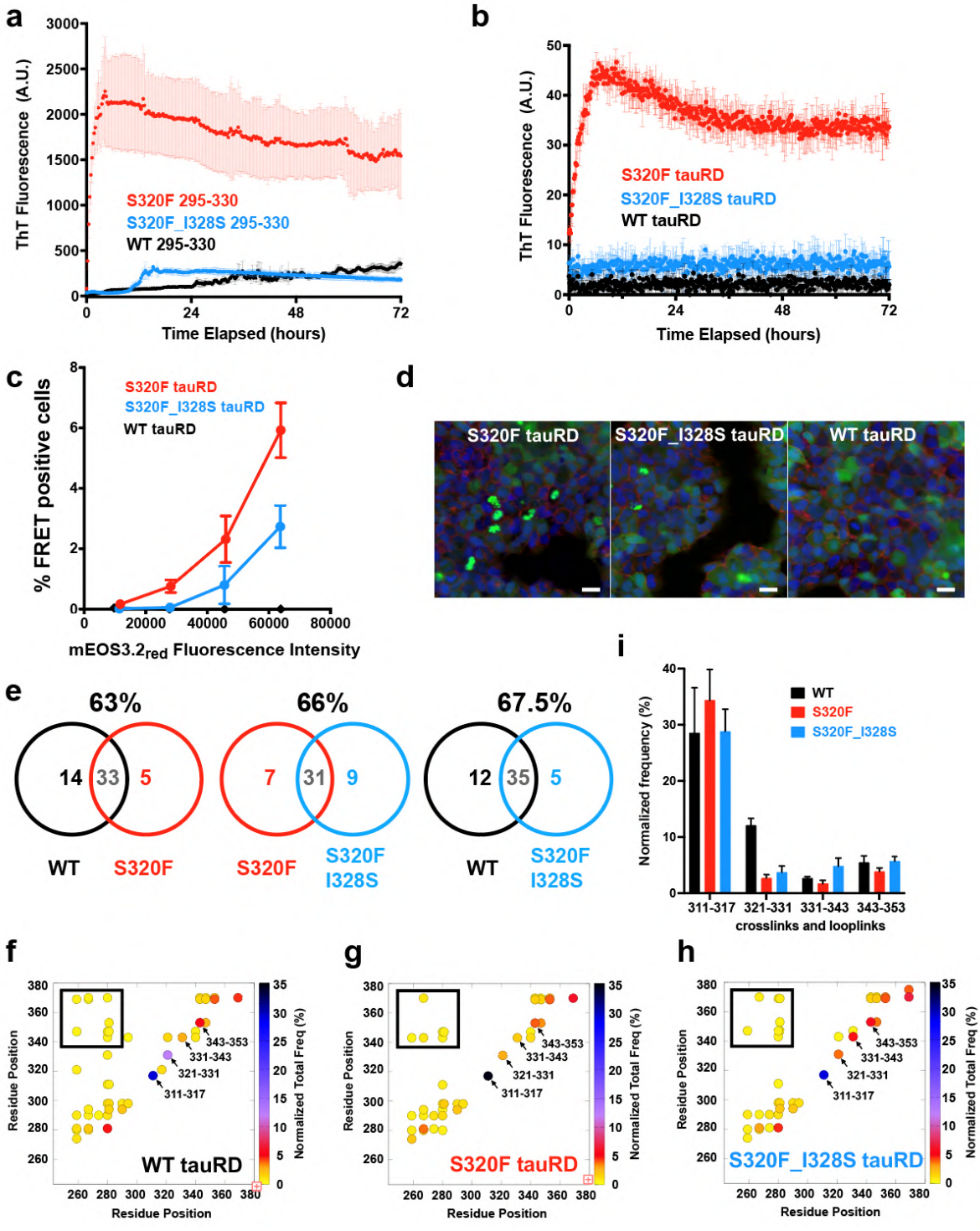

a
ThT Fluorescence (A.U.)
Time Elapsed (hours)
S320F 295-330
S320F_I328S 295-330
WT 295-330
b
ThT Fluorescence (A.U.)
Time Elapsed (hours)
S320F tauRD
S320F_I328S tauRD
WT tauRD
c
% FRET positive cells
mEOS3.2red Fluorescence Intensity
S320F tauRD
S320F_I328S tauRD
WT tauRD
d
S320F tauRD
S320F_I328S tauRD
WT tauRD
e
63%
14
33
5
WT
S320F
66%
7
31
9
S320F
S320F I328S
67.5%
12
35
5
WT
S320F I328S
i
Normalized frequency (%)
WT
S320F
S320F_I328S
311-317
321-331
331-343
343-353
crosslinks and looplinks
f
Residue Position
Normalized Total Freq (%)
343-353
331-343
321-331
311-317
WT tauRD
g
S320F tauRD
h
S320F_I328S tauRD

Fig. 4: I328S was able to partially repair the effect of S320F in vitro and in cells. a ThT fluorescence assay on $S320F_{295-330}$ (red), $S320F/I328S_{295-330}$ (blue) and $WT_{295-330}$ (black) at 200 μM, 37 °C. **b** ThT fluorescence assay on S320F tauRD (red), S320F/I328S (blue) and WT (black) at 25 uM, 37 °C. **c** HEK293T cells expressing S320F, S320F/I328S or WT tauRD-mEos3.2 were fixed at Day 4. FRET (tauRD-CFP/tauRD-mCherry) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition on S320F (red), P301S (green)), or WT (black). Comparison was conducted at multiple fluorescent intensity levels of $mEos3.2_{red}$. Error bars represent a 95% CI of each condition. **d** Representative images of S320F, S320F/I328S, and WT tauRD-mEos3.2 expressed HEK293T cells prior to photoconversion. mEos3.2 (green), Hoechst33342 (blue, nuclei stain), and Wheat Germ Agglutinin (red, cell membrane stains) fluorescence signals are shown in green and blue, respectively. Scale bar, 15 μm, is shown in white. **e** Number of unique DSS linkages (crosslinks and looplinks) in conditions of tauRD WT, S320F or S320F/I328S and their overlaps. **F-h** Consensus DSS linkages (circles) are shown in contact maps color-coded by summed frequency across replicates which is normalized by the total number of linkages in each condition. Boxed linkages are defined as long-range contacts. Arrows point to linkage pairs that show distinct differences across the three conditions. **i** Bar plots showing the arrow-pointed pairs from **f-h**. Error bars represent a 95% CI of each condition.

Computational design of nonpolar contacts produces spontaneously aggregating tau sequences

It remains unknown how to control tau folding into discrete fibril conformations. Our data suggest that the S320F allosterically changes how the amyloid motif is exposed thus leading to conformational changes that lead to spontaneous aggregation. To test this hypothesis more precisely we wanted to ascertain whether nonpolar substitutions at and in proximity to 320 may lead to a similar aggregation phenotype observed with S320F. FTD-tau driven by the S320F appears to be associated with both 3R and 4R tau isoforms (Rosso, 2002). Staining of FTD-tau S320F tissues with 4R antibodies shows pathology consistent with perinuclear inclusion similar to Pick bodies described in Picks disease or CBD (Fig 5a). Leveraging the available CBD and PiD cryo-EM structures we wondered whether a S320F

could be compatible with this fibril conformation (Falcon, 2018; Zhang, 2020). S320 in both the CBD and PiD conformations is placed in a cluster of nonpolar contacts that include L325 and I328 (Fig 5b and supplementary Fig 5a). We wondered whether we could use these structures to computationally optimize these local clusters and test whether such as nonpolar cluster may promote aggregation and stabilize fibrils. We employed the rosetta design module to computationally engineer the tau sequence using the CBD and PiD structures as templates. For the CBD fibril we optimized the 320 and 328 positions allowing sampling of all possible amino acid combinations yielding a matrix of 400 mutants. For PiD, we combined optimization of 325, 328 and 320 but restricted the search to nonpolar residues to reduce the number of possible substitutions. Our rosetta approach allowed us to identify low energy amino acid substitutions for both the CBD and PiD designs (Fig 5c and supplementary Fig. 5x). For CBD, two of the top scoring substitutions involved S320I and S320I with I328V (Fig. 5d, top and middle panels) yielding structures with low energies (Fig. 5c; S320I_I328, DREU=-17.7 and S320I_I320V, DREU=-19.6) and rmsds (Fig. 5c, of 0.631 Å for S320I_I328 and 0.633 Å for S320I_I320V), compared to the input conformation (Fig. 5d). Similarly, the minimized wildtype CBD structure retained a low rmsd (0.625 Å) relative to the input cryo-EM structure indicating that our method can maintain the correct fibril geometry with near native side-chain rotamers (Fig. 5d, bottom panel). We performed a similar mutagenesis calculation on the PiD structure and identified S320V as the most optimal solution with a low energy (DREU=-8.18) and rmsd (1.02 Å) relative to the PiD input structure (Supplementary Fig. 5b). Based on our hypothesis, we predict that stabilizing these nonpolar clusters should yield tau sequences that are similarly aggregation-prone

similarly to S320F. We tested this directly by producing tauRD fragments and measured their aggregation properties in vitro. What we find is that 25uM tauRD S320I and S320I_I328V aggregated readily in an in vitro assay with $t_{1/2max}$ values of 4.4 ± 0.9 hrs and 5.4 ± 2.0 hrs, respectively, and similar to S320F ($t_{1/2max}$ =3.0 ± 0.4 hrs) while the WT tauRD remained flat over the course of the experiment ($t_{1/2max}$ > 72hrs) (Fig. 5e). 25uM tauRD S320V aggregated with a $t_{1/2max}$ of 7.3 ± 1.2 hrs while WT tauRD remained flat ($t_{1/2max}$ > 72hrs) (Supplementary Fig. 5c). Repeating the aggregation experiment using 100uM reduces the $t_{1/2max}$ for all designed sequences and increases the amplitude of the ThT fluorescence (Supplementary Fig. 5d, S320F $t_{1/2max}$ =1.2 ± 0.2 hrs, S320I $t_{1/2max}$ =3.0 ± 0.8 hrs, S320I_I328V $t_{1/2max}$ =1.4 ± 0.6 hrs and S320V $t_{1/2max}$ =2.5 ± 0.7 hrs). For both the CBD and PiD designed tau sequences we were able to confirm the presence of fibrils using TEM (Fig. 5f and supplementary Fig. 5e). We finally tested the behavior of tauRD in our in cell spontaneous aggregation assay. We produced polyclonal HEK293 cell lines that express CBD-derived tauRD S320I and S320I_I328V and the PiD-derived tauRD S320V constructs as fusions to mE0S3.2. The cells were grown for 4 days, a population of the mEO3.2 was converted using UV and the percentage of the population with FRET signal was quantified. We find that in this experiment the WT and P301S tauRD-mEO3.2 constructs remained FRET negative even at high tauRD expression levels while the designed mutants increased as a function tauRD expression. The CBD- (Fig. 5g) and PiD-derived (Supplementary Fig. 5f) tauRD constructs revealed robust spontaneous aggregation in 75% as measured by FRET positivity while equivalent expression of FTD-tau S320F only approached 25% of cells with FRET positive signal. Our data suggest that S320F aggregation properties are driven by disruption of

protective nonpolar contacts and that our in silico approach yielded designed tau sequences that mimic S320F using other nonpolar residues that maintained the aggregation-prone phenotype.

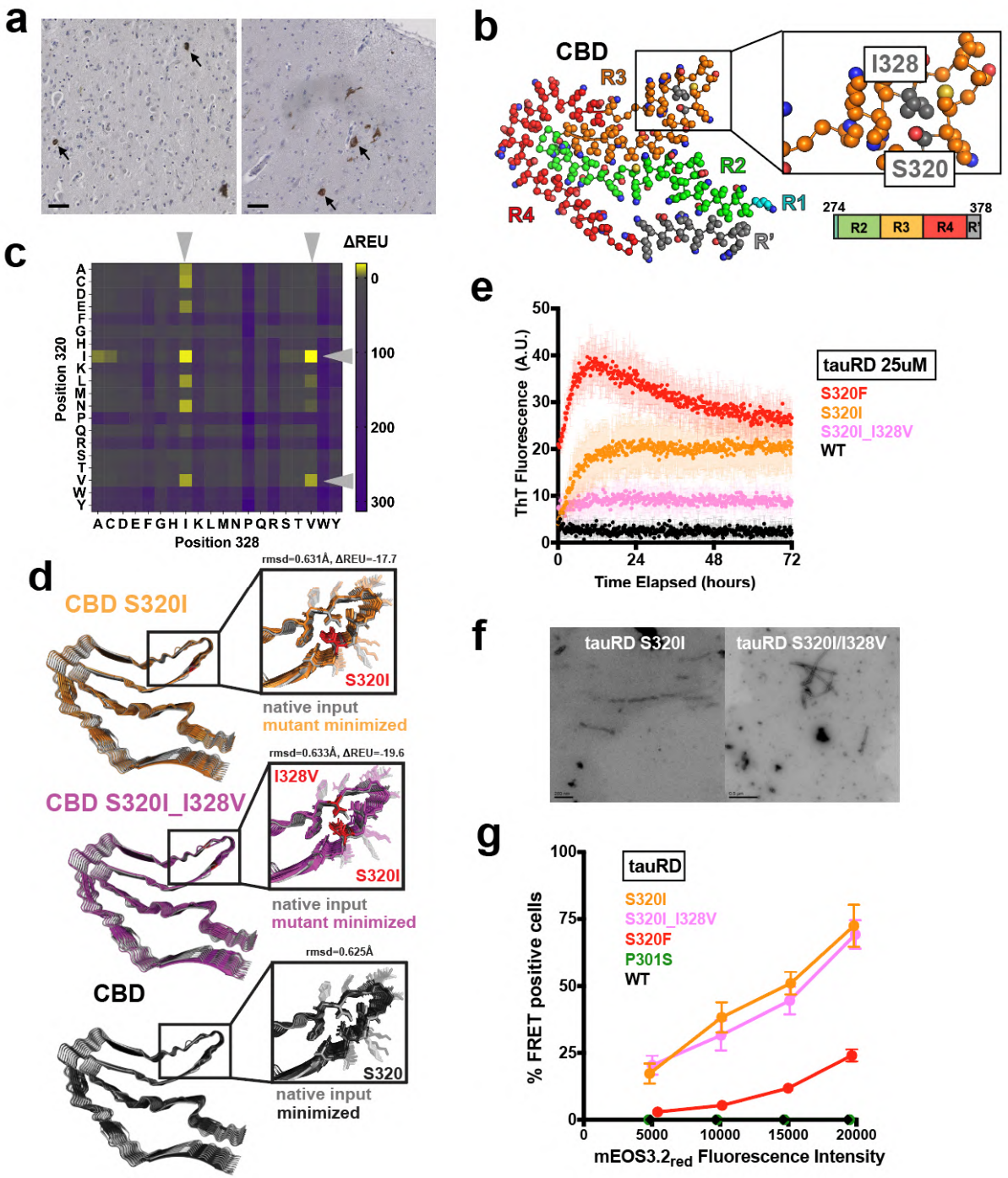


Fig. 5: Stabilizing local hydrophobic cluster promotes aggregation. a Immunohistochemistry staining of cingulum and hippocampus sections using 4R antibody. Arrows point to Pick-like bodies. Scale bar, 100 μm, is shown in black. **b** Top viewed, atomic model of the Cryo-EM resolved tau fibril from CBD brains. Residues are colored according to the color scheme of the tauRD schematic. A zoomed in view of the local region encompassing S320 and I328 is shown, where S320 and I328 are colored in gray. **c** ΔREU (REU = Rosetta Energy Unit, REU_{mutant} - REU_{WT}) map for relaxed structures with residues mutated at position 320 and 328 in the context of CBD tau fibril. Color bar indicates the value of ΔREU where negative ΔREU will be in yellow and positive ΔREU will be in purple. Gray arrow heads point to mutations of interest. **d** Representative models of an energy-minimized S320I 9-mer (orange), S320I/I328 9-mer (pink) and WT 9-mer overlaid with the WT native input in the context of CBD tau fibril structure. **e** ThT fluorescence assay on tauRD S320F (red), S320I (orange), S320I/I328V (pink), and WT (black) at 25 μM, 37 °C. **f** Representative TEM images of the endpoints of ThT assays on tauRD S320I and S320I/I328V at 25 uM. **g** FRET (tauRD-CFP/tauRD-mCherry) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition of S320F (red), S320I (orange), S320I/I328V (pink), P301S (green), or WT (black). Comparison was conducted at multiple fluorescent intensity levels of $mEos3.2_{red}$. Error bars represent a 95% CI of each condition.

Discussion

In this study, we elucidate the molecular mechanism of how S320F, a FTDP-17 *MAPT* mutation that drives spontaneous aggregation of tau in vitro and in cells. Leveraging a combination of in silico, in vitro and cell model experiments, we unveiled the local structural change that leads to the spontaneous aggregation behavior of S320F tau. Building upon our previous work (Chen, 2019), we propose that homologous sequence motifs preceding ^{301}PGGG^{304} and ^{332}PGGG^{335} in repeat 2 (^{295}DNIKHV^{300}) and repeat 3 (^{325}LGNIHH^{330}) shield the amyloid motif ^{306}VQIVYK^{311} in a synergistic manner (Fig. 6). The introduction of S320F mutation attracts the sequence motif preceding ^{332}PGGG^{335} in repeat 3 to form stabilized local hydrophobic cluster and that allosterically disrupts the protection of ^{306}VQIVYK^{311}, the

exposure of which leads to self-assembly of tau (Fig. 6). The understanding on the structural basis of how S320F mutation causes tau to aggregate enabled us to design sequences beyond disease-associated mutations that allow tau to aggregate. The engineered mutants of tau with one or two amino acid substitutes suggests the importance of local structure stabilized by hydrophobic clustering in facilitating tau aggregation.

These hydrophobic clusters are seen in the monomer within the tau fibril structures determined by cryo-EM, such as residues from ^{306}VQIVYK^{311} (V306, I308, and Y310) form contacts with ^{373}K**L**T**F**RE^{378} in AD and CTE, or with ^{337}**V**E**V**KSE^{342} in PiD and CBD (Fitzpatrick, 2017; Falcon, 2019; Falcon, 2018; Zhang, 2020; Vaquer, 2021). These interactions might be important as the driving force of tau templating into specific shapes (Vaquer, 2021). A group using molecular dynamics simulations integrated with experimental data found that a fraction of the tau RD ensemble presents local structures matching that of the CBD tau fibril (Stelzl, 2022). This implies the possibility of a sub population of monomer to adopt local structures compatible with the fibrillar end product which might therefore facilitate intermolecular templating. To validate if local structure that drives tau aggregation is preserved in the end fibril structures, cryo-EM can be conducted on S320F tau fibrils as a next step. If the connection is built, a systematic design of diagnostics or therapeutics for various types of tauopathies based on solved tau fibril structures will be possible.

We do not think that loss of a phosphorylation site due to the S320F mutation underlies its aggregation mechanism since there is no evidence that S320 is phosphorylated in tau from either normal or disease tau. However, S320F was shown to reduce the ability of

recombinant tau to promote microtubule assembly (Rosso, 2002). This property together with the aggregation enhancement potentially underlie the cause of disease.

A homologous site of S320 in repeat 4, S352 has been discovered that if mutated to Leucine (S352L) also causes tauopathy (Nicholl et al., 2003). Repeat 4 also harbors the homologous of ^{306}VQIVYK^{311} amyloid motif, ^{337}VEVKSEK^{342}, at the beginning of the repeat. S352L mutation is also a polar to hydrophobic mutation, which possibly facilitate tau aggregation using a similar mechanism as S320F. The fact that the homologous position of 320 in a different repeat is found to encode a disease-causing mutation and physical similarity between S352L and S320F imply the importance of these homologous positions in tau fibril assembly.

Evidence from other studies demonstrated the importance of how local structures in tau regulate its property. Local conformational changes in intrinsically disordered protein tau may not transform its global conformation but might be sufficient to alter the behavior of tau, such as in aggregation (Chen, 2019) or microtubule-associated function (Cario, 2021). It was characterized that the disease associated R5L mutation of tau induced local structural change which affected the formation of "tau patches" and hence altered the behavior of tau on microtubules (Cario, 2021). Through NMR, it was demonstrated that the mutation which perturbs only 15 residues of N-terminal affected up to 50 residues in the presence of microtubule (Cario, 2021). This suggests mutation in tau might only affect the structure locally but can trigger more prominent changes with regard to its biochemical property.

This study together with our previous work (Chen, 2019) have emphasized the importance of exposure of amyloid motif ^{306}VQIVYK^{311} in aggregation and strengthened the

therapeutic significance of shielding the amyloid motif to mitigate aggregation. These insights suggest the possibility of using the exposed states of amyloid motif as molecular targets for diagnostic and therapeutic development.

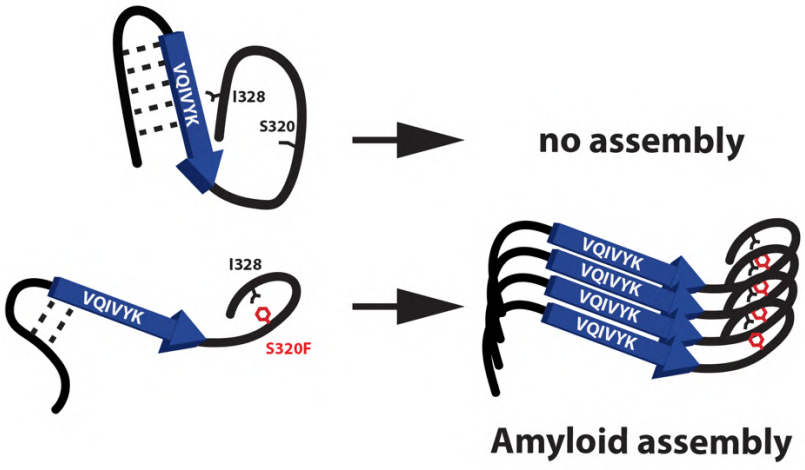
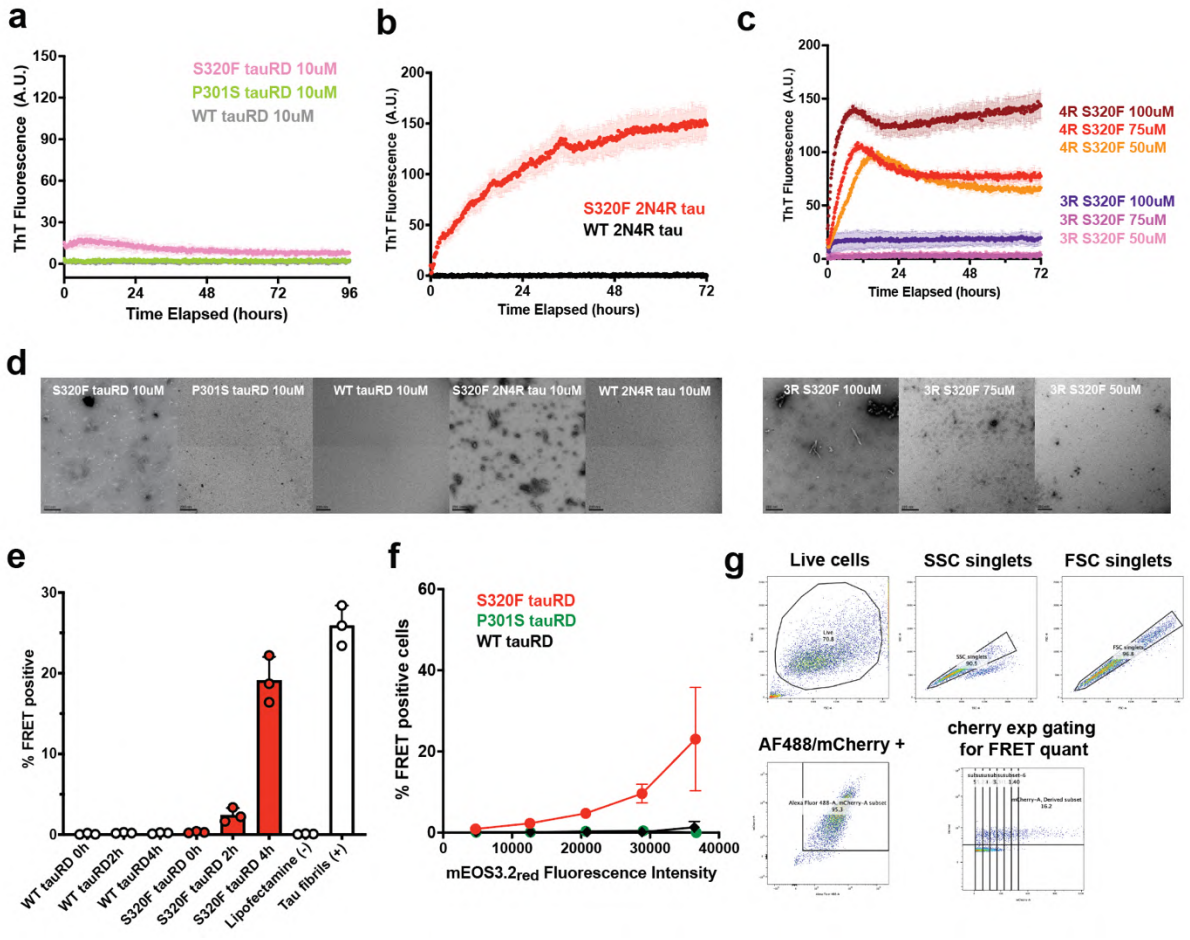


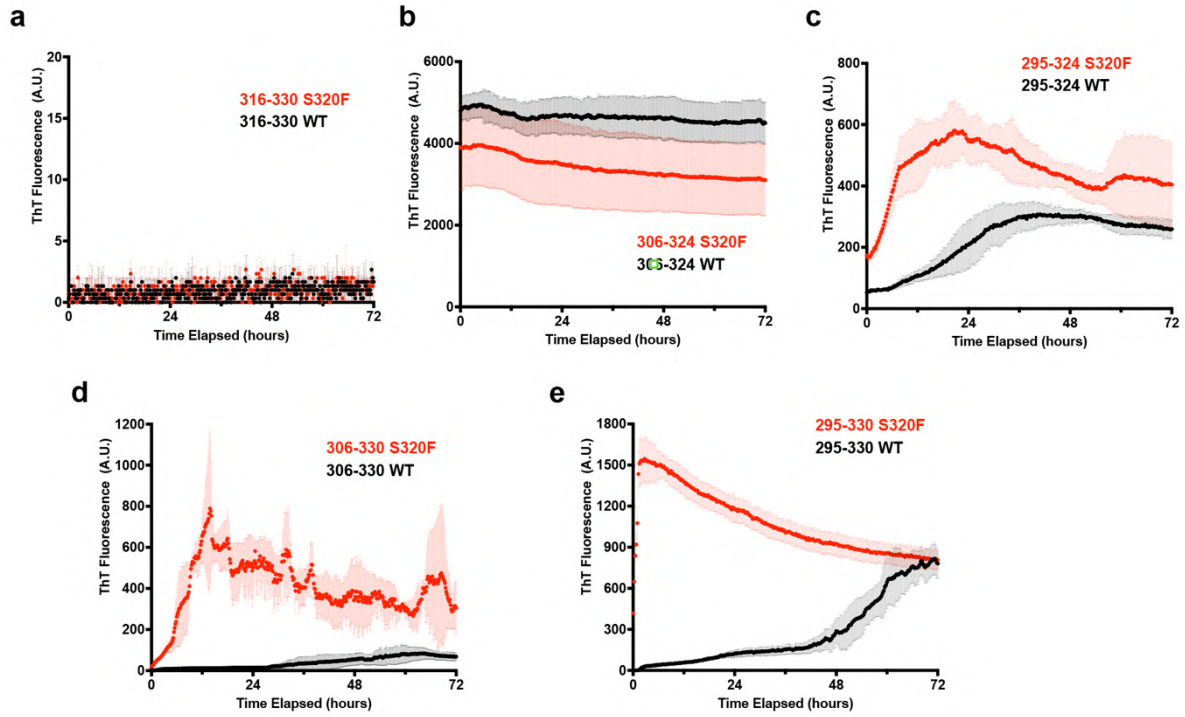
Fig. 6. Molecular model of the structural differences between WT and S320F within region 295-330 of tau and their subsequent consequences. WT (top) forms a relatively collapsed structure surrounding the amyloid motif ^{306}VQIVYK^{311}, where the preceding sequence motifs at the end of repeat 2 and repeat 3 shield ^{306}VQIVYK^{311} synergistically. In S320F mutant, the sequence motif at the end of repeat 3 is pulled back to instead form local hydrophobic cluster with S320F. The loss of the interaction from the sequence motif of repeat 3 also destabilizes the interaction of the sequence motif at the end of repeat 2 with ^{306}VQIVYK^{311}. This allows partial or higher frequency of the exposure of ^{306}VQIVYK^{311} amyloid motif and thus enhances aggregation propensity, leading to amyloid formation.

Supplemental Figures

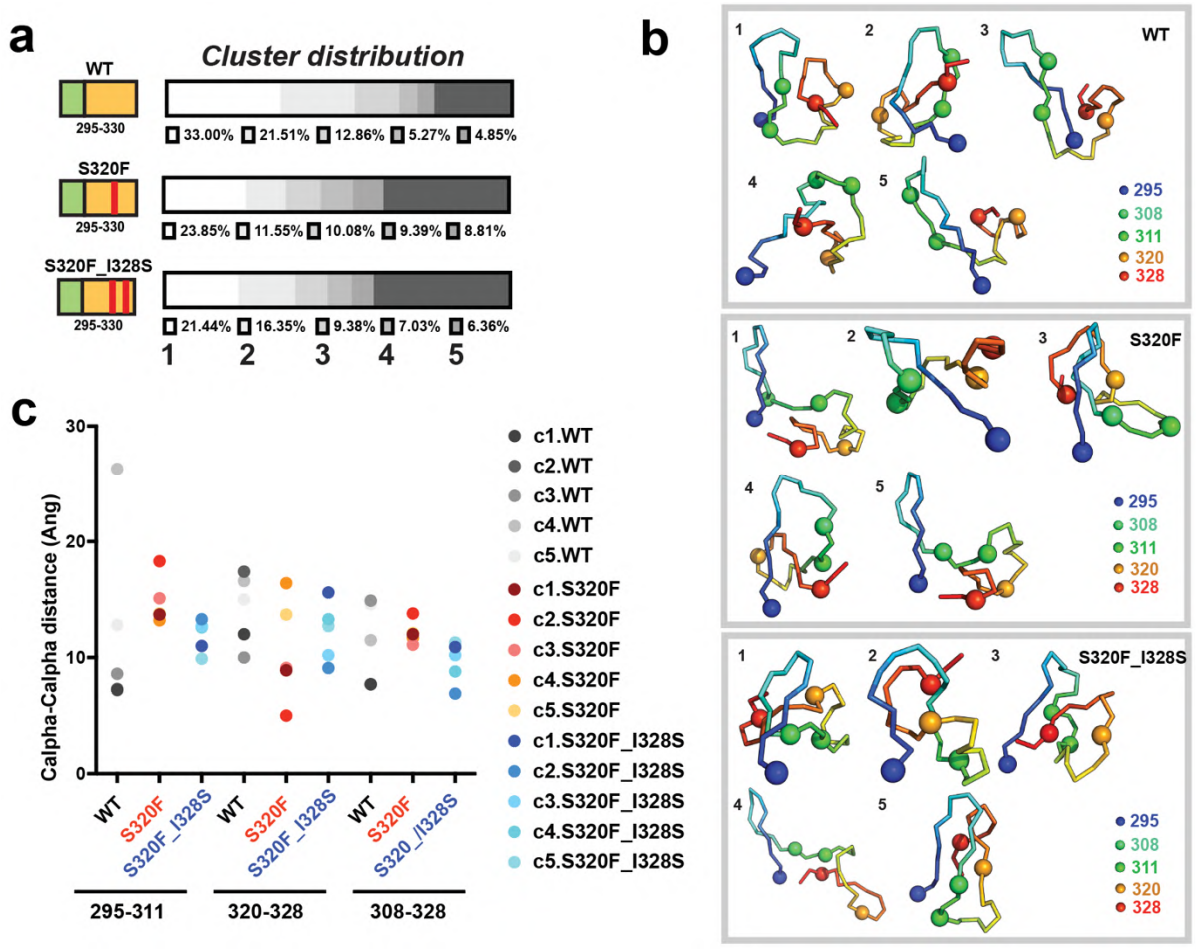


Supplementary Fig. 1: additional in vitro and in cell assays on tauRD, 2N4R tau and tauRD (3R). a ThT fluorescence assay on tauRD S320F (red and magenta), P301S (green and olive) and WT (black and gray) at 62.5 uM and 25 µM, 37 °C. **b** ThT fluorescence assay of full-length 2N4R tau S320F (red) and WT (black). **c** ThT fluorescence assay on tauRD (4R) and tauRD (3R) S320F. Light to dark colors indicate increasing concentrations. **d** TEM images of the endpoint ThT assay on the above mentioned conditions. The black bars represent 200 nm distance. **e** tauRD WT and S320F incubated at 2h, 4h or with no incubation were transduced in parallel with lipofectamine (negative control) and tau fibrils (positive control) into the tau biosensor cells. FRET signal indicating amount of tau aggregates 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. **f** HEK293T cells expressing S320F, P301S or WT tauRD-mEos3.2 were fixed at Day 7. FRET (tauRD-CFP/tauRD-mCherry) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition on S320F (red), P301S (green)), or WT (black). Comparison was conducted at multiple fluorescent intensity levels

of mEos3.2$_{red}$. Error bars represent a 95% CI of each condition. **g** Flow Jo procedures to extract live, single, Alexa-Fluor 488 and mCherry double-positive cells followed by expression level based gatings for FRET quantifications.

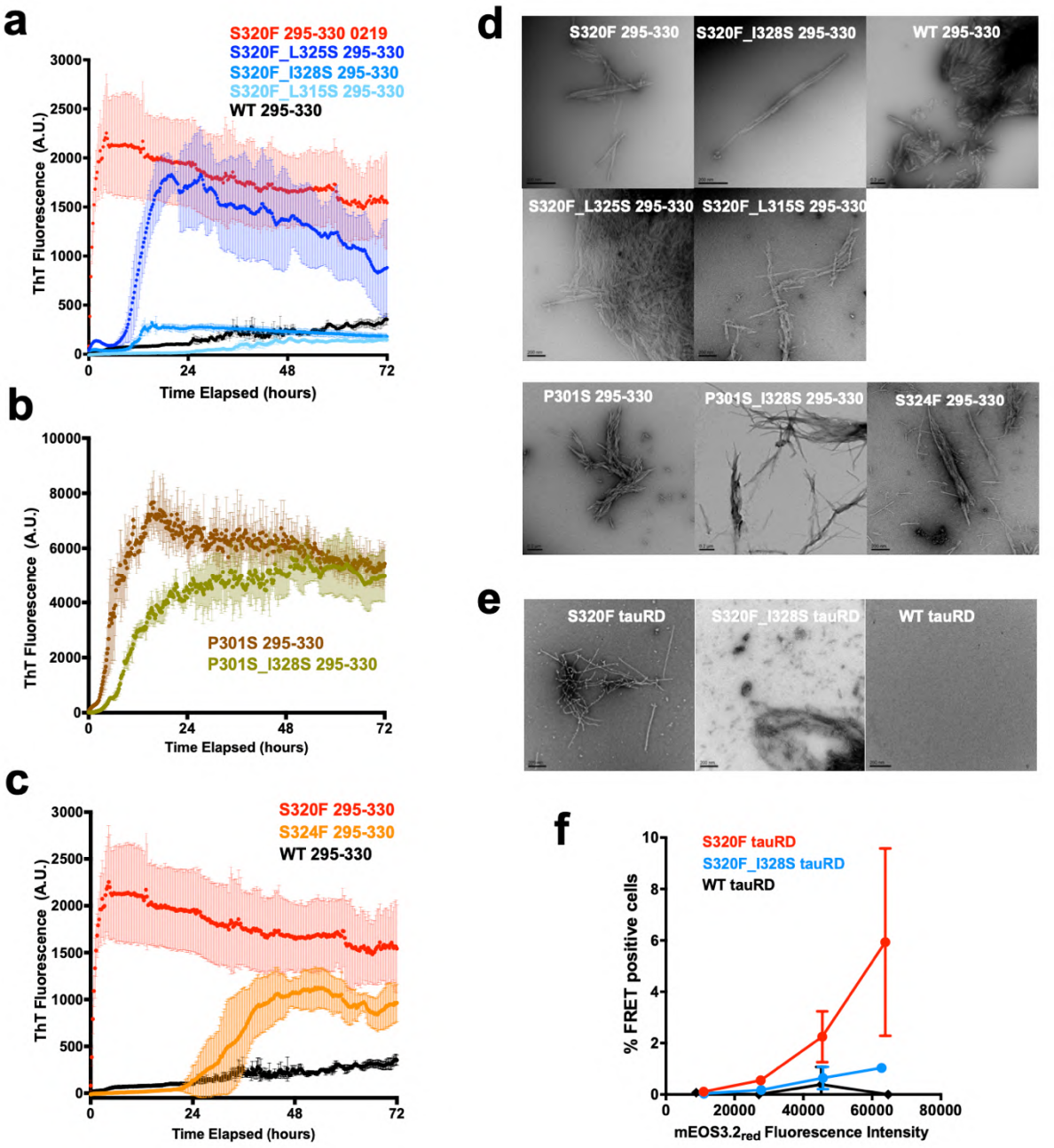


Supplementary Fig. 2: VQIVYK is essential in S320F facilitated aggregation and the surrounding sequence regulates aggregation. Raw ThT curves comparing WT and S320F for **a** 316-330, b 306-324, **c** 295-324, **d** 306-330, and **e** 295-330.



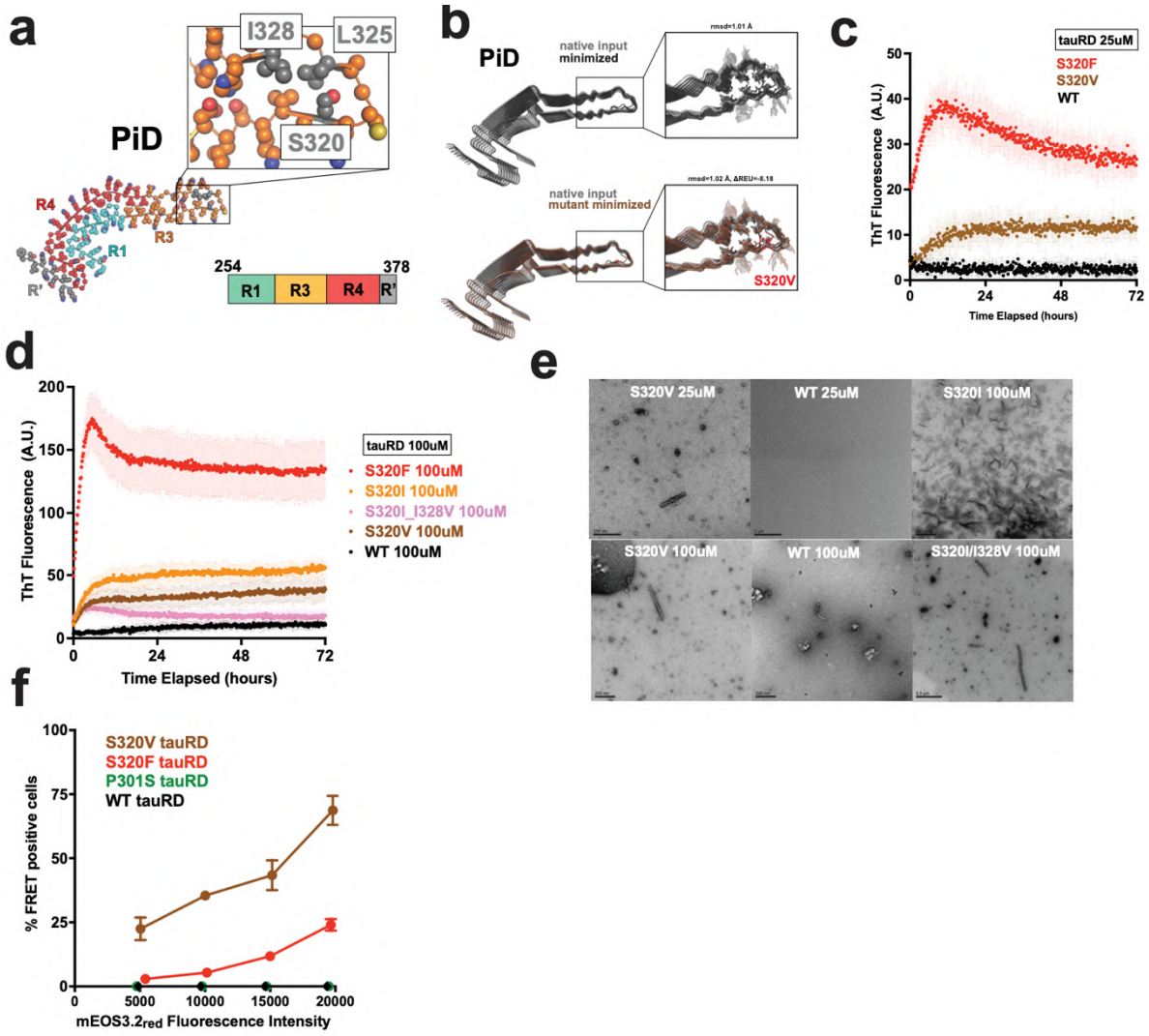
Supplementary Fig. 3: Supporting data for MD simulations on WT, S320F and S320F/I328S 295-330 fragment.

a cluster distribution of MD simulations on $WT_{295-330}$, $S320F_{295-330}$, and $S320F/I328S_{295-330}$. The percentages for the top five clusters are shown. **b** The middle structure of each of the top five clusters that are within 0.6 nm RMSD for simulation on WT, S320F, and S320F/I328S. In addition to 295, 311 and 320 that are marked in the main figure, 308 and 328 are shown in lime and red spheres. **c** Cα distance measured of the pairs of interest for the middle structures of the top five clusters. c1 = cluster 1, etc.



Supplementary Fig. 4: Aggregation assays of additional mutant peptide controls and cell model assay at Day 7. a ThT fluorescence assay on $S320F_{295-330}$ (red), $S320F/L325S_{295-330}$ (dark blue), $S320F/I328S_{295-330}$ (blue), $S320F/L315S_{295-330}$, and $WT_{295-330}$ (black) at 200 µM, 37 °C. **b** ThT fluorescence assay on $P301S_{295-330}$ (brown) and $P301S/I328S_{295-330}$ (olive) at 200 µM, 37 °C. **c** ThT fluorescence assay on $S320F_{295-330}$ (red), $S324F_{295-330}$ (orrange) and $WT_{295-330}$ (black) at 200 µM, 37 °C. The S324F condition was tested in the same experiment as in **a** and thus S320F and WT are used again here. **d** Representative TEM images of the endpoint ThT assay on 295-330 sequence fragments from **a**, **b** and **c**. **e** Representative

TEM images of the endpoint ThT assay on tauRD S320F, S320F/I328S and WT. **f** HEK293T cells expressing S320F, S320F/I328S or WT tauRD-mEos3.2 were fixed at Day 4. FRET (tauRD-CFP/tauRD-mCherry) was measured by flow cytometry on three biological triplicates of at least 10,000 cells per condition on S320F (red), P301S (green)), or WT (black). Comparison was conducted at multiple fluorescent intensity levels of $mEos3.2_{red}$. Error bars represent a 95% CI of each condition.



Supplementary Fig. 5: Aggregation assay on stabilizing mutation predicted from the PiD tau fibril structure and additional data on stabilizing mutations predicted from the CBD tau fibril structure. a Top viewed, atomic model of the Cryo-EM resolved tau fibril from PiD brains. Residues are colored according to the color scheme of the tauRD schematic. A zoomed in view of the local region encompassing S320, L325, and I328 is shown, where S320, L325, and I328 are colored in gray. **b** Representative models of an energy-minimized S320V 9-mer (brown), WT 9-mer overlaid with the WT native input in the context of CBD tau fibril structure. **c** ThT fluorescence assay on tauRD S320F (red), S320V (brown), and

WT (black) at 25 μM, 37 °C. The S320V condition was tested in the same experiment as in **Fig. 5e** and thus S320F and WT are used again here. **d** ThT fluorescence assay on tauRD S320F (red), S320I (orange), S320I/I328V (pink), S320V (brown), and WT (black) at 100 μM, 37 °C. **e** Representative TEM images of the endpoint ThT assay on tauRD S320V and WT at 25 uM, S320I, S320V, S320F/I328S and WT at 100 uM.

Materials and 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 S320F mutant of full-length tau in pet28b backbone was produced by Twist Bioscience. 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 sopropyl 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 Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4).

Plasmids of tau-RD WT and the mutant forms, S320F and S320F+I328S in pet28b backbone but without the C-terminal His-tag were produced by Twist Bioscience. Tau RD

wildtype and mutants were expressed the same way as full-length tau. Purification procedures for tau-RD were adopted from Dr. Paul Seidler (USC). The cells were harvested and lysed in 20mM 2-(*N*-morpholino) ethanesulfonic acid (MES), 1mM (Ethylenediaminetetraacetic acid) EDTA, 1mM $MgCl_2$, 5mM β-mercaptoethanol, pH6.8, and appropriate amounts of cOmplete™ EDTA-free Protease Inhibitor Cocktail tablets (Sigma), using an Omni Sonic Ruptor 400 at 4 °C. The lysates were centrifuged, and the supernatant was boiled in a flask with 500 mM NaCl for 20 min in a water bath. The boiled supernatant was centrifuged at 15,000 xg for 15 min. The supernatant after centrifugation was dialyzed against a 20-fold volume of 20mM MES, 50 mM NaCl, 5mM β-mercaptoethanol, pH6.8. The dialysis buffer was changed once after 4 hours and left overnight. The dialyzed lysate was filtered using a 0.22 μm filter and loaded on a 5 ml HiTrap SP HP (GE) and eluted with a 50 mM–800 mM NaCl gradient in 20mM MES, 50 mM NaCl, 5mM β-mercaptoethanol, pH6.8. Tau RD containing fractions were concentrated on an Amicon-15 concentrator (EMD Millipore) 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 Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Aliquots were all stored at − 80 °C in 1 × PBS.

Peptide synthesis

All sequence fragments 316-330, 306-324, 306-330, and 295-330 were synthesized as ordered by Genscript with N-terminal acetylation and C-terminal amidation modifications and > 95% purity.

ThT fluorescence aggregation assays

WT or mutant FL tau and tau RD protein was diluted in 1 × PBS and filtered with 0.22 μm centrifuge filters before use. A final concentration of 25 μM protein, 2 mM TCEP, and 25 μM ThT were mixed and aliquoted into a 384-well clear bottom plate, each well with 55 uL volume. Peptides were disaggregated as previously described (O'Nuallain, 2006). In brief, lyophilized peptides were dissolved in 200 uL 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 or CO_2 gas, and then immediately placed under lyophilizer to remove any residual volatile solvents. The peptide residue was resuspended in 1 × PBS and 2 mM TCEP to a 200 μM concentration to adjust the peptide to buffered reaction conditions and the sample was adjusted to pH 7 with NaOH. ThT was added to the samples at a final concentration of 25 μM at the end. 55 uL of master mix was added in triplicates in a 384-well clear bottom plate. All conditions were done in triplicates at RT. ThT kinetic scans were run every 10 min on a Tecan Spark plate reader at 446 nm Ex (5 nm bandwidth), 482 nm Em (5 nm bandwidth). Values of the blank wells containing buffer and ThT were subtracted from values of the experimental groups.

Transmission electron microscopy

An aliquot of 5 μL of sample was placed onto a glow-discharged Carbon 300-mesh copper grids for 2 min, washed with distilled water for 30 s, and then negatively stained with 2% uranyl acetate for 2 min. Images were acquired on a Tecnai G^2 spirit transmission

electron microscope (FEI, Hillsboro, OR), serial number: D1067, equipped with a LaB_6 source at 120 kV using a Gatan ultrascan CCD camera.

Seeding assay on tau RD biosensor cells

Stable HEK293T cell line expressing tau RD P301S – CFP and tau RD P301S – YFP was plated at a density of 35,000 cells per well in a 96-well plate 18 – 24 hr before treatment. 25 μM of tau RD WT and S320F after incubation at 37 °C for various time periods were transduced by Lipofectamin 2000. Specifically, for each well, 10uL of each condition and 10uL of transduction reagent [9.5 μL Opti-MEM (Gibco) +0.5 μL Lipofectamine 2000 (Invitrogen)] were mixed and incubated for 20 min at room temperature. All conditions were done in triplicates with 20 μL total volume treated per well. After 48 hr incubation, cells were harvested with 0.05% trypsin then fixed in 2% paraformaldehyde (Electron Microscopy Services) for 10 min, after which resuspended in 1 x PBS for flow cytometry analysis.

Cell expression

FM5-CMV constructs of tauRD WT, P301S, S320F, and S320F/I328S fused with mEOS3.2 at the C-terminal were used for cell expression. Virus of each construct was produced in Lenti-X™ 293T Cell Line. Specifically, 400ng PSP, 1200ng VSVG, 400ng of plasmid of interest. 7.5uL TransIT and Opti-MEM were mixed to a final volume of 150 μL and incubated at room temperature for 30 min before adding to the Lenti-X™ 293T Cell Line (Takara, Cat. #: 632180). Media (10% FBS, 1% Pen/Strep, 1% GlutaMax in Dulbecco's modified Eagle's medium) containing virus was collected after 48 hrs and concentrated by

50-fold following the protocol of Lenti-X™ Concentrator (Takara, Cat. #: 631232). The concentrated virus in 40uL volume was treated to the HEK293T cells plated several hours ahead started with 80,000 cells in one well of the 24-well plate. Fluorescence expression was check 24 hrs later and cells were passed to one well of a 6-well plate once reach ~80% - 90% confluency. Cells were harvested on Day 7 after virus treatment with 0.05% trypsin then fixed in 2% paraformaldehyde (Electron Microscopy Services) for 10 min, after which resuspended in 1 x PBS. Immediately before flow cytometry, cells was photoconverted under UV for 25 min, according to empirical evidence on optimal ratio of green and red fluorescence for FRET measurement (Khan, 2018).

In cell imaging on tauRD cell lines

A split of the HEK293T cells were treated with virus expressing either tau RD S320F, S320F/I328S or P301S were plated at 20,000 cells per well in media (10% FBS, 1% Pen/Strep, 1% GlutaMax in Dulbecco's modified Eagle's medium) in a 96-well clear bottom plate (Corning, Product # 3603). After 24 h, cells were stained with Hoechst33342 and Wheat Germ Agglutinin at a final concentration of 5 μg/mL in media for 10 min at 37 °C, protected from light. The staining solution was removed and substituted with 1x PBS afterwards. The plate was placed on an IN Cell 6000 Analyzer (GE Healthcare) with a heated stage and 50 fields of view were imaged under 4′,6-diamidino-2-phenylindole (DAPI) and FITC channels at ×60 magnification (Nikon ×60/0.95, Plan Apo, Corr Collar 0.11–0.23, CFI/60 lambda). Images were exported as TIFF files for downstream analysis.

Flow cytometry

A BD-LSR Fortessa instrument was used to perform FRET flow cytometry. For seeding assay, CFP, YFP and FRET were measured. 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 MACSQuantify Software from Miltenyi Biotec. Because some YFP-only cells exhibit emission in the FRET channel, we introduced and 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 triplicates. Data analysis was performed using FlowJo v10 software (Treestar). For the mEOS3.2 cell expression system, tauRD-mEOS biosensor cells were first photoconverted under UV for 30 min. 10,000 singlet events corresponding to donor (non-photoconverted mEOS3.2) and acceptor (photoconverted mEOS3.2) positive cells were collected for each sample. Collection parameters are detailed in Supplemental Table 1. FCS files were exported

from the FACSDiva data collection software and analyzed using FlowJo v10 software (Treestar). Compensation was manually applied to correct for donor bleed-through into the FRET channel guided by a sample with non-aggregated and photoconverted tauRD-mEOS. Samples were gated on the acceptor intensity such that cells with similar concentrations of tauRD-mEOS were analyzed to mitigate the contribution of differences in concentration leading to apparent changes in the fraction of FRET positive cells in each condition. FRET positive cells were quantified by gating double-positive singlet events with a ratio of FRET to donor signal higher than that of a population of tauRD-mEOS photoconverted cells without aggregates.

XL-MS sample preparation and mass spectrometry

Preparation of tauRD was diluted in 1 × PBS with 1 mM DTT and cross-linked at a total protein concentration of 60 μM using 300 μg of starting material for WT, S320F and S320F_I328S tauRD. The cross-linking reaction was initiated by adding DSS stock solution (200 mM DSS-d_0 and -d_{12}, Creative Molecules, dissolved in DMF) to a final concentration of 10 mM. Samples were incubated at 37 °C for 30 s with 350 RPM shaking. The crosslinking reactions were quenched by addition of ammonium bicarbonate to 100 mM final concentration and incubation at 37 °C 350 RPM for 30 min. Aliquots of the crosslinked samples were resolved on an SDS-PAGE gel and the monomer band was excised and digested in the gel. In detail, gel pieces were sliced into 1 mm^2 cubes and sonicated in 25mM NH_4HCO_3/50% ACN (acetonitrile) for 5 min twice and each time the supernatant was discarded. Gel pieces were then washed with 100% ACN with occasional vortex for 5min

until shrank and became white. Gel pieces were subsequently evaporated to dryness by lyophilization. Gel pieces were then incubated in 25 mM NH_4HCO_3 (extra volume to ensure gel pieces were still covered after expanding) with 10 mM DTT at 56°C for 40 min. After removal of supernatant, gel pieces were incubated in 25 mM NH_4HCO_3 with 55 mM iodoacetamide and sat in dark for 30 min at room temperature. Gel pieces were then washed with 25 mM NH_4HCO_3, 25 mM NH_4HCO_3 in 50% ACN, and lastly 100% ACN, each step with intermittent vortexing and sitting for 5 min. Gel pieces were again evaporated to dryness by lyophilization. Total of 60 ug of trypsin (mass spectrometry grade, NEB) was diluted in 25 mM NH_4HCO_3 and added to dried gel pieces (3x volume of the gel volume). In gel proteolysis were carried out at 37°C 600 RPM overnight. Proteolysis solution was saved and gel pieces were incubated in 5% (v/v) formic acid diluted in H_2O at 37°C for 10 min followed by 5 min sonication twice. 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/ACN/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 C_{18} 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 S320F or S320F_I328S 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 > 27, 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. 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. The

nseen (frequency) of each residue position modified by cross-link or loop-link was summed and normalized to the total number of cross-link and loop-link modifications across all residues in each condition. The contact maps were plotted using an in-house Gnuplot script where the color scheme indicates the normalized total frequency in %. The normalized total frequency in % for the four interesting DSS modified pairs was plotted against the residue positions in Prism with error bars representing a 95% CI.

Molecular dynamics simulations

Two separate molecular dynamics simulations were generated that differentiated with setup preminimization and final simulations. Preminimization simulations were performed to obtain energetically minimized conformations of initial linear peptide conformations were used as starting structures for final simulations. Both simulations were performed for three peptides: $WT_{295\text{-}330}$, $S320F_{295\text{-}330}$, $S320F\_I328S_{295\text{-}330}$. In the case of preminimization simulations, the systems were built based on a fully extended conformations of peptides constructed in Pymol. The AMBER99sb-ildn force-field (Lindorff-Larsen, 2010) and TIP3P water model (Jorgensen, 1983) were used for all simulations. To prepare the system a dodecahedron box (constructed with a minimum 1 nm boundary condition – including about 30 000 of water molecules) with inclusion of NaCl (150 mM physiological ionic strength). To perform energy minimization of constructed set up the steepest decent algorithm was used to obtain maximum force below 1000.0 kJ/mol/nm. Then 10 ns of NVT and 20 ns of NPT (first 10 ns with Berendsen coupling (Berendsen, 1984) and the last 10 ns with Parrinello-Rahman pressure coupling (Parrinello, 1981)) equilibrations were performed. The subsequent

production level trajectories are based on 2 fs time steps (Feenstra, 1999). Production level trajectories were obtained for an NPT ensemble with Parrinello-Rahman barostat, and periodic boundary conditions with Particle Mesh Ewald (PME) (Darden, 1993) summation for long-range electrostatics. A total of 1.5 μs trajectories were generated through the production of five 100 ns simulations per peptide. These trajectories were analyzed using the Gromacs rms function – based on the results extended but energetically minimized conformations were extracted. Final simulations were performed with the same minimization, equilibration and production workflow with only a few modifications. The size of the box was decreased – because of the differences in compactness of the starting structures, size of the boxes was normalized to include 10 000 water molecules. A total of 45 μs trajectories were generated through the production of five 3 μs trajectories per peptide. All simulations were performed on UTSW's bioHPC computing cluster.

Computational Structural Minimization and Mutation Energetics Calculation in Rosetta

Using Pymol (version 2.5), fibril PDB structures were created with nine layers using the CBD_T1 fibril (PDB ID: 6tjo) and Pick's Disease (PDB ID: 6gx5) tau fibril structures. Structural alignment was used to superimpose the top two chains with the bottom two chains from a duplicated fibril assembly, preserving the geometry of the assembly while extending the fibril length. Overlapping chains were removed and chains were renamed to a 9-layer fibril assembly with chain lettering increasing in alphabetical order from the top to the bottom layer. These assemblies were then used as input for the subsequent mutagenesis and minimization in Rosetta. Changes in assembly energy were calculated using a method

adapted from the Flex ddG protocol (Barlow, K.A., et al., 2018). From the input assembly, a set of pairwise atom constraints with a maximum distance of 9 Ang are generated with a weight of 1, using the fa_talaris2014 score function. Using this constrained score function, the structure is then minimized. After minimization, the residues within 8 angstroms of the mutation site are subjected to backrub sampling, creating a set of sampled structures capturing backbone variation. These sampled structures are either only repacked and minimized, or the desired mutation(s) is introduced, followed by repacking and minimization. This is repeated for thirty-five independent replicates. The Rosetta interface change in energy mover is used as in the Flex ddG protocol to allow additional analysis of the $\Delta\Delta G^{interface}$ by defining a fibril interface, the bound WT and mutant structures reported by the interface ddG mover were used for estimating the change in assembly energy due to mutations. The lowest energy bound mutant and bound WT structure energies from each replicate are extracted, subtracted to give a $\boxed{\Delta \mathrm{REU}^{\mathrm{assembly}}_{\mathrm{mut-wt}}}$, and averaged over the 35 replicates to yield a $\overline{\Delta REU^{assembly}_{mut-wt}}$ of mutation for the given substitution(s).

$$\boxed{\Delta \mathrm{REU}^{\mathrm{assembly}}_{\mathrm{mut-wt}} = \frac{1}{n\_replicates} \sum_{replicate} bound_{mut} - bound_{wt}}$$

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.

Supplemental Table 1. Parameters used on flow cytometry data collection of tauRD-mEOS3.2 expression system.

Collected parameter	Voltage	Excitation (nm)	Emission filter (nm)
Forward Scatter	340	488	-
Side Scatter	190	488	488/10 BP
Acceptor, "mCherry"	340	561	595 LP 610/20 BP
Donor, "Alexa Fluor 488"	200	488	505 LP 530/30 BP
FRET, "PerCP"	340	488	595 LP 610/20 BP

Author Contributions

D.C. and L.A.J. initiated the project. D.C. purified all proteins involved in the study. D.C. performed all peptide and protein aggregation experiments. D.C. performed TEM of tau fibrils. D.C. and A.W. performed MD calculations and all structural analyses. D.C. and J.V.A. carried out the cell-based aggregation experiments. D.C. and V.M. performed the Rosetta design calculations on tauopathy fibrils. Finally, D.C. and L.A.J. conceived of and directed the research as well as wrote the manuscript. All authors contributed to the revisions of the manuscript. S.M. performed immunostaining of brain tissues. H. S. and J.C.S. provided knowledge on disease genetics and instructions on immunostaining. M.I.D. provided valuable insights.

Acknowledgement

This chapter was replicated in full from my first author manuscript in preparation.

CHAPTER FIVE
Concluding Remarks and Discussion

Conclusion and Future Directions

The intrinsically disordered nature of tau impedes the biophysical study to understand mechanisms underlie its biological function and in disease. Building upon the established method of crosslinking mass spectrometry (XL-MS) on studying protein complexes, I developed a method that couples XL-MS with temperature denaturation to examine protein unfolding with sequence resolution. It is anticipated that our approach can be generalized to other intrinsically disordered proteins and will help discover critical structural elements to better understand important biological questions including protein aggregation.

Using XL-MS and together with in silico simulations, in vitro and in cell aggregation assays, we characterized the stability of local structures shielding the amyloid motif in P301L/S mutants and WT. Our data supported the hypothesis that destabilization of the protective local structure by mutations for example can allow the amyloid motif ^{306}VQIVYK^{311} to sample more exposed conformations and hence aggregation (Chen et. al, 2019). The S320F mutation despite located far from ^{306}VQIVYK^{311}, was demonstrated to control the ^{306}VQIVYK^{311} driving aggregation allosterically through competitive interaction with the protective sequence motif of ^{306}VQIVYK^{311}. This was again tested by a combination of computational and experimental methods. The consensus on the aggregation mechanism where ^{306}VQIVYK^{311} is essential in both mutations proximate to and far from the inter-repeat interfaces highlights the possibly unifying theme in tau aggregation.

For future directions on tau mutation studies, several follow-up projects can be conducted to better understand the mechanism of tau aggregation. First, as we have demonstrated that stabilizing local hydrophobic cluster present in disease fibrils can facilitate tau aggregation, it will be valuable to check the presence of this local structure in the fibrillar end product. To accomplish this, Cryo-EM can be done to solve the structure of the S320F, S320I, or S320I/I328V tau mutant filament. This will allow us to test the connection between the aggregation-promoting structural component in mutant monomer and its fibrillar end product. Secondly, the aggregation mechanism in response to mutations in other repeats causing inherited tauopathy can also be investigated to expand our understanding on tau aggregation. Mutations that are interesting to look into include S352L, which is at the homolog position of S320F and is also a polar to hydrophobic substitute. Molecular Dynamics can be leveraged here again to gain insights into local structures encompassing the mutation; heat denaturation with XL-MS can be applied to measure the stability of those local structure; total crosslinker modifications can be analyzed to examine the solvent accessibility; the aggregation kinetics of sequence fragments can be tested to reveal the important elements in its aggregation. Aggregation mechanism of mutations such as V337M at the interface of R3 and R4 can also be investigated to further confirm the importance of local structure in protecting amyloid motifs.

Together, studies on tauopathy-causing mutations will help us understand the structural principals behind tau aggregation and bridge the gap between aggregation and disease.

Discussion

Thoughts on 3R vs. 4R in Tauopathy

Proband with S320F tau mutation showed Pick's Disease like clinical phenotype, and there was presence of Pick-like bodies revealed by immunopathology. However, not all the tau-positive inclusions resemble Pick bodies and there was presence of "more diffused" staining of nuclei (Rosso, 2002). Moreover, 3R and 4R tau were found equally present in the sarkosyl insoluble fraction (Rosso, 2002), which is distinct from Pick's Disease (PiD) where only 3R tau is present in tau filament. This suggests S320F mutation results in a disease different from the PiD. It is possible that the S320F mutant tau filament formed its own inclusion while it's still the WT 3R tau that forms the Pick Bodies in those patients. An immunohistochemistry assay using 3R and 4R antibodies will aid in this identification.

For a long time, it was unclear if disease follows tau dysfunction or vice versa until the discovery of FTDP-17 causing tau mutations (Goedert, 2005). Among the 62 disease associated *MAPT* mutations, about half of the have their effect on the protein level while the other half on the RNA level. Although tau is among many microtubule-associated proteins, its presence is found to be essential, as four independently created tau knockout mouse lines showed behavioral impairment and motor deficits when aged, albeit no overt phenotype in early stage (Ke, 2012). For mutations having primary effect on protein level, most of them in accordance with their location in microtubule-binding domain, were found to impair the ability of tau to promote microtubule assembly. S305N and Q336R are exceptions as they to the opposite slightly increase the ability of promoting microtubule assembly (Hasegawa, 1999; Pickering-Brown, 2004). However, S305N was found to cause an increased ratio of

4R/3R tau expression while Q336R was found to promote tau aggregation (Hasegawa, 1999; Pickering-Brown, 2004). The imbalance of 4R tau expression and increased tau aggregation propensity could have overcome the enhanced microtubule binding effect and thus led to disease.

Most of the mutations affecting alternative mRNA splicing cause more frequent exon 10 inclusion, which results in an excess expression of 4R tau isoforms compared to 3R tau. 4R tau is found to interact more strongly with microtubule as well as to better promote assembly of microtubule than 3R tau (Jakes, 1990). This argues against the hypothesis of loss of function as a disease mechanism. In most cases, toxic gain of function seems to be the culprit, as 4R tau was found to show greater aggregation propensity than 3R tau. However, because of the few exceptions of disease associated mutations such as K280del where 3R other than 4R tau is overproduced (Momeni, 2007), it is still unknown how those mutations that affect alternative splicing trigger disease as the protein expressed is WT, which was found resistant to aggregation in in vitro studies. There seem to be other factors concomitant with the imbalance of 4R/3R tau that drive tauopathies other than simply loss of function from microtubule binding or toxic gain of function from tau aggregation. A hypothesis is that an imbalance of 3R and 4R tau could cause disruption of microtubule dynamics, which could lead to apoptosis. However, details remained to be revealed.

Solid-state NMR has identified intermolecular disulfide bond between C322 residues in heparin-induced 3R tau filaments (Daebel, 2012). The filament formation of 4R tau was found inhibited under oxidizing conditions, whereas 3R tau forms filaments more readily under oxidizing conditions than under reducing conditions (Barghorn and Mandelkow, 2002;

Schweers et al., 1995; Wille et al., 1992). In the publication of cryo EM structure for heparin-induced tau, Zhang and others found that heparin-induced 3R tau dimer harbors C322 within a disulfide bond distance although their filaments are formed in reducing conditions (Zhang, 2019). This suggests that under oxidizing conditions, the formation of disulfide bond can potentially facilitate 3R tau fibril formation. However, no disulfide bond is identified in either the structure of 4R/3R (Fitzpatrick, 2018) or 3R only tau filaments (Falcon, 2019) from diseases. The wide Pick filament (WPF) from PiD has two intermolecular cysteines (from different filaments) seemingly in close distance to one another, but the authors commented that the interface of the two protofilaments in WPF is very likely connected by van der Waals interaction and not a disulfide bond given the distance (Falcon, 2019). Their argument was supported by the observation that WPFs were stable under reducing conditions. Although no intermolecular disulfide bond interaction was identified in 3R tau filament from PiD, we cannot rule out the possibility of its potential involvement in the initiation of 3R filament formation.

In sporadic PiD patients, only 3R tau was found in the tau filament while in CBD patients, only 4R tau was involved in the filament. This is very confounding because without the presence of any mutation that could alter splice isoform expression, 4R tau and 3R tau are supposedly both expressed to comparable level as they are in healthy controls. In fact, a study that analyzed the isoform expression level in multiple sporadic tauopathies showed that several cases of PiD actually had slightly higher 4R/3R tau expression and in CBD, PSP or AD cases the expressions of 4R and 3R tau are similar to those of control cases (Umea, 2004). Why do specific isoforms selectively aggregate in different diseases? It’s intuitive that 3R

and 4R doesn't seed onto each other in 3R only or 4R only filaments, as the filament core structure involves an isoform-specific region, for example R1R3 (without R2) in the 3R context or R2R3 in the 4R context. Nevertheless, there hasn't been compelling evidence in why only 3R tau forms filaments while 4R tau doesn't in one disease or vice versa as they only differ by having or not having the second pseudo-repeat and can both fibrillize in vitro. One possibility supported by a study conducted in a mouse model is that not all cells express 3R and 4R tau equally and it was those cells predominantly expressing 3R or 4R where selective aggregation of specific isoforms in diseases originates from (McMillan, 2008). It is likely that once those initial seeds with a particular shape is formed, 3R tau filament will not be able to template 4R tau or vice versa.

Additional Thoughts on Tau Filament Formation in Hereditary Tauopathy

The Cryo-EM studies on tau filaments revealed that AD, CTE, PiD, CBD, and PSP are distinguishable in terms of the tau filament shape which is consistent in individuals of the same disease. This suggests different shapes of tau filament are in correlation with clinical differences in tauopathies. However, for tauopathies caused by *MAPT* mutations and show clinical phenotypes of PiD, it might be a different story. The core structure of tau filament in PiD is composed of residue 254-378 without R2 (Falcon, 2019). Disease-associated mutations located within this region that give rise to Pick's clinical phenotype are K257T, G272V, K280del, S320F/Y, Q336H/R, K369I. Among these mutation sites, K257, Q336 and K369 are solvent accessible in the core structure of PiD filament so mutations on these residues might not alter the structure and might present in a very similar tau filament shape as

in PiD. G272V and S320F/Y, on the other hand, are facing inward in the filament structure which can potentially cause a clash given how densely packed at around 272 and the bulkiness of F/Y residue. K280del (or K311 in 3R isoform) will also very likely alter the filament structure as it is a deletion. These facts suggest that patients with G272V, S320F/Y or K280del might form tau filaments with a different fold regardless the difference is large or nuanced. However, none of the patients with the above listed mutations were diagnosed with clear PiD but rather tauopathies with symptoms consistent with PiD. Resolving the tau filament structure from these patients will help further reveal the connection between clinical symptoms and tau filament fold, which will provide a stronger basis for design of diagnostics and therapeutics.

As FTDP-17 mutations cause autosomal dominant disorders, patients with those mutations can have both mutant and WT tau expressed. Some of these mutations, such as P301L and P301S will only affect 20–25% of tau molecules with the rest being wild-type (Goedert, 2005). There hasn't been a structure solved for fibrils of tau with missense mutation from human brains. We don't know if the mutant and WT tau form filaments exclusively or can template each other within the same filament in disease. However, we do know WT fibril can seed mutant tau or vice versa in cells. Immunochemistry studies on brains from two R406W patients also showed that mutant tau and WT tau co-localize in neurofibrillary tangles (Miyasaka, 2001). Resolving the tau filament structures from these patients can unravel this conundrum.

Same mutation within kindreds does not always manifest the same clinical phenotype. For example, a member of the family carrying P301S mutation on tau presented Frontal

Temporal Dementia while his son carrying the same mutation presented CBD (Bugiani, 1999). This suggests that simply identifying mutation does not translate to disease diagnosis. With publications on the structure of tau filaments extracted from AD, CBD, PD, CTE where multiple patients with same disease were found to have same structure of tau filaments, we tend to think filament structure has close relation with disease phenotype. However, the fact that kindreds with same mutation present different phenotype implicates that other factors besides tau fibrilization should be considered for disease diagnosis, such as pathways of how tau fibrils are spread.

BIBLIOGRAPHY

Alzforum Mutation Databse. https://www.alzforum.org/mutations/mapt. Accessed 2018.

Abraham, M. J. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1–2. doi:10.1016/j.softx.2015.06.001

Arakhamia, T., Lee, C. E., Carlomagno, Y., Duong, D. M., Kundinger, S. R., Wang, K., . . . Fitzpatrick, A. W. P. (2020). Posttranslational Modifications Mediate the Structural Diversity of Tauopathy Strains. Cell, 180(4), 633-644.e612. doi:10.1016/j.cell.2020.01.027

Barducci, A., Bussi, G., & Parrinello, M. (2008). Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. Physical Review Letters, 100(2), 020603. doi:10.1103/PhysRevLett.100.020603

Barghorn, S., & Mandelkow, E. (2002). Toward a Unified Scheme for the Aggregation of Tau into Alzheimer Paired Helical Filaments. Biochemistry, 41(50), 14885-14896. doi:10.1021/bi026469j

Baughman, H. E. R., Clouser, A. F., Klevit, R. E., & Nath, A. (2018). HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. J. Biol. Chem., 293. doi:10.1074/jbc.M117.803411

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics, 81(8), 3684-3690. doi:10.1063/1.448118

Bergen, M. (2001). Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local beta-structure. J. Biol. Chem., 276. doi:10.1074/jbc.M105196200

Bugiani, O., Murrell, J. R., Giaccone, G., Hasegawa, M., Ghigo, G., Tabaton, M., . . . Ghetti, B. (1999). Frontotemporal Dementia and Corticobasal Degeneration in a Family with a P301S Mutation in Tau. Journal of Neuropathology & Experimental Neurology, 58(6), 667-677. doi:10.1097/00005072-199906000-00011

Cario, A., Savastano, A., Wood, N. B., Liu, Z., Previs, M. J., Hendricks, A. G., . . . Berger, C. L. (2022). The pathogenic R5L mutation disrupts formation of Tau complexes on the microtubule by altering local N-terminal structure. Proceedings

of the National Academy of Sciences, 119(7), e2114215119. doi:doi:10.1073/pnas.2114215119

Chen, D., Drombosky, K. W., Hou, Z., Sari, L., Kashmer, O. M., Ryder, B. D., . . . Joachimiak, L. A. (2019). Tau local structure shields an amyloid-forming motif and controls aggregation propensity. Nat Commun, 10(1), 2493. doi:10.1038/s41467-019-10355-1

Chirita, C. N., Congdon, E. E., Yin, H., & Kuret, J. (2005). Triggers of full-length tau aggregation: a role for partially folded intermediates. Biochemistry, 44. doi:10.1021/bi0500123

Cleveland, D. W., Hwo, S. Y., & Kirschner, M. W. (1977). Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. J. Mol. Biol., 116. doi:10.1016/0022-2836(77)90214-5

Daebel, V., Chinnathambi, S., Biernat, J., Schwalbe, M., Habenstein, B., Loquet, A., . . . Lange, A. (2012). β-Sheet Core of Tau Paired Helical Filaments Revealed by Solid-State NMR. Journal of the American Chemical Society, 134(34), 13982-13989. doi:10.1021/ja305470p

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys., 98. doi:10.1063/1.464397

Ding, F., Jha, R. K., & Dokholyan, N. V. (2005). Scaling behavior and structure of denatured proteins. Structure, 13. doi:10.1016/j.str.2005.04.009

Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology, 6(3), 197-208. doi:10.1038/nrm1589

Eichner, T., & Radford, S. E. (2009). A generic mechanism of B2-microglobulin amyloid assembly at neutral ph involving a specific proline switch. J. Mol. Biol., 386. doi:10.1016/j.jmb.2009.01.013

Elbaum-Garfinkle, S., & Rhoades, E. (2012). Identification of an aggregation-prone structure of tau. J. Am. Chem. Soc., 134. doi:10.1021/ja305206m

Eliezer, D., Barré, P., Kobaslija, M., Chan, D., Li, X., & Heend, L. (2005). Residual structure in the repeat domain of tau: echoes of microtubule binding and paired helical filament formation. Biochemistry, 44(3), 1026-1036. doi:10.1021/bi048953n

Falcon, B., Zhang, W., Murzin, A. G., Murshudov, G., Garringer, H. J., Vidal, R., . . . Goedert, M. (2018). Structures of filaments from Pick's disease reveal a novel tau protein fold. Nature, 561(7721), 137-140. doi:10.1038/s41586-018-0454-y

Falcon, B., Zhang, W., Schweighauser, M., Murzin, A. G., Vidal, R., Garringer, H. J., . . . Goedert, M. (2018). Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. Acta Neuropathol, 136(5), 699-708. doi:10.1007/s00401-018-1914-z

Falcon, B., Zivanov, J., Zhang, W., Murzin, A. G., Garringer, H. J., Vidal, R., . . . Scheres, S. H. W. (2019). Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. Nature, 568(7752), 420-423. doi:10.1038/s41586-019-1026-5

Feenstra, K. A., Hess, B., & Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. Journal of Computational Chemistry, 20(8), 786-798. doi:https://doi.org/10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B

Fischer, D. (2007). Structural and microtubule binding properties of tau mutants of frontotemporal dementias. Biochemistry, 46. doi:10.1021/bi061318s

Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., . . . Scheres, S. H. W. (2017). Cryo-EM structures of tau filaments from Alzheimer’s disease. Nature, 547(7662), 185-190. doi:10.1038/nature23002

Furman, J. L., Holmes, B. B., & Diamond, M. I. (2015). Sensitive Detection of Proteopathic Seeding Activity with FRET Flow Cytometry. Journal of visualized experiments : JoVE(106), e53205-e53205. doi:10.3791/53205

Ghetti, B., Oblak, A. L., Boeve, B. F., Johnson, K. A., Dickerson, B. C., & Goedert, M. (2015). Invited review: Frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging. Neuropathol Appl Neurobiol, 41(1), 24-46. doi:10.1111/nan.12213

Goedert, M., & Jakes, R. (1990). Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization. Embo j, 9(13), 4225-4230.

Goedert, M., & Jakes, R. (2005). Mutations causing neurodegenerative tauopathies. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1739(2), 240-250. doi:https://doi.org/10.1016/j.bbadis.2004.08.007

Goedert, M., Wischik, C. M., Crowther, R. A., Walker, J. E., & Klug, A. (1988). 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. doi:10.1073/pnas.85.11.4051

Goldschmidt, L., Teng, P. K., Riek, R., & Eisenberg, D. (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A, 107(8), 3487-3492. doi:10.1073/pnas.0915166107

Goode, B. L., Chau, M., Denis, P. E., & Feinstein, S. C. (2000). Structural and functional differences between 3-repeat and 4-repeat tau isoforms. Implications for normal tau function and the onset of neurodegenetative disease. J Biol Chem, 275(49), 38182-38189. doi:10.1074/jbc.M007489200

Gustafson, C. L. (2017). A slow conformational switch in the BMAL1 transactivation domain modulates circadian rhythms. Mol. Cell, 66. doi:10.1016/j.molcel.2017.04.011

Hasegawa, M., Smith, M. J., Iijima, M., Tabira, T., & Goedert, M. (1999). FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. FEBS Letters, 443(2), 93-96. doi:https://doi.org/10.1016/S0014-5793(98)01696-2

Holmes, B. B., Furman, J. L., Mahan, T. E., Yamasaki, T. R., Mirbaha, H., Eades, W. C., . . . Diamond, M. I. (2014). Proteopathic tau seeding predicts tauopathy in vivo. Proceedings of the National Academy of Sciences, 111(41), E4376-E4385. doi:10.1073/pnas.1411649111

Hutton, M. (1998). Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature, 393. doi:10.1038/31508

Ismail, T., & Kanapathipillai, M. (2018). Effect of cellular polyanion mimetics on tau peptide aggregation. J. Pept. Sci., 24. doi:10.1002/psc.3125

Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R., & Frydman, J. (2014). The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. Cell, 159. doi:10.1016/j.cell.2014.10.042

Kadavath, H. (2015). Folding of the Tau protein on microtubules. Angew. Chem. Int. Ed. Engl., 54. doi:10.1002/anie.201501714

Karamanos, T. K., Kalverda, A. P., Thompson, G. S., & Radford, S. E. (2015). Mechanisms of amyloid formation revealed by solution NMR. Prog. Nucl. Magn. Reson Spectrosc., 88–89. doi:10.1016/j.pnmrs.2015.05.002

Kaufman, S. K. (2016). Tau prion strains dictate patterns of cell pathology, progression rate, and regional vulnerability in vivo. Neuron, 92. doi:10.1016/j.neuron.2016.09.055

Kellogg, E. H., Hejab, N. M. A., Poepsel, S., Downing, K. H., DiMaio, F., & Nogales, E. (2018). Near-atomic model of microtubule-tau interactions. Science, 360(6394), 1242-1246. doi:10.1126/science.aat1780

Kier, B. L., Shu, I., Eidenschink, L. A., & Andersen, N. H. (2010). Stabilizing capping motif for beta-hairpins and sheets. Proc. Natl. Acad. Sci. USA, 107. doi:10.1073/pnas.0913534107

Kontsekova, E., Zilka, N., Kovacech, B., Skrabana, R., & Novak, M. (2014). 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. doi:10.1186/alzrt277

Kuret, J. (2005). Evaluating triggers and enhancers of tau fibrillization. Microsc. Res. Tech., 67. doi:10.1002/jemt.20187

Lange, O. F. (2012). 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. doi:10.1073/pnas.1203013109

Lee, S.-H., Le Pichon, C. E., Adolfsson, O., Gafner, V., Pihlgren, M., Lin, H., . . . Ayalon, G. (2016). Antibody-Mediated Targeting of Tau In Vivo Does Not Require Effector Function and Microglial Engagement. Cell Reports, 16(6), 1690-1700. doi:https://doi.org/10.1016/j.celrep.2016.06.099

Leitner, A. (2012). The molecular architecture of the eukaryotic chaperonin TRiC/CCT. Struct./Fold. Des., 20. doi:10.1016/j.str.2012.03.007

Leitner, A., Joachimiak, L. A., Unverdorben, P., Walzthoeni, T., Frydman, J., Förster, F., & Aebersold, R. (2014). Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. Proceedings of the National Academy of Sciences of the United States of America, 111(26), 9455-9460. doi:10.1073/pnas.1320298111

Leitner, A., Walzthoeni, T., & Aebersold, R. (2014). Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. Nature Protocols, 9(1), 120-137. doi:10.1038/nprot.2013.168

Lim, J. (2008). Pin1 has opposite effects on wild-type and P301L tau stability and tauopathy. J. Clin. Investig., 118. doi:10.1172/JCI34777

Lin, M. M. (2017). Leveraging symmetry to predict self-assembly of multiple polymers. Chem. Phys. Lett., 683. doi:10.1016/j.cplett.2017.05.011

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins, 78(8), 1950-1958. doi:10.1002/prot.22711

McMillan, P., Korvatska, E., Poorkaj, P., Evstafjeva, Z., Robinson, L., Greenup, L., . . . D'Souza, I. (2008). Tau isoform regulation is region- and cell-specific in mouse brain. The Journal of comparative neurology, 511(6), 788-803. doi:10.1002/cne.21867

Meng, W., Lyle, N., Luan, B., Raleigh, D. P., & Pappu, R. V. (2013). Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. Proc. Natl. Acad. Sci. USA, 110. doi:10.1073/pnas.1216979110

Mirbaha, H., Chen, D., Morazova, O. A., Ruff, K. M., Sharma, A. M., Liu, X., . . . Diamond, M. I. (2018). Inert and seed-competent tau monomers suggest structural origins of aggregation. eLife, 7. doi:10.7554/eLife.36584

Miyasaka, T., Morishima-Kawashima, M., Ravid, R., Heutink, P., van Swieten, J. C., Nagashima, K., & Ihara, Y. (2001). Molecular Analysis of Mutant and Wild-Type Tau Deposited in the Brain Affected by the FTDP-17 R406W Mutation. The American Journal of Pathology, 158(2), 373-379. doi:https://doi.org/10.1016/S0002-9440(10)63979-X

Mocanu, M. M. (2008). 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. doi:10.1523/JNEUROSCI.2824-07.2008

Mok, S. A. (2018). Mapping interactions with the chaperone network reveals factors that protect against tau aggregation. Nat. Struct. Mol. Biol., 25. doi:10.1038/s41594-018-0057-1

Momeni, P., Pittman, A., Lashley, T., Vandrovcova, J., Malzer, E., Luk, C., . . . de Silva, R. (2009). Clinical and pathological features of an Alzheimer's disease patient with the MAPT Delta K280 mutation. Neurobiology of aging, 30(3), 388-393. doi:10.1016/j.neurobiolaging.2007.07.013

Mukrasch, M. D. (2005). Sites of tau important for aggregation populate {beta}-structure and bind to microtubules and polyanions. J. Biol. Chem., 280. doi:10.1074/jbc.M501565200

Mukrasch, M. D. (2007). Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J. Am. Chem. Soc., 129. doi:10.1021/ja0690159

Mukrasch, M. D. (2007). The 'jaws' of the tau-microtubule interaction. J. Biol. Chem., 282. doi:10.1074/jbc.M607159200

Mukrasch, M. D. (2009). Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol., 7. doi:10.1371/journal.pbio.1000034

Mylonas, E., Hascher, A., Bernadó, P., Blackledge, M., Mandelkow, E., & Svergun, D. I. (2008). Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry, 47(39), 10345-10353. doi:10.1021/bi800900d

Nicholl, D. J., Greenstone, M. A., Clarke, C. E., Rizzu, P., Crooks, D., Crowe, A., . . . Heutink, P. (2003). An English kindred with a novel recessive tauopathy and respiratory failure. Ann Neurol, 54(5), 682-686. doi:10.1002/ana.10747

Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., & Fersht, A. R. (1997). The folding pathway of a protein at high resolution from microseconds to seconds. Proceedings of the National Academy of Sciences, 94(3), 826-830. doi:10.1073/pnas.94.3.826

O'Nuallain, B. (2006). Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. Methods Enzymol, 413. doi:10.1016/S0076-6879(06)13003-7

Ovchinnikov, S., Park, H., Kim, D. E., DiMaio, F., & Baker, D. (2018). Protein structure prediction using Rosetta in CASP12. Proteins, 86. doi:10.1002/prot.25390

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics, 52(12), 7182-7190. doi:10.1063/1.328693

Pastorino, L. (2006). The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. Nature, 440. doi:10.1038/nature04543

Pickering-Brown, S. M., Baker, M., Nonaka, T., Ikeda, K., Sharma, S., Mackenzie, J., . . . Mann, D. M. (2004). Frontotemporal dementia with Pick-type histology

associated with Q336R mutation in the tau gene. Brain, 127(Pt 6), 1415-1426. doi:10.1093/brain/awh147

Raman, S. (2010). NMR structure determination for larger proteins using backbone-only data. Science, 327. doi:10.1126/science.1183649

Reumers, J., Maurer-Stroh, S., Schymkowitz, J., & Rousseau, F. D. (2009). Protein sequences encode safeguards against aggregation. Hum. Mutat., 30. doi:10.1002/humu.20905

Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L. N., Beck, M., Schmidt, A., . . . Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. Nature Methods, 5(4), 315-318. doi:10.1038/nmeth.1192

Rizzu, P. (1999). 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. doi:10.1086/302256

Rosso, S. M., van Herpen, E., Deelen, W., Kamphorst, W., Severijnen, L. A., Willemsen, R., . . . van Swieten, J. C. (2002). A novel tau mutation, S320F, causes a tauopathy with inclusions similar to those in Pick's disease. Ann Neurol, 51(3), 373-376. doi:10.1002/ana.10140

Sanders, D. W. (2014). Distinct Tau prion strains propagate in cells and mice and define different tauopathies. Neuron, 82. doi:10.1016/j.neuron.2014.04.047

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., . . . Eisenberg, D. (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature, 447(7143), 453-457. doi:10.1038/nature05695

Scheres, S. H. W., Zhang, W., Falcon, B., & Goedert, M. (2020). Cryo-EM structures of tau filaments. Current Opinion in Structural Biology, 64, 17-25. doi:https://doi.org/10.1016/j.sbi.2020.05.011

Schoch, K. M. M. (2016). Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse Model. Neuron, 90. doi:10.1016/j.neuron.2016.04.042

Schweers, O., Mandelkow, E. M., Biernat, J., & Mandelkow, E. (1995). Oxidation of cysteine-322 in the repeat domain of microtubule-associated protein tau controls the in vitro assembly of paired helical filaments. Proceedings of the National Academy of Sciences, 92(18), 8463. doi:10.1073/pnas.92.18.8463

Sharma, A. M., Thomas, T. L., Woodard, D. R., Kashmer, O. M., & Diamond, M. I. (2018). Tau monomer encodes strains. eLife, 7. doi:10.7554/eLife.37813

Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. (2009). TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR, 44. doi:10.1007/s10858-009-9333-z

Shi, Y., Zhang, W., Yang, Y., Murzin, A. G., Falcon, B., Kotecha, A., . . . Scheres, S. H. W. (2021). Structure-based classification of tauopathies. Nature, 598(7880), 359-363. doi:10.1038/s41586-021-03911-7

Stelzl, L. S., Pietrek, L. M., Holla, A., Oroz, J., Sikora, M., Köfinger, J., . . . Hummer, G. (2022). Global Structure of the Intrinsically Disordered Protein Tau Emerges from Its Local Structure. JACS Au, 2(3), 673-686. doi:10.1021/jacsau.1c00536

Stöhr, J. (2017). A 31-residue peptide induces aggregation of tau's microtubule-binding region in cells. Nat. Chem., 9. doi:10.1038/nchem.2754

Strang, K. H., Croft, C. L., Sorrentino, Z. A., Chakrabarty, P., Golde, T. E., & Giasson, B. I. (2018). Distinct differences in prion-like seeding and aggregation between Tau protein variants provide mechanistic insights into tauopathies. J Biol Chem, 293(7), 2408-2421. doi:10.1074/jbc.M117.815357

Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D., & Eisenberg, D. (2006). The 3D profile method for identifying fibril-forming segments of proteins. Proceedings of the National Academy of Sciences of the United States of America, 103(11), 4074-4078. doi:10.1073/pnas.0511295103

Torbeev, V. Y., & Hilvert, D. (2013). Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate 2-microglobulin amyloid assembly. Proc. Natl. Acad. Sci. USA, 110. doi:10.1073/pnas.1310414110

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014). PLUMED 2: New feathers for an old bird. Computer Physics Communications, 185(2), 604-613. doi:https://doi.org/10.1016/j.cpc.2013.09.018

Umeda, Y., Taniguchi, S., Arima, K., Piao, Y.-S., Takahashi, H., Iwatsubo, T., . . . Hasegawa, M. (2004). Alterations in human tau transcripts correlate with those of neurofilament in sporadic tauopathies. Neuroscience Letters, 359(3), 151-154. doi:https://doi.org/10.1016/j.neulet.2004.01.060

Vaquer-Alicea, J., Diamond, M. I., & Joachimiak, L. A. (2021). Tau strains shape disease. Acta Neuropathol, 142(1), 57-71. doi:10.1007/s00401-021-02301-7

Vogel, M., Bukau, B., & Mayer, M. P. (2006). Allosteric regulation of Hsp70 chaperones by a proline switch. Mol. Cell, 21. doi:10.1016/j.molcel.2005.12.017

von Bergen, M., Barghorn, S., Li, L., Marx, A., Biernat, J., Mandelkow, E. M., & Mandelkow, E. (2001). Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local beta-structure. J Biol Chem, 276(51), 48165-48174. doi:10.1074/jbc.M105196200

Walker, S., Ullman, O., & Stultz, C. M. (2012). Using intramolecular disulfide bonds in tau protein to deduce structural features of aggregation-resistant conformations. J. Biol. Chem., 287. doi:10.1074/jbc.M111.336107

Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Förster, F., . . . Aebersold, R. (2012). False discovery rate estimation for cross-linked peptides identified by mass spectrometry. Nature Methods, 9(9), 901-903. doi:10.1038/nmeth.2103

Wille, H., Drewes, G., Biernat, J., Mandelkow, E. M., & Mandelkow, E. (1992). Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau in vitro. Journal of Cell Biology, 118(3), 573-584. doi:10.1083/jcb.118.3.573

Williams, D. R. (2006). Tauopathies: Classification and clinical update on neurodegenerative diseases associated with microtubule-associated protein tau. Intern. Med. J., 36. doi:10.1111/j.1445-5994.2006.01153.x

Yanamandra, K., Jiang, H., Mahan, T. E., Maloney, S. E., Wozniak, D. F., Diamond, M. I., & Holtzman, D. M. (2015). Anti-tau antibody reduces insoluble tau and decreases brain atrophy. Ann Clin Transl Neurol, 2(3), 278-288. doi:10.1002/acn3.176

Yoshiyama, Y. (2007). Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. Neuron, 53. doi:10.1016/j.neuron.2007.01.010

Zhang, W., Falcon, B., Murzin, A. G., Fan, J., Crowther, R. A., Goedert, M., & Scheres, S. H. (2019). Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. eLife, 8. doi:10.7554/eLife.43584

Zhang, W., Tarutani, A., Newell, K. L., Murzin, A. G., Matsubara, T., Falcon, B., . . . Scheres, S. H. W. (2020). Novel tau filament fold in corticobasal degeneration. Nature, 580(7802), 283-287. doi:10.1038/s41586-020-2043-0

Zhang, X., Lin, Y., Eschmann, N. A., Zhou, H., Rauch, J. N., Hernandez, I., . . . Han, S. (2017). RNA stores tau reversibly in complex coacervates. PLoS Biol, 15(7), e2002183. doi:10.1371/journal.pbio.2002183

Zhao, J. (2017). Glycan determinants of Heparin-Tau interaction. Biophys. J., 112. doi:10.1016/j.bpj.2017.01.024