Tau local structure shields amyloid motif and controls aggregation propensity

Dailu Chen[1,5*], Kenneth W. Drombosky[1*$], Zhiqiang Hou[1], Levent Sari[3,4], Omar M. Kashmer[1], Bryan D. Ryder[1,5], Valerie A. Perez[1,5], DaNae R. Woodard[1], Milo M. Lin[3,4], Marc I. Diamond[1], Lukasz A. Joachimiak[1,2#]

[1] Center for Alzheimer's and Neurodegenerative Diseases, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States

[2] Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States

[3] Green Center for Molecular, Computational, and Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States

[4] Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States

[5] Molecular Biophysics Graduate Program, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States

[*] These authors contributed equally

[$] current address: Mylan NV, 3711 Collins Ferry, Morgantown, WV 26505

[#] To whom correspondence should be addressed: Lukasz.Joachimiak@utsouthwestern.edu

ABSTRACT

Tauopathies are neurodegenerative diseases characterized by intracellular amyloid deposits of tau protein. Missense mutations in the tau gene (*MAPT*) correlate with aggregation propensity and cause dominantly inherited tauopathies, but their biophysical mechanism driving amyloid formation is poorly understood. Many disease-associated mutations localize within tau's repeat domain at inter-repeat interfaces proximal to amyloidogenic sequences, such as ^{306}VQIVYK^{311}. Using cross-linking mass spectrometry, intramolecular FRET, recombinant protein and synthetic peptide systems, *in silico* modeling, and cell models, we conclude that the aggregation prone ^{306}VQIVYK^{311} motif forms metastable compact structures with the upstream sequence that modulates aggregation propensity. Disease-associated mutations, isomerization of a critical proline, or alternative splicing are all sufficient to destabilize this local structure and trigger spontaneous aggregation. These findings provide a biophysical framework to explain the basis of early conformational changes that may underlie genetic and sporadic tau pathogenesis.

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[1]. In contrast, the tau repeat domain (tau RD), which spans residues 243 to 365, is predicted to be more structured[2], forms the core of amyloid fibrils[3], and is the minimal region to propagate tau prion strains[4]. 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*[5] and promotes pathology *in vivo*[6]. Nuclear Magnetic Resonance (NMR) experiments on tau indicate that in solution the ^{306}VQIVYK^{311} motif adopts a β-strand conformation[2,7]. Recent cryo-EM studies of tau patient-derived fibrils have shown that ^{306}VQIVYK^{311} mediates important contacts in these structures[3,8]. 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*[9-11]. 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[12]. Double electron-electron resonance (DEER) experiments indicated an expansion of this region upon heparin binding[9]. Cryo-EM structures also suggested an extended conformation of tau when bound to tubulin[13]. Other work mapping the recruitment of molecular chaperones to tau indicated that many chaperones, including Hsp40, Hsp70 and Hsp90, localize around ^{306}VQIVYK^{311} [14]. 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[15]. 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[16]. M_s itself adopts multiple stable structures that encode different tau prion strains[17], which are unique amyloid assemblies that faithfully replicate in living systems. Based on extrapolations, the existence of an aggregation prone monomer of tau had been previously proposed [18,19] but our study was the first to biochemically isolate and characterize this species[16]. Different forms of M_s have been purified from recombinant protein, and tauopathy brain lysates[16,17]. Using multiple low-resolution structural methods, we have mapped critical structural changes that differentiate M_i from M_s to near the ^{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[16]. 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[20]. 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[21]. This was recently supported by predicted models guided by experimental restraints from crosslinking mass spectrometry[16] and is consistent with independent NMR data[22,23].

Based on our prior work[16] 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**), such as P301L and P301S. These mutations have no definitive biophysical mechanism of action, but are nevertheless widely used in cell and animal models[25,26]. Solution NMR experiments on tau RD encoding a P301L mutation have shown local chemical shift perturbations surrounding the mutation resulting in an increased β-strand propensity[27]. 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[28].

101 As with disease-associated mutations, alternative splicing also changes the sequence N-terminal to
102 ^{306}VQIVYK^{311}. Tau is expressed in the adult brain primarily as two major splice isoforms: 3-repeat and 4-
103 repeat[29]. The truncated 3-repeat isoform lacks the second of four imperfectly repeated segments in tau
104 RD. Expression of the 4-repeat isoform correlates with the deposition of aggregated tau tangles in
105 many tauopathies[30] and non-coding mutations that increase preferential splicing or expression of the 4-
106 repeat isoform cause dominantly inherited tauopathies[30-32]. It is not obvious why the incorporation or
107 absence of the second repeat correlates with disease, as the primary sequences, although imperfectly
108 repeated, are relatively conserved.

109 Previous reports have focused on the structure of a repeat with the assumption that each repeat
110 functions independently within tau RD[33]. These have described a relationship between the length of a
111 repeat fragment, its propensity to spontaneously aggregate, and its seeding capacity in cells[33].
112 However, inter-repeat interactions may also influence aggregation given that both alternative splicing
113 and many disease-associated mutations cluster around the repeat interfaces (**Figure 1 a**). Our prior
114 work suggested that wild-type tau aggregates less efficiently because the flanking sequence shields
115 ^{306}VQIVYK^{311} [16]. We hypothesize that the intrinsically disordered tau protein evolved to minimize
116 aggregation by adopting local structure that shields the ^{306}VQIVYK^{311} amyloid motif from interactions
117 leading to seed formation and amyloid propagation. We predict that disease-associated mutations,
118 alternative splicing, or other factors can destabilize this local structure and expose ^{306}VQIVYK^{311} leading
119 to self-assembly. Therefore, we employed an array of *in silico*, *in vitro*, and cellular assays to elucidate
120 the molecular interactions and physiological consequences of ^{306}VQIVYK^{311} within tau.

121

122 RESULTS

123 **Disease-associated mutations at P301 promote rapid aggregation *in vitro* and in cells**. Missense
124 mutations that change proline 301 to leucine or serine cause dominantly inherited tauopathy[34] and are
125 associated with neurodegeneration in model systems[26,35], though the biophysical mechanism is not
126 understood. We studied changes in aggregation propensity driven by mutations at P301 in full-length
127 (FL) tau (2N4R; amino acids 1-441) and tau repeat domain (tau RD; amino acids 244-380). First, we
128 monitored aggregation of FL wild-type (WT) tau and mutant (P301L) tau using a Thioflavin T (ThT)
129 fluorescence assay induced with stoichiometric amounts of heparin. We observed that P301L tau ($t_{1/2}$ =
130 41.6 ± 0.5 hrs) aggregated more rapidly compared to WT tau ($t_{1/2}$ = 75 ± 0.3 hrs) (**Figure 1 c**). Next, we
131 compared changes in heparin-induced aggregation of the tau RD, comparing WT, P301L and P301S
132 mutants. We again observed that the two mutants aggregated faster (P301L tau RD, $t_{1/2}$ = 5.2 ± 0.1 hrs;
133 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**).
134 Consistently, we found that mutations at position 301 (from proline to either leucine or serine) increased
135 aggregation rates by ~2-fold compared to WT in both FL tau and tau RD constructs. Thus, *in vitro,* tau
136 RD recapitulates key aspects of aggregation observed in FL tau.

137 The inert conformation of monomeric tau (M_i) requires cofactors, such as heparin, to spontaneously
138 aggregate *in vitro*, while the seed-competent monomer (M_s), derived from recombinant protein or
139 Alzheimer's patient brain material, readily self-assembles to form amyloid[16]. Previously we determined
140 that M_s converts FL tau into fibrils at sub-stoichiometric ratios, in contrast to the stoichiometric amounts
141 necessary in heparin-containing reactions[16]. In this study, we evaluated the aggregation propensity of
142 the P301L mutant compared to WT when incubated in the presence of recombinantly produced M_s. We
143 incubated FL tau with sub-stoichiometric amounts of M_s (1:133) and monitored aggregation using ThT.
144 In comparison, we observed that M_s-seeded P301L tau self-assembled more rapidly (P301L tau, $t_{1/2}$ =
145 8.5 ± 0.6 hrs) than the WT protein (WT tau, $t_{1/2}$ = 40 ± 1.1 hrs) (**Figure 1 e).** P301L tau aggregated
146 faster than WT tau with a 4-fold increase in rate after seeding by M_s. Independent of induction – heparin
147 or M_s – P301L assembled into ThT positive aggregates more rapidly. Moreover, tau appeared to be
148 more sensitive to M_s seeded aggregation compared to heparin, given the sub-stoichiometric ratios
149 needed for robust aggregation. Based on these observations we next sought to determine why M_s

150 seeds appear to be more efficient than heparin at converting tau into fibrils, and why mutations at P301
151 increase tau's sensitivity to seeded aggregation.

152 We evaluated the change in conformation between M_i and M_s using fluorescence resonance energy
153 transfer (FRET). Tau encodes two cysteine residues at positions C291 in repeat 2 and C322 in repeat 3
154 (**Methods**, **Table 1, Supplemental Figure 1 a**), which serendipitously flank the amyloid motif
155 ^{306}VQIVYK^{311}**.** We labeled M_i and M_s at C291 and C322 with donor and acceptor fluorophores using
156 maleimide chemistry and measured their steady-state FRET intensities. M_i, which is predicted to be in a
157 more closed conformation[16], showed higher FRET efficiencies (E = 0.58 ± 0.12) compared to M_s (E =
158 0.25 ± 0.078) (**Supplemental Figure 1 b**). Denaturing M_i and M_s at 95 °C overnight yielded similar
159 FRET efficiencies (E = 0.24 ± 0.056, E = 0.29 ± 0.04, respectively) (**Supplemental Figure 1 b**).
160 Importantly, FRET efficiencies calculated for free dyes were low and remained constant at low and high
161 temperatures (**Supplemental Table 2**). FRET efficiencies for both 95 °C denatured proteins were only
162 modestly above the background FRET efficiency for the free dye negative control (**Supplemental Table**
163 **2**), suggesting a minimal interaction between these two regions. Furthermore, FRET from M_s was
164 reduced even at room temperature, likely reflecting a similarly extended conformation of the repeat 2
165 and 3 region versus a relatively compact conformation in M_i. Thus, the effectiveness of M_s to seed
166 aggregation of M_i may be explained by a direct change in conformation proximal to the ^{306}VQIVYK^{311}
167 motif, which is central to tau aggregation. This suggested that mutations at the P301 may exacerbate
168 aggregation by unfolding the region surrounding the ^{306}VQIVYK^{311} motif, thereby producing a more
169 compatible conformation for the similarly expanded aggregation-prone M_s seed.

170 To test the structural compatibility of aggregates formed by *in vitro* tau models, we employed tau
171 biosensor HEK293 cells that stably express tau RD (P301S) fused to cyan or yellow fluorescent
172 proteins [25]. These cells sensitively report a FRET signal (tau RD-CFP/tau RD-YFP) only when
173 aggregated in response to tau amyloid seeds, and are unresponsive to aggregates formed by other
174 proteins, such as huntingtin or α-synuclein[36]. Each sample formed amyloid fibril morphologies
175 confirmed by transmission electron microscopy, except for samples not incubated with heparin or M_s
176 and the low-concentration M_s, where no large ordered structures were found (**Supplemental Figure 2**).
177 The tau biosensor cells responded to FL tau fibrils created by exposure to heparin and showed an

178 increase in seeding activity for the P301L mutant compared to WT fibrils (**Figure 1 f**). Next, we
179 compared seeding for the tau RD heparin-induced fibrils and again found that P301L and P301S
180 mutants produced higher seeding activity relative to WT (**Figure 1 g**). Lastly, the seeding activity for the
181 M_s-induced FL tau fibrils showed a 2-fold higher activity for P301L compared to WT (**Figure 1 h**). WT
182 FL tau and tau RD control samples (no heparin or M_s) did not produce seeding activity in cells while
183 P301 mutants, both FL and tau RD, showed hints of seeding activity despite not yielding ThT signal *in*
184 *vitro* (data not shown), perhaps due to the formation of oligomers not captured by ThT. As expected,
185 33nM M_s control exhibited seeding activity at the onset and did not change after 5 days, but overall
186 signal was low due to the low concentrations used in the aggregation experiments. Interestingly, WT
187 tau induced with 33nM M_s seeded at similar levels to concentrated control (200nM) M_s samples
188 highlighting efficient conversion of WT tau into seed-competent forms (**Figure 1 h**). Thus, P301
189 mutations promote aggregation *in vitro* and in cells across different constructs. Importantly, these
190 effects are conserved between FL tau and tau RD.
191
192 **Mutation at P301 destabilizes native tau structure.** To determine how the P301L mutation drives
193 conformational changes we labeled WT tau RD and P301L tau RD with fluorophores at C291 and
194 C322, flanking the ^{306}VQIVYK^{311} motif (**Figure 2 a**), and monitored FRET in a heat denaturation time-
195 course experiment. FRET efficiency for WT tau RD gradually decreased at 65ºC or 75ºC, consistent
196 with denaturation of a metastable local structure (**Figure 2 b**). We next evaluated P301L tau RD and
197 found no detectible difference between monomeric WT tau RD FRET at physiological temperature
198 (37ºC). However, unlike WT, P301L tau RD rapidly lost FRET signal with heat denaturation (**Figure 2**
199 **c**), indicating perturbation of local structure. The magnitude of destabilization was consistent with the
200 differences observed for seeded elongation (**Figure 1 g**) which may not be readily detected by global
201 ensemble measurements like NMR.
202 We next used an orthogonal approach to test these ideas. Cross-linking mass spectrometry (XL-MS)
203 defines amino acid contacts in proteins and can thus guide the determination of structure for large
204 complexes, transient interactions and dynamics of intrinsically disordered proteins [16,37,38]. To preclude
205 the formation of intermolecular crosslinks between monomers, low concentration samples of WT,

206 P301L and P301S tau RD were incubated at different temperatures, reacted with disuccinimidyl
207 suberate (DSS; a primary amine crosslinker) for 1 minute and quenched (**Figure 2 d**). The crosslinked
208 protein monomers were confirmed by SDS-PAGE (**Supplemental Figure 3 a**). Cross-linked samples
209 were trypsin digested, analyzed by mass spectrometry and the spectra were searched using the
210 Xquest[39] to identify intramolecular protein contact pairs (**Methods and Supplemental Table 2**). In each
211 dataset, the cross-links reported represent consensus data across five independent samples with a low
212 false discovery rate (**Methods, Supplemental Table 2 and 3 and Supplemental Figure 4**). XL-MS of
213 recombinant WT tau RD acquired at 37ºC revealed three clusters of interactions that localize within the
214 N-terminus (residues 243-310; N-term), within the C-terminus (residues 311-380; C-term) and span N-
215 and C-termini (N-C; between residues 243-310 and 311-380) (**Figure 2 e, f**). Importantly, the
216 experimentally observed crosslinks represent only a small subset of all theoretical lysine pairs
217 suggesting that tau RD samples discrete modes of contacts (**Figure 2 f, grey circles**). Heating the
218 samples to 50ºC or even 75ºC decreased the number of N-C long-range and N-term short-range
219 contacts identified (**Figure 2 e, g, h and Supplemental Figure 3 b).** The data acquired for WT tau RD
220 in physiological conditions is consistent with a loose metastable structure comprised of weak long-
221 range and short-range contacts that are sensitive to temperature.
222 In contrast, XL-MS of recombinant tau RD with the P301L mutation revealed an increased susceptibility
223 to heat denaturation. At 37ºC, the crosslinks found in P301L tau RD (**Figure 2 i**) were similar in pattern
224 to WT, except for fewer N-C terminal long-range contacts **(Figure 2 e and Supplemental Figure 3 b,**
225 **c)**. However, samples incubated at 50ºC and 75ºC revealed a significant reduction in both long-range
226 and short-range contacts in P301L tau RD compared to WT **(Figure 2 e)**. The loss of short-range
227 contacts, both the total number of crosslinks and the abundance of each crosslink, were detected
228 particularly within the N-terminal sector, which sits upstream of P301 (**Figure 2 e, i-k and**
229 **Supplemental Figure 3 b, c**). In contrast, the C-term local contacts sample many theoretical lysine-
230 lysine pairs and remained relatively constant across temperatures for both WT and P301L tau RD
231 possibly suggesting a higher degree of disorder that is independent of the mutation site. Moreover, a
232 stepwise loss of the inter-repeat crosslinks between repeat 2 and 3 was seen in the heat denaturation
233 of WT tau RD and was more pronounced with the P301L mutation, indicating an unfolding of local

234 structure at the interface of repeat 2 and 3, encompassing ^{306}VQIVYK^{311} (**Figure 2 f-k, inset box**).
235 Thus, while the number of crosslinks identified was comparable between WT and P301L tau RD at
236 37ºC, P301L tau RD retained approximately half as many crosslinks as WT when heated. Similar
237 crosslinking profiles were observed for the P301S tau RD sample (**Supplemental Figure 3 d**). Thus,
238 the XL-MS data were consistent with findings from FRET labeling: P301L tau RD lacks the
239 thermostability observed in WT tau RD.
240 ***In silico* modeling of tau RD shows inter-repeat elements adopt local structure.** To gain insight
241 into what types of local structures the inter-repeat elements can form, we used ROSETTA to predict
242 structures of tau RD. We built 5,000 models using two parallel strategies in ROSETTA: *ab initio* which
243 employed fragment libraries derived from experimental structures[40] and CS-ROSETTA which leveraged
244 available chemical shifts for tau RD to produce fragment libraries[41]. Both approaches led to a diversity
245 of models consistent with experimentally determined radii of gyration[42] **(Supplemental Figure 5 a-d)**.
246 Analysis of each structural ensemble showed a propensity to form hairpin-like structures across R1R2,
247 R2R3, R3R4 and R4R' repeat interfaces centered on the ^{271}PGGG^{274}, ^{301}PGGG^{304}, ^{333}PGGG^{336} and
248 ^{366}PGGG^{369} sequences (**Supplemental Figure 5 e and 6**). Previously published solution NMR data
249 have shown that the PGGG sequences in tau can adopt type II β-turns[7], and the ^{301}PGGG^{304} sequence
250 preceding ^{306}VQIVYK^{311} is compatible with the formation of a β-hairpin. We illustrated the R2R3
251 ^{306}VQIVYK^{311}-containing fragment derived from low energy expanded models produced by each
252 method (**Supplemental Figure 5 c, d**). The ^{306}VQIVYK^{311}-containing interface has the highest
253 frequency of disease-associated mutations, particularly P301L and P301S (**Figure 1 a**). Other potential
254 amyloid-forming regions, such as ^{275}VQIINK^{280}, which is capable of aggregation (**Supplemental Figure**
255 **7**), is also preceded by ^{271}PGGG^{274} and predicted to form a β-hairpin (**Supplemental Figure 5 d and**
256 **Supplemental Figure 6**), however, it is absent in recent cryo-EM structures of tau aggregates[3,8].
257 Mapping known missense mutations onto the *ab initio* β-hairpin structure at R2R3 interface
258 (**Supplemental Figure 5 f**), we hypothesized that this cluster of disease-associated mutations could
259 destabilize the β-hairpin secondary structure, thus exposing the amyloid motif ^{306}VQIVYK^{311} and
260 enabling aggregation. This model is compatible with recent cryo-EM findings that indicate a

261 disengagement of the ^{306}VQIVYK^{311} N-terminal flanking sequence in a fibril structure[3]. Thus, we
262 focused our studies on the R2R3 motif of tau that contains ^{306}VQIVYK^{311}.

263 **P301L promotes extended forms of tau.** *In silico* modelling corroborated recent biochemical
264 findings[16] and suggested a minimal sequence necessary to form a collapsed structure around
265 ^{306}VQIVYK^{311}. To understand how these structures might self-assemble, we employed molecular
266 dynamics (MD) simulations of two tau peptide fragments comprising the minimally structured fragment
267 centered around the R2R3 interface (^{295}DNIKHVPGGGSVQIVYK^{311}): R2R3-WT and R2R3-P301L
268 (**Table 1 and Figure 4 b**). To enable sufficient sampling of oligomer structures, we employed an
269 unbiased algorithm based on a recently-developed symmetry-constraint approach[43]. The trimer
270 conformations obtained in simulations are depicted on a root-mean-square deviation (RMSD) matrix for
271 both R2R3-WT (**Figure 3 a**) and the R2R3-P301L mutant peptide fragments (**Figure 3 b**). For R2R3-
272 WT peptide fragment, we observed a dominant population of trimeric conformations composed of
273 hairpins, while the P301L disease-associated mutation stabilized an extended fibrillar form. The energy
274 basin for R2R3-WT peptide fragment was predicted to be 5 to 6 kJ/mol lower in a collapsed state than
275 an extended state, whereas the R2R3-P301L peptide fragment was 3 kJ/mol lower in an extended state
276 than a collapsed state (**Figure 3 c**). Additionally, the free energy surface suggested an energy barrier of
277 approximately 5 kJ/mol to convert R2R3-WT peptide fragment from collapsed to extended. That same
278 barrier however was less than 1 kJ/mole for the R2R3-P301L peptide fragment, suggesting a faster rate
279 of kinetic conversion from collapsed hairpin to extended fibrillar form. Thus, MD predicts that P301L
280 mutation promotes amyloid assembly by destabilizing monomeric hairpin structures.

281 **Tau amyloid formation is governed by flanking residues**. In tau RD, ^{306}VQIVYK^{311} is necessary for
282 amyloid formation[5,6]. In solution, ^{306}VQIVYK^{311} hexapeptide aggregates spontaneously and rapidly as
283 measured by ThT fluorescence intensity (**Supplemental Figure 7**) whereas the upstream N-terminal
284 sequence ^{295}DNIKHV^{300} does not aggregate. To experimentally test the prediction of a local hairpin
285 structure encompassing ^{306}VQIVYK^{311}, we employed a mutagenesis study on synthetic peptide systems
286 that recapitulate the minimal hairpin sequence (**Table 1**). Consistent with the prediction from MD
287 simulation, R2R3-WT peptide fragment did not aggregate readily, with no ThT detected within 96 hrs
288 (**Figure 4 a, c**). By contrast, single disease-associated mutations (**Figure 4 b**) substituted into the

289 R2R3 peptide fragment were sufficient to promote spontaneous amyloid formation: R2R3-P301S ($t_{1/2}$ =
290 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-
291 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 ±
292 1.3 hours**, Figure 4 c**). Each of these peptides was confirmed to form amyloid-like fibril morphologies
293 by transmission electron microscopy, except for the WT R2R3 peptide fragment where no large
294 structures were found (**Figure 5 b-h**).

295 To test the structural compatibility of peptide aggregates formed by *in vitro* tau models, we again
296 employed tau biosensor cells[25]. The tau biosensor cells responded to all disease-associated tau
297 peptide fragments that aggregated spontaneously *in vitro*, but not to the wild-type R2R3 peptide
298 fragment (which did not aggregate *in vitro*) (**Figure 5 a**). Qualitatively, biosensor cells retained their
299 diffused tau localization when untreated or exposed to a wild-type R2R3 peptide fragment but formed
300 fluorescent puncta when cultured with aggregated mutant peptides (**Figure 5 i-p**). Interestingly, the
301 biosensor cells responded to disease-associated mutant peptides with varying degrees of sensitivity
302 and created distinct aggregate morphologies. This is consistent with amyloid structures that act as
303 distinct templates and form the basis of tau prion strains[4,44]. Thus, R2R3 peptide fragment model
304 system responds to mutations *in vitro* and in cells similarly to the FL tau and tau RD system, suggesting
305 that local conformational changes in tau can be recapitulated using shorter fragments.

306 **Tau splice variants reveal differential aggregation propensity**. Tau is expressed in the adult brain
307 as 6 major splice isoform types that include either 3 or 4 repeated segments within RD (**Figure 6 a**). 3R
308 tau lacks the second of four imperfect repeats. 4R tau correlates strongly with aggregation in most
309 tauopathies[30] and mutations that increase splicing of the 4R isoform cause dominantly inherited
310 tauopathies[30-32]. We examined whether this splice isoform affects the propensity of ^{306}VQIVYK^{311}-
311 mediated aggregation due to the different composition of upstream flanking sequence. We constructed
312 a series of peptide fragments to encompass the R1R3 interface (**Table 1**). This wild type peptide
313 fragment R1R3 mimicking a 3R splice isoform did not spontaneously aggregate (data not shown).
314 Surprisingly, an R1R3 peptide fragment with a corresponding P301L mutation (R1R3-P270L) also did
315 not aggregate (**Figure 6**). We hypothesized that R1 leading sequence stabilizes the amyloid motif
316 ^{306}VQIVYK^{311} resulting in the aggregation resistance in the presence of disease-associated mutations.

317 The R1 leading sequence ^{264}**E**N**L**KH**Q**PGGG**K**^{273} differs from R2 ^{295}**D**N**I**KH**V**PGGG**S**^{304} at four amino
318 acid positions. To identify which amino acid(s) governed R1's stronger inhibitory effects, we constructed
319 16 peptides with a P301L mutation to represent every combinatorial sequence between the two leading
320 strands (**Figure 6 b**) and measured their aggregation kinetics (**Supplemental Figure 8)**. We identified
321 a general trend where the R2R3-P301L peptide fragment aggregates in hours with zero or one R1
322 substitutions. With two R1 substitutions, R2R3-P301L peptide aggregation was delayed roughly an
323 order of magnitude to tens of hours. With three R1 substitutions, R2R3-P301L peptide fragment
324 aggregation was further delayed to hundreds of hours. With all four R1 substitutions in the peptide
325 (R1R3-P301L), no ThT signal was observed within a week (**Figure 6 b, Supplemental Figure 8**). Thus,
326 all four amino acids contributed to the ability of the R1 leading sequence to delay ^{306}VQIVYK^{311}-
327 mediated spontaneous aggregation in a 3R splice isoform. This may explain the differential aggregation
328 propensities of tau isoforms in human pathology.

329 **Stabilizing β-hairpin structure blocks P301L-mediated aggregation.** Our model predicted shielding
330 of the ^{306}VQIVYK^{311} motif in tau *via* local β-structure, and thus we hypothesized that artificially stabilizing
331 the termini of this local structure would promote a more inert, closed conformation. To test this, we
332 flanked the R2R3-P301L peptide fragment with a tryptophan zipper (Trp-R2R3-P301L-Trp, **Table 1 and**
333 **Figure 7 a**), which stabilizes a β-hairpin structure approximately -2.5 to -7 kJ/mol[45]. Consistent with our
334 model, Trp-R2R3-P301L-Trp peptide fragment does not spontaneously aggregate *in vitro* (**Figure 7 b**).
335 To ensure that this effect wasn't a result of adding bulky tryptophan residues, we constructed control
336 peptide fragments that contain only the N-term (Trp-R2R3-P301L) or the C-term (R2R3-P301L-Trp)
337 portion of the tryptophan zipper sequence (**Figure 7 a**). Both half-sequence controls spontaneously
338 aggregated, implying that a tryptophan at either position is insufficient to block aggregation (**Figure 7**
339 **b**). Only a fully intact tryptophan zipper that stabilizes a β-hairpin conformation ameliorates aggregation
340 propensity. Alternative methods to stabilize a β-hairpin architecture, such as introducing isoelectric
341 interactions, also delayed aggregation: peptides containing two additional aspartic acid on the N-
342 terminus and two lysine on the C-terminus (R2R3-IEZip, **Table 1**) retarded R2R3-P301L peptide

343 fragment aggregation over an order of magnitude ($t_{1/2}$ = 7 hours to $t_{1/2}$ > 70 hours, **Supplemental**
344 **Figure 9**).

345 To test this effect in cells, we engineered tau RD (P301S) biosensor cells encoding tryptophan zipper
346 motifs that flank the R2R3 element. These biosensors had a significantly diminished capacity to be
347 seeded; R2R3-P301S peptide fragment aggregates triggered aggregation in 14.5 ± 1.6% of tau
348 biosensor cells, but only 0.3 ± 0.12% of the tryptophan zipper stabilized biosensor cells (**Supplemental**
349 **Figure 10**).

350 **Proline 301 *cis-trans* isomerization modulates aggregation.** Many proteins in the cell utilize proline
351 isomerization as a molecular switch, such as heat shock protein activation[46] or cell cycle regulation[47]. In
352 some proteins, proline isomerization directly induces or mitigates aggregation into amyloid[48-50]. Proline
353 isomerization events in tau have been proposed to play a role in aggregation and disease[48], but P301
354 isomerization has not been linked to tau aggregation and pathology. With the fact that serine or leucine
355 substitutions at P301 proximal to ^{306}VQIVYK^{311} drastically alter aggregation propensity, we
356 hypothesized that P301 plays a crucial role inducing a β-turn in a PGGG motif, which mediates a
357 collapsed structure. To test whether isomerization of P301 could influence spontaneous amyloid
358 formation, we constructed a series of R2R3 peptide fragments with proline analogs that preferentially
359 populate either: (1) a *cis* rotamer (2S,4S)-fluoroproline; (2) a *trans* rotamer (2S,4R)-fluoroproline; or (3)
360 an analog that easily interconverts between *cis* and *trans* (4,4)-difluoroproline (**Table 1, Figure 7 c**).
361 Only R2R3-Trans peptide spontaneously aggregated (**Figure 7 d**), indicating the potential for proline
362 isomerization events in tau pathogenesis.

363

364 DISCUSSION

365 Here we establish the molecular and functional basis for how a series of prominent tau mutations drive
366 aggregation. Integrating experimental and computational approaches, we independently and directly
367 probed the local structural changes within tau. We identified metastable local structures within the inter-
368 repeat junction of tau RD (the repeat 2-repeat 3 interface), which encompasses the amyloidogenic
369 ^{306}VQIVYK^{311} motif. This R2R3 interface becomes less stable when a disease-associated mutation is

370 present, such as P301L, which is commonly employed in cell and animal models of tauopathy. Thus,
371 P301L and similar mutations decrease the threshold for local structural expansion, especially in the
372 presence of stressors (heat, seeds, heparin, or high concentration). This in turn is predicted to enhance
373 the conversion of tau into a seed-competent form[16]. Thus, the proposed model rationalizes the
374 fundamental molecular mechanisms of aggregation for P301L and at least 5 other mutations, explains
375 why P301L spontaneously aggregates in animal and cellular models, and defines how splice isoforms
376 of tau and proline isomerization at P301 may contribute to aggregation. Ultimately, these insights may
377 inform the mechanisms of tauopathy in human disease and potential molecular targets for therapeutic
378 development.
379

380 **Mechanism of tau seeding**

381 *In vitro* induction of tau aggregation is typically achieved by the addition of polyanionic molecules such
382 as heparin, arachidonic acid or nucleic acids[9-11]. It is thought that heparin binding to tau expands the
383 local conformation of the repeat 2 and repeat 3 regions, thereby exposing amyloidogenic sequences for
384 subsequent aggregation[9,12,16]. This process, however, requires stoichiometric amounts of polyanion and
385 is not a physiological condition since heparin is not present intracellularly. Our recent work has
386 elucidated a seed competent form of tau monomer that can promote tau aggregation. This seed
387 competent monomeric tau is found in AD patient brains and is likely the incipient species contributing to
388 pathology[16]. We find that sub-stoichiometric amounts of M_s (1:133) enhance the rate of WT tau
389 aggregation relative to heparin. Parallel experiments with P301L tau show an even more dramatic
390 enhancement. Our data support that the ^{306}VQIVYK^{311} motif is preferentially exposed in M_s or P301L
391 mutant in contrast to normal tau where it is relatively shielded. Thus, the dramatic sensitivity of P301L
392 to seeds can be explained by an increased exposure of the aggregation-prone ^{306}VQIVYK^{311}
393 sequence. These data suggest that M_s functions catalytically to convert normal tau into aggregates.
394 Thus, the proposed seeding mechanism of M_s may be generalized to tauopathies that are not caused
395 by mutations.
396

397 **Perturbing metastable tau structures**

398 Ensemble averaging methods, such as NMR, have had limited success in understanding the solution
399 conformations of tau under physiological conditions. They have revealed secondary structure
400 propensities of key regions and proposed the existence of local contacts[2,7,22,23,51]. However, capturing
401 more transient or low population local conformations has been difficult. This is confounded by poor
402 signal to noise, requiring long acquisition times at high concentrations, and non-physiological
403 temperatures to suppress protein aggregation. As such, capturing transient but important local
404 structural signatures have been challenging with classical structural biology methods. Both experiment
405 and simulation have shown that weak local structure may play key roles in limiting aggregation of
406 globular proteins during translation and that these structural elements may play even bigger roles in
407 intrinsically disordered proteins[52,53]. Thus, local structures that bury proximal amyloid sequences may
408 be a general evolutionary design principle that controls aggregation.
409 Our study has suggested that local structure encompassing the amyloid motif ^{306}VQIVYK^{311} regulates
410 aggregation of tau and that the P301L mutation increases susceptibility to conformational changes that
411 expose the ^{306}VQIVYK^{311} amyloid motif. While these differences are subtle, we observe that P301L-
412 mediated structural rearrangements only manifest under moderate stress conditions (i.e. heat, seed).
413 Hence, as compared to NMR, real-time assays, such as FRET, and XL-MS that kinetically trap
414 conformations are more appropriate to detect metastable sub-populations. These data may explain the
415 elusiveness of a biophysical basis of the cluster of pathogenic mutations near ^{306}VQIVYK^{311}.
416 Simulations predict that repeat interfaces could encode local structures that are compatible with a β-
417 hairpin and that the P301L mutation, dramatically shifted the equilibrium away from collapsed hairpins
418 to extended fibril-like conformations. Our findings are consistent with published NMR data -- PGGG
419 sequences in tau can adopt type II β-turns[7] and that the P301L mutation increases local β-strand
420 propensity[27]. Thus, our work supports the structural and functional findings that metastable local
421 structures in tau are destabilized by disease-associated mutations.

422

423 **Peptides capture local structure observed in tau**

424 Guided by our simulations, we predicted that a local fragment spanning the interface between repeat 2
425 and 3 should encode a minimal structure necessary to replicate this aggregation phenomenon. We
426 examined whether structural perturbations influenced aggregation propensity in a peptide model
427 system that captures this local structural element. The WT tau interface peptide model containing
428 ^{306}VQIVYK^{311} did not aggregate spontaneously; however, single point substitutions of 6 disease-
429 associated mutations immediately N-terminal to ^{306}VQIVYK^{311} consistently induced spontaneous
430 aggregation. Given that destabilization of local structure around ^{306}VQIVYK^{311} promotes aggregation,
431 stabilizing local structure should rationally mitigate aggregation. By promoting a β-hairpin structure *via*
432 tryptophan zipper motifs or by using isoelectric forces, a P301L-containing tau peptide had an inhibited
433 propensity to aggregate.
434 Our data support the hypothesis that local forces are key to preventing aggregation of tau by
435 maintaining specific local structures. Tau is generally considered to be an intrinsically disordered
436 protein, and therefore long-range contacts are unlikely to play a significant role in stability. Published
437 NMR experiments support local structure formation of these regions in tau. Spectra of tau RD (K18;
438 amino acids 244 – 372) overlaps with a N- and C-terminally expanded tau RD (K32; amino acids 198 –
439 394) and even with the splice isoform of tau RD missing repeat 2 (K19; amino acids 244-372 with 275-
440 306 deleted)[7,51], suggesting that adding residues and even deleting an entire repeat have minimal
441 effects on the local structure. Thus, the conformations of local structures in tau are disproportionally
442 more important to its properties compared with structured proteins. This suggests that peptide fragment
443 models are a valid surrogate and can encapsulate the most relevant endogenous structural elements
444 for investigating aggregation of tau.
445
446 **Tau mutations drive disease by promoting rapid aggregation**
447 Disease-associated mutations found near tau's amyloid motif, such as P301L or P301S, have no
448 definitive biophysical mechanism but are nevertheless widely used in cell and animal models of
449 disease[25,26]. Using our peptide, tau RD and FL-tau model systems, we observe that key mutations
450 dramatically alter aggregation rates on similar time scales *in vitro* and seed in cell models. We therefore
451 provide an explanation for the toxic-gain-of-function for several mutations.

452 Previous reports have studied intra-repeat interactions, with the assumption that each repeat functions
453 independently within tau RD[33]. Their peptide models have shown a relationship between the length of a
454 peptide fragment and the seeding capacity of tau to define repeat 3 as the minimal sequence
455 necessary to act as a fully functional seed[33]. Our model defines a minimal regulatory sequence that
456 limits spontaneous aggregation and suggests that inter-repeat structural elements modulate
457 aggregation propensity. The composition of these inter-repeat sequences, governed by missense
458 mutations, directly impacts the stability of local structures and aggregation propensity. It is tempting to
459 speculate that local structure surrounding each of the four inter-repeat regions plays independent roles
460 in the exposure of amyloid sequences. This modular nature of the tau RD region may explain how
461 these independent regions can lead to distinct structures that define tau prion strains[4]. A more
462 comprehensive structure-function analysis of other repeat interfaces may help explain how each inter-
463 repeat element contributes to the formation of different tau structures.

464

465 **Splice variants modulate VQIVYK aggregation propensity**

466 The expression levels of the two major isoform types of tau in the central nervous system – 3R and 4R
467 – are similar in the adult brain[29]. However, the 3R:4R ratio of aggregate deposits is disproportionally
468 shifted towards 4R in most tauopathies[30]. Mutations in tau that affect alternative splicing and generate
469 excess 4R isoforms correlate with some genetic tauopathies[31,32]. The N-terminal flanking sequences
470 leading into ^{306}VQIVYK^{311} differ by four amino acids between the two isoforms. We find that these two
471 isoforms have drastically different aggregation propensities in the presence of disease-associated
472 mutations ($t_{1/2}$ = 7 hours vs $t_{1/2}$ > 100 hours, respectively). Chimeras of R1R3 / R2R3 transition from
473 aggregation-resistant to aggregation-prone as they lose R1 N-terminal flanking character. The ability of
474 an R1 leading strand to mitigate ^{306}VQIVYK^{311} aggregation may explain why 4R tau correlates more
475 closely with pathology. Thus, inter-repeat contacts may explain aggregation propensities of tau isoforms
476 in disease. Encouraging data for a tau vaccine targeting a ^{300}HXPGGG^{304} sequence suggests it's
477 possible to utilize inter-repeat regions to select between pathogenic and non-pathogenic conformations
478 of tau[54].

479

480 ***Cis-trans* isomerization and prolyl isomerases**

481 Studying the missense mutations in tau has generated valuable disease models [26,35]; however, the
482 majority of human tauopathies have no observed genetic mutation in tau[34]. Critical proline residues N-
483 terminal to the amyloid motif can isomerize into *cis* or *trans* rotamers spontaneously or through
484 unidentified cellular mechanisms. We observe that P301 *cis* and *trans* rotamers have distinct
485 aggregation propensities *in vitro*. In fact, the aggregation kinetics for a *trans* rotamer of P301 are on par
486 with some well-defined disease mutants (N296Δ, V300I). The concept of proline isomerization
487 triggering aggregation into amyloid is not novel, as this is an accepted mechanism of β2-microglobulin
488 aggregation in kidney dialysis amyloidosis[55]. Other proline residues outside of the tau repeat domain
489 have also been proposed to undergo proline isomerization[48]. Our proposed model suggests a possible
490 mechanism whereby WT tau aggregation could be controlled *in vivo*: specific prolyl isomerization
491 events – possibly triggered by cellular proline isomerases – could trigger spontaneous aggregation by
492 modulating inter-repeat structural elements.

493

494 **Conclusions**

495 We propose that sequences N-terminal to tau's amyloid motif forms local contacts consistent with a β-
496 hairpin-like compact structure. This shields the amyloid motif and mitigates aggregation (**Figure 8**). This
497 represents a simple yet comprehensive model of tau aggregation that unifies key observations
498 throughout tau literature.

499 Algorithms that identify potential amyloid nucleating regions, such as TANGO, have indicated that
500 nearly 75% of aggregation nucleating regions in the human proteome use two or more "gatekeeper"
501 residues, with proline being the most common single gatekeeping residue[56]. These gatekeeping
502 residues are more likely than average to be the site of disease-associated missense mutations and are
503 consistent with our identification of gatekeeping residues near tau's amyloid motif. Thus, local flanking
504 sequences and their structural contacts may play an important role in mitigating aggregation propensity
505 in tau and likely other intrinsically disordered proteins.

506 Finally, the identification and characterization of metastable compact structures encompassing
507 ^{306}VQIVYK^{311} may itself prove to be a valuable therapeutic target. One might be able to shift the
508 structural rearrangement of tau amyloid motif from exposed (aggregation prone) to buried (inert) using
509 small molecules, antibodies, or cellular co-factors. Our results indicate that subtle changes in local
510 structure have immense functional ramifications; therefore, small molecules that shift this structural
511 equilibrium modestly may have significant benefits.
512

513

514 METHODS

515 **Recombinant Full-Length tau and tau RD Production.** We utilized several forms of recombinant tau.
516 Full-length (FL) wildtype and P301L tau were cloned into pet28b vector (a kind gift from Dr. David
517 Eisenberg) using NcoI and XhoI sites. The plasmid was transformed into BL21-Gold (DE3) cells. Cells
518 were grown in 1x Terrific Broth media to OD600 1.4 and induced with 1 mM IPTG for 3 hours at 37ºC.
519 The cells were harvested and lysed in 50 mM Tris, 500 mM NaCl, 1 mM β-Mercaptoethanol, 20 mM
520 imidazole, 1 mM PMSF, pH 7.5, using an Omni Sonic Ruptor 400. The lysates were centrifuged, the
521 supernatant was applied to a Ni-NTA column and eluted with 50 mM Tris, 250 mM NaCl, 1 mM β-
522 Mercaptoethanol, 300 mM imidazole. Eluting fractions containing tau were desalted into 50 mM MES,
523 50 mM NaCl, 1mM β-Mercaptoethanol (pH 6.0) by PD-10 column GE. Exchanged fractions were
524 applied to a HiTrap SP HP (GE) and eluted with a 50mM-1M NaCl gradient. Tau containing fractions
525 were concentrated on an Amicon-15 concentrator and applied to a Superdex 200 Increase 10/300 GL
526 (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).
527 Tau RD wildtype and mutants were expressed and purified as previously described [57]. Aliquots were all
528 stored at -80ºC in 1x PBS. Tau seeding monomer (M_s) was produced as previously described [16].

529 **ThT Fluorescence Aggregation Assays**. Wildtype or mutant FL tau and tau RD protein was diluted in
530 1x PBS with 5μM β-Mercaptoethanol and boiled at 95ºC for 5 min. A final concentration of 4.4μM
531 heparin or 33nM M_s seed was added to 4.4μM tau or tau RD protein in a 60μL volume mixed with 25
532 μM ThT and aliquoted into a 96-well clear bottom plate. Peptides were disaggregated as previously
533 described[58], lyophilized, and resuspended in 2x PBS. 25 μM ThT was added to 200uL of 200 μM
534 peptide in a 96-well clear bottom plate. All conditions were done in triplicates (except for the R2R3-
535 IEZip experiment) at room temperature. ThT kinetic scans were run every 5 minutes on a Tecan M1000
536 plate reader at 446 nm Ex (5 nm bandwidth), 482 nm Em (5 nm bandwidth). Blank wells containing
537 buffer and ThT were subtracted from experimental values. Samples producing signal to background
538 (ThT only) with ratios only greater than 2:1 were considered and these values were normalized to the
539 maximum amplitude in each condition. The data were plotted, and the $t_{1/2}$ values were derived from a
540 non-linear regression model fitting in GraphPad Prism.

541 **Transmission electron microscopy (TEM).** An aliquot of 5 μl of sample was placed onto a glow-
542 discharged Formvar-coated 400-mesh copper grids for 30 seconds, washed with distilled water, and
543 then negatively stained with 2% uranyl acetate for 1 min. Images were acquired on a Tecnai G^2 spirit
544 transmission electron microscope (FEI, Hillsboro, OR), serial number: D1067, equipped with a LaB_6
545 source at 120kV using a Gatan ultrascan CCD camera.

546 **Tau Biosensor Cells**. Biosensor cells were plated into 96-well plates at 20,000 cells per well. For tau
547 and tau RD experiments, after 5 days of incubation with heparin or M_s, 10μL of 4.4μM aggregated
548 protein material was mixed with 1.25μL lipofectamine and 8.75μL Opti-MEM, incubated at room
549 temperature for 30 min, and added to cell media. The "t=0" samples were prepared in the same way but
550 straight from the freezer aliquots. After 2 days, cells were harvested with 0.05% trypsin, then
551 resuspended in Flow buffer (1x HBSS, 1% FBS, 1 mM EDTA, 1x DPBS) and analyzed by flow
552 cytometry. For peptide experiments, 10 μg of aggregated peptide material was added to 0.5 uL
553 lipofectamine and Opti-MEM to a total volume of 10 μL, incubated at room temperature for 30 min, and
554 added directly to cell media. After 3 days, cells were harvested with 0.05% trypsin, then resuspended in
555 Flow buffer and analyzed by flow cytometry. All conditions were done in triplicates. The Trp Zip
556 biosensor cells expressing the tryptophan zipper motifs flanking the R2R3 element in tau RD were
557 generated as previously described[25].

558 **Tau intramolecular FRET.** 10-fold excess TCEP was added to 100 μM tau RD, the atmosphere was
559 deoxygenated with N_2 gas, and incubated at RT for 30 minutes. To label, C_5 maleimide Alexa-488 and
560 C_2 maleimide Alexa-647 were dissolved in DMSO and added in 5-fold excess to tau RD, and incubated
561 at 4°C overnight. Alexa-labelled tau RD was purified using an 8 kDa cut-off mini dialysis exchanged into
562 1x PBS overnight. FRET measurements were carried out at 100nM tau RD labeled protein samples. Mi
563 and Ms were purified as described above. For M_i and M_s intramolecular FRET experiments, 4uM
564 protein samples were reduced with 10-fold excess TCEP, the atmosphere was deoxygenated with N_2
565 gas, and incubated at RT for 30 minutes. To label, C_5 maleimide Alexa-488 and C_2 maleimide Alexa-647
566 were dissolved in DMSO and added in 5-fold excess to M_i and M_s, and incubated at 4°C overnight.
567 Alexa-labelled M_i and M_s was purified using an 8 kDa cut-off mini dialysis exchanged into 1x PBS
568 overnight. FRET measurements were carried out at 100nM M_i and M_s labeled protein samples. FRET

569 efficiencies were calculated as a function of acceptor emission, given $E = (I_{AD}\varepsilon_{AA}-I_{AD}\varepsilon_{AD})/I_{AA}\varepsilon_{DD}$ where I_{AA}
570 is Acceptor intensity following Acceptor excitation, I_{AD} is Acceptor intensity following Donor excitation,
571 and ε is the extinction coefficient of the fluorophores at given excitations and emissions. Alexa 488
572 Ex/Em was measured at 490 nm/520 nm, Alexa 647 Ex/Em 647 nm/670 nm with a 5 nm bandwidth for
573 both. As a control, spectra for free 100nM Alexa 647 and Alexa 488 were acquired at 37°C, 50°C and
574 75°C at two time points (t=0 and t=30 minutes). FRET efficiencies remained flat and were not perturbed
575 by heating (**Supplemental Table 2**).

576 **Crosslinking, sample processing and LC-MS/MS analysis.** Preparation of tau RD was crosslinked
577 at a total protein concentration of 1.0 mg/mL using 100 µg of starting material. The crosslinking buffer
578 was 1X PBS with 3mM DTT. Five replicates for each condition (37°C, 50°C, and 75°C) were prepared.
579 Samples for 50°C and 75°C conditions were equilibrated at the appropriate temperature for 1 hour
580 before crosslinking. The crosslinking reaction was initiated by adding disuccinimidyl suberate (DSS)
581 stock solution (25 mM DSS-d_0 and –d_{12}, Creative Molecules) in DMF to a final concentration of 1 mM.
582 Samples were further incubated at 37°C, 50°C, or 75°C for 1 min with 350 RPM shaking. Excess
583 reagent was quenched by addition of Tris (pH 7.5) to 100 mM and incubation at 37ºC for 30 min, and
584 subsequently flash frozen by liquid nitrogen and evaporated to dryness by lyophilization. Proteins were
585 resuspended in 8 M urea, reduced with 2.5 mM TCEP (37°C, 30 min) and alkylated with 5mM
586 iodoacetamide (30min, room temperature, protected from light). The sample solutions were diluted to 1
587 M urea with 50 mM ammonium hydrogen carbonate and trypsin (Promega) was added at an enzyme-
588 to-substrate ratio of 1:50. Proteolysis was carried out at 37ºC overnight followed by acidification with
589 formic acid to 2% (v/v). Samples were then purified by solid-phase extraction using Sep-Pak tC18
590 cartridges (Waters) according to standard protocols. Samples were evaporated to dryness and
591 reconstituted in water/acetonitrile/formic acid (95:5:0.1, v/v/v) to a final concentration of approximately
592 0.5 µg/µl. 2µL each were injected for duplicate LC-MS/MS analyses on an Eksigent 1D-NanoLC-Ultra
593 HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system. Peptides were separated on self-
594 packed New Objective PicoFrit columns (11cm x 0.075mm I.D.) containing Magic C_{18} material
595 (Michrom, 3µm particle size, 200Å pore size) at a flow rate of 300nL/min using the following gradient. 0-
596 5min = 5 %B, 5-95min = 5-35 %B, 95-97min = 35-95 %B and 97-107min = 95 %B, where A =

597 (water/acetonitrile/formic acid, 97:3:0.1) and B = (acetonitrile/water/formic acid, 97:3:0.1). The mass
598 spectrometer was operated in data-dependent mode by selecting the five most abundant precursor ions
599 (m/z 350-1600, charge state 3+ and above) from a preview scan and subjecting them to collision-
600 induced dissociation (normalized collision energy = 35%, 30ms activation). Fragment ions were
601 detected at low resolution in the linear ion trap. Dynamic exclusion was enabled (repeat count 1,
602 exclusion duration 30sec).
603 **Analysis of mass spectrometry data.** Thermo .raw files were converted to the open .mzXML format
604 using msconvert (proteowizard.sourceforge.net) and analyzed using an in-house version of xQuest[50].
605 Spectral pairs with a precursor mass difference of 12.075321 Da were extracted and searched against
606 the respective FASTA databases containing tau (TAU_HUMAN P10636-8) or with a P301L/S
607 substitution. xQuest settings were as follows: Maximum number of missed cleavages (excluding the
608 crosslinking site) = 2, peptide length = 5-50 aa, fixed modifications = carbamidomethyl-Cys (mass shift
609 = 57.021460 Da), mass shift of the light crosslinker = 138.068080 Da, mass shift of mono-links =
610 156.078644 and 155.096428 Da, MS^1 tolerance = 10 ppm, MS^2 tolerance = 0.2 Da for common ions
611 and 0.3 Da for crosslink ions, search in ion-tag mode. Post-search manual validation and filtering was
612 performed using the following criteria: xQuest score > 25, mass error between -2.2 and +3.8ppm, %TIC
613 > 10, and a minimum peptide length of six aa. In addition, at least four assigned fragment ions (or at
614 least three contiguous fragments) were required on each of the two peptides in a crosslink. False
615 discovery rates (FDR's) for the identified crosslinks were estimated using xprophet[59] and estimated to
616 be 1.3 - 10% (**Supplemental Figure 4, Supplemental Table 3**). At each temperature, the 5 replicate
617 datasets were compared and only crosslinks present in 5 of the 5 datasets were used to generate a
618 consensus dataset (**Supplemental Table 4**). Crosslink data with information of crosslinked residue
619 positions and nseen (frequency) was visualized using customized gnuplot scripts.
620 **Model generation of tau RD using ROSETTA.** The backbone NH, N, CA, CB and C=O chemical shift
621 assignments for the tau fragment from 243-368 (bmrbId=19253) were used in CS-Rosetta to generate
622 fragment libraries for subsequent model refinement [60]. First, chemical shift parameters were used to
623 predict backbone torsional angles using TALOS to generate a CS-guided fragment library representing
624 the conformations of the protein[60]. For the *ab initio* ROSETTA calculations, the tau RD sequence was

625 used to generate 3-mer and 9-mer fragments derived from the protein data bank using the fragment
626 picker tool[40]. The Rosetta energy function was used to assemble and iteratively refine 5000 structural
627 models using each set of fragments[40,41,61]. Cα–based root mean square deviations were computed in
628 Rosetta for tau RD and hairpin fragments. Clustering analysis of the tau RD ensemble showed similar
629 results yielding median root mean square deviations of 19.5 Ang^2 and 19.75 Ang^2 for *ab initio* and CS-
630 ROSETTA simulations, respectively (data not shown). Ensemble wide calculation of cα-cα end to end
631 distances between residues 264-280, 295-311, 327-343 and 359-375 were carried out using a python
632 script. All simulations were performed on UTSW's biohpc computing cluster. All plots were generated
633 with gnuplot. Images were created using Pymol.

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

637 **Flow Cytometry**. A BD LSRFortessa was used to perform FRET flow cytometry. To measure CFP and
638 FRET, cells were excited with the 405 nm laser, and fluorescence was captured with a 405/50 nm and
639 525/50 nm filter, respectively. To measure YFP, cells were excited with a 488 laser and fluorescence
640 was captured with a 525/50 nm filter. To quantify FRET, we used a gating strategy where CFP bleed-
641 through into the YFP and FRET channels was compensated using FlowJo analysis software. We then
642 created a plot of FRET vs. CFP and introduced a triangular gate to assess the number of FRET-positive
643 cells, as previously described[16]. For each experiment, at least 10,000 single cells per replicate were
644 analyzed. Data analysis was performed using FlowJo v10 software (Treestar).

645 **MD Simulations.** Well-Tempered Metadynamics[62] was employed to enable accelerated conformational
646 sampling and to construct the associated free energy surface. Metadynamics was performed on a two-
647 dimensional space of parallel-β sheet content and anti-parallel sheet content. To increase search
648 efficiency in oligomeric space, we have incorporated conformational symmetry constraints, which have
649 been shown to enable sampling of multi-polymer landscapes[43]. The initial dodecahedron simulation box
650 was constructed from a trimer of a randomly unfolded structure of 295-311 by adding 7587 SPCE
651 explicit waters and 3 neutralizing Cl ions (one for each monomer). The AMBER99sb-ildn force-field[63]

652 was used for all simulations. After an initial 1009 steepest descent steps of converged energy
653 minimization, 10 ns of NVT and 20 ns of NPT (first 10 with Berendsen[64] and the last 10 with Parrinello-
654 Rahman[65] barostats) equilibrations were performed. The subsequent production level trajectories are
655 based on 5 fs time steps using hydrogen-only virtual sites[66]. Production level trajectories were obtained
656 for an NPT ensemble with Parrinello-Rahman barostat, and periodic boundary conditions with Particle
657 Mesh Ewald (PME) [67] summation for long-range electrostatics. The tuned well-tempered metadynamics
658 parameters are 10, 1.4 kJ/mole, and 0.3 for bias factor, Gaussian height, collective variable space
659 Gaussian widths, respectively. The Gaussian perturbations were included into MD every 2.5 ps using
660 the PLUMED package[68] as an external patch to Gromacs-5.0.4[69]. A total of 18 μs trajectories were
661 generated, 9 μs for wild type and 9 μs for the P301L mutant, over a total of 6 independent runs. All
662 simulations were done on UTSW's bioHPC computing cluster.

663 **Statistics:** All statistics were calculated using GraphPad Prism 8.0. Three independent ThT
664 experiments were run for each condition. The data were normalized to the highest amplitude and
665 averages and standard deviations were plotted. Plots were fitted to a non-linear regression model, from
666 which $t_{1/2}$ values were derived. $t_{1/2}$ error represents a 95% CI. Flow cytometry cell aggregation was
667 conducted in 3 independent experiments, whose values are plotted. Error bars represent a 95% CI.

668 **Data Availability:** ThT data for tau, tau RD and peptide experiments is available in Supplemental Table
669 1. Raw crosslinking mass spectrometry data is available in Supplemental Table 3 and 4. Other
670 supporting data available upon request from the authors.

671 **Acknowledgements**

672 This work was supported by grants from the Tau Consortium and NIH grants awarded to
673 1R01NS071835 (M.I.D.), the Effie Marie Cain Endowed Scholarship (L.A.J.) and the Heising-Simons
674 Foundation (M.M.L.). We appreciate the help of the Structural Biology Laboratory and Proteomics Core
675 Facility at the University of Texas Southwestern Medical Center.

676

677 **Author Contributions**

678 K.D. M.I.D., and L.A.J. conceived and designed the overall study. K.D. performed *in vitro* peptide
679 assays, intramolecular FRET, flow cytometry, and cell models. D.C. performed *in vitro* protein assays,
680 cell models, crosslink mass spectrometry, and ROSETTA simulations. Z.H. developed the tau *in vitro*
681 seeded aggregation assays. O.M.K. assisted with tau and tau RD purifications. V.A.P. helped with cell
682 seeding assays. B.D.R. and D.R.W. performed electron microscopy. L.S. and M.M.L. performed
683 molecular dynamics simulations. K.D., D.C. and L.A.J. wrote the manuscript, and all authors
684 contributed to its improvement.
685
686
687

688 **Table 1: List of Peptide and Protein Sequences**

Name	Amino Acid Sequence
VQIINK	----------VQIINK
VQIVYK	----------VQIVYK
R1R2 (264-280)	------ENLKHQPGGGKVQIINK
R2R3 (295-311)	------DNIKHVPGGGSVQIVYK
R2R3-N296Δ	------D-IKHVPGSSSVQIVYK
R2R3-V300I	------DNIKH**I**PGGGSVQIVYK
R2R3-P301L	------DNIKHV**L**GGGSVQIVYK
R2R3-P301S	------DNIKHV**S**GGGSVQIVYK
R2R3-G303V	------DNIKHVPG**V**GSVQIVYK
R2R3-S305N	------DNIKHVPGGG**N**VQIVYK
R1R3	------ENLKHQPGGGSVQIVYK
R1R3-P270L	------ENLKHQ**L**GGGSVQIVYK
R1R3-P270S	------ENLKHQ**S**GGGSVQIVYK
R1R3-G272V	------ENLKHQPG**V**GSVQIVYK
R1R3-G273R	------ENLKHQPGG**R**SVQIVYK
Trp-R2R3-P301L-Trp	**WTGK**SKDNIKHV**L**GGGSVQIVYK**EGGW**
Trp-R2R3-P301L	**WTGK**SKDNIKHV**L**GGGSVQIVYKPVDL
R2R3-P301L-Trp	SKCGSKDNIKHV**L**GGGSVQIVYK**EGGW**
R2R3-IEZip	----DDDNIKHVPGGGSVQIVYKKK
R2R3-IEZip-P301L	----DDDNIKHV**L**GGGSVQIVYKKK
R2R3-Cis[1]	------DNIKHV**P**GGGSVQIVYK
R2R3-Trans[1]	------DNIKHV**P**GGGSVQIVYK
R2R3-Ea[1]	------DNIKHV**P**GGGSVQIVYK
WT tau RD[2]	R1: $_{244}$ QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK $_{274}$ R2: $_{275}$ VQIINKKLDLSN-VQSK**C**GSKDNIKHVPGGGS $_{305}$ R3: $_{306}$ VQIVYKPVDLSK-VTSK**C**GSLGNIHHKPGGGQ $_{336}$ R4: $_{337}$ VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN $_{368}$ R′: $_{369}$ KKIETH $_{374}$
WT tau RD[3]	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[3]	R1: $_{244}$ QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK $_{274}$ R2: $_{275}$ VQIINKKLDLSN-VQSKCGSKDNIKHV**L**GGGS $_{305}$ R3: $_{306}$ VQIVYKPVDLSK-VTSKCGSLGNIHHKPGGGQ $_{336}$ R4: $_{337}$ VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN $_{368}$ R′: $_{369}$ KKIETHKLTFRE $_{380}$
P301S tau RD[3]	R1: $_{244}$ QTAPVPMPDLKN-VKSKIGSTENLKHQPGGGK $_{274}$ R2: $_{275}$ VQIINKKLDLSN-VQSKCGSKDNIKHV**S**GGGS $_{305}$ R3: $_{306}$ VQIVYKPVDLSK-VTSKCGSLGNIHHKPGGGQ $_{336}$ R4: $_{337}$ VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGN $_{368}$ R′: $_{369}$ KKIETHKLTFRE $_{380}$

WT tau[3]	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEEPG SETSDAKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTPSLEDEAAG HVTQARMVSKSKDGTGSDDKKAKGADGKTKIATPRGAAPPGQKGQANATRIPAKTPPAPK TPPSSGEPPKSGDRSGYSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAK SRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV PGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNI THVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV DSPQLATLADEVSASLAKQGLLEHHHHHH
P301L tau[3]	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEEPG SETSDAKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTPSLEDEAAG HVTQARMVSKSKDGTGSDDKKAKGADGKTKIATPRGAAPPGQKGQANATRIPAKTPPAPK TPPSSGEPPKSGDRSGYSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAK SRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV **L**GGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNI THVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV DSPQLATLADEVSASLAKQGLLEHHHHHH

[1]Fluorinated proline analogs were used to preferentially populate *cis, trans,* or lowered E_a barrier conformers at the positions indicated as **P**.
[2]Protein used in intramolecular FRET assay. Cysteines labeled are marked as **C**.
[3]Proteins used in ThT, seeding, and XL-MS experiments.

Table 2: List of AlzForum Disease-Associated Mutations

Name	Amino Acid Sequence
Tau RD AlzForum Mutations	R1: 244 QTAPVPMPDLKN-V**K**SK**I**GSTEN**L**KHQP**GG**GK 274 R2: 275 VQII**N**KKLD**LS**N-VQSKCGSKD**N**IKH**VP**GG**G**G**S** 305 R3: 306 VQIVYKPVD**LSK**-**V**TSKCGSLGNIHHK**P**GG**GQ** 336 R4: 337 **V**EVKSE**K**LDFKDRVQ**S**KIG**S**LDN**I**TH**VP**GG**G**GN 368

700 FIGURE LEGENDS

701

702 **Figure 1**. **Tauopathy mutations cluster to inter-repeat regions**. **a.** Disease-associated
703 mutation frequency found in human tauopathies. Most mutations are found within the repeat
704 domain (tau RD) (repeat 1 = red; repeat 2 = green; repeat 3 = blue; repeat 4 = purple).
705 Amyloidogenic sequence ^{306}VQIVYK^{311} is shown in the inset cartoon. **b.** Detailed mutation
706 frequencies found near the ^{306}VQIVYK^{311} amyloid motif. **c.** FL WT tau and mutant P301L tau at
707 a 4.4 μM concentration were mixed with stoichiometric amounts of heparin (4.4 μM), and allowed
708 to aggregate in the presence of ThT at room temperature. Control WT and P301L tau in the
709 absence of heparin yielded no detectible ThT signal change (less than 2-fold ratio to background
710 signal) over the course of the experiment (see Supplemental Table 1). ThT fluorescence was
711 normalized to the maximum for each condition. **d**. WT tau RD and mutant P301L and P301S tau
712 RD at a 4.4 μM concentration were each mixed with equimolar amounts of heparin (4.4 μM), and
713 allowed to aggregate in the presence of ThT at room temperature. Control WT, P301L and
714 P301S tau RD in the absence of heparin yielded no detectible ThT signal change (less than 2-
715 fold ratio to background signal) over the course of the experiment (see Supplemental Table 1).
716 **e**. WT FL tau and mutant P301L tau at a 4.4 μM concentration were mixed with substoichiometric
717 M_s tau seed (33nM) and allowed to aggregate in the presence of ThT at room temperature.
718 Control WT and P301L tau in the absence of M_s yielded no detectible ThT signal change (less
719 than 2-fold ratio to background signal) over the course of the experiment (see Supplemental
720 Table 1). All ThT experiments were carried out in triplicate. The data are shown as the average
721 with standard deviation and are colored according to mutation. **f-h**. After 120 hours of *in vitro*
722 incubation, proteins from previous ThT experiments were transduced into tau biosensor cells *via*
723 lipofectamine (**Methods**). FRET signal from each condition (tau RD-CFP/tau RD-YFP) was
724 measured by flow cytometry on 3 biological triplicates of at least 10,000 cells per condition. Error
725 bars represent a 95% CI of each condition.

726

727 **Figure 2. Tau RD encodes global and local structure. a**. Cartoon schematic of tau RD used
728 for FRET studies. For FRET studies, endogenous cysteines are labelled with Alexa-488 and
729 Alexa-647 (red and blue circles), which flank the ^{306}VQIVYK^{311} amyloid motif. **b-c**. FRET was
730 measured for dye modified monomeric WT and P301L tau RD at a series of temperatures over
731 a time course. **d**. Cartoon schematic of tau RD used for XL-MS studies colored according to
732 repeat domain. Recombinant WT and P301L tau RD were heated at 37ºC, 50ºC or 75ºC for one
733 hour, then chemically crosslinked using DSS. After cross-linking, trypsin fragmentation, and LC-
734 MS/MS analysis were performed. Each sample was carried out in 5 technical replicates. **e**. Total
735 consensus crosslinks parsed by temperature and location in WT and P301L tau RD: within N-
736 terminus (blue; residues 243-310; N-term), within C-terminus (orange; residues 311-380; C-
737 term) and span N- and C-terminus (magenta; between residues 243-310 and 311-380; N-C). **f-**
738 **h**. Consensus crosslinks (circles) are shown in contact maps color coded by average frequency
739 across replicates. The theoretical lysine pairs are shown in the background as gray circles.
740 Crosslink contacts within the N-term (blue), C-term (red) and across N- to C-term (purple) are
741 shown as sectors. The x- and y-axis are colored according to repeat number as in Figure 1. The
742 dashed boxes define inter-repeat crosslinks observed between repeat 2 and repeat 3. **i-k**. Same
743 as f-h above, except with tau RD that contains a P301L mutation.

744

745 **Figure 3. Wildtype and mutant peptides differentially populate collapsed and extended**
746 **conformations. a**. Conformations obtained for a trimer of WT peptide fragment (R2R3-WT) with
747 the sequence ^{295}DNIKHVPGGGSVQIVYK^{311}. Two-dimensional root-mean-squared-differences
748 (RMSD's) are calculated between all pairs of conformations visited during MD simulations.
749 Snapshots of trimeric structures are depicted for selected metastable basins, with each peptide
750 monomer represented by a different color. **b**. The same analysis as above, but for the P301L
751 substituted trimer. **c**. The free energy surface as a function of deviation from a canonical hairpin

752 structure. Two distinct basins, corresponding to collapsed and extended sub-ensembles, are
753 found in WT and P301L peptide fragment respectively.
754
755 **Figure 4. Tauopathy mutations drive aggregation propensity**. **a**. Schematic of tau RD and
756 the derived peptides representing the minimal structural element around ^{306}VQIVYK^{311}. **b**. WT
757 and mutant peptides were disaggregated, resuspended to 200 μM, and allowed to aggregate in
758 the presence of ThT at room temperature. WT R2R3 and R1R2 fragment peptides yielded no
759 detectible ThT signal change (less than 2-fold ratio to background signal) over the course of the
760 experiment (see Supplemental Table 1). ThT signals are shown as average of triplicates with
761 standard deviation, are colored according to mutation and are normalized to the maximum for
762 each condition.
763
764 **Figure 5. Peptides form amyloid structures and seed *in vivo*. a**. After 96 hours of *in vitro*
765 incubation, peptides from previous ThT experiments (**Figure 4 c**) were transduced into tau
766 biosensor cells *via* lipofectamine (**Methods**). FRET signal from each condition (tau RD-CFP/tau
767 RD-YFP) was measured by flow cytometry on 3 biological triplicates of at least 10,000 cells per
768 condition. Error bars represent a 95% CI of each condition. Solid and dashed horizontal lines
769 represent the mean and 95% error from untreated biosensor cells, respectively, for ease of
770 statistical comparison. **b-h**. Electron microscopy images of each peptide from previous ThT
771 experiments (**Figure 4 c**). The black bar represents 200 nm distance in each image **i-p**.
772 Qualitative fluorescence microscopy images of tau biosensor cells immediately prior to Flow
773 Cytometry experiments. i shows a representative image of untreated biosensor cells. j – p each
774 shows a representative image of biosensor cells treated with samples from b-h respectively.
775
776 **Figure 6. Alternative splicing modulates aggregation propensity. a**. Cartoon schematic for
777 tau 4R and 3R splice isoforms illustrate the difference in primary amino acid sequence leading

778 into the amyloidogenic ^{306}VQIVYK^{311} motif. **b**. A full combinatorial panel of R2R3-P301L and
779 R1R3-P301L chimeras were aggregated *in vitro*. ^{306}VQIVYK^{311} is shown in blue, amino acids
780 common between the splice isoforms are shown in gray, amino acids unique to an R3 isoform
781 are colored red, amino acids unique to an R4 isoform are colored green. The aggregation
782 kinetics, represented as $t_{1/2}$ in hours with 95% CI, are listed in the right-side column alongside
783 its respective peptide.
784
785 **Figure 7. Enhancing β-hairpin structure rescues spontaneous aggregation phenotypes.**
786 **a**. Cartoon schematic representation of the tryptophan zipper motif (green bar) and controls used
787 to stabilize a β-hairpin structure in an R2R3-P301L peptide fragment (**Table 1**). **b**. Aggregation
788 reactions of the tryptophan zipper peptide and controls measured by ThT fluorescence. The Trp-
789 R2R3-P301L-Trp fragment peptide yielded no detectible ThT signal change (less than 2-fold
790 ratio to background signal) over the course of the experiment (see Supplemental Table 1) ThT
791 signals are shown as average of triplicates with standard deviation and were normalized to the
792 maximum for each condition. **c**. Schematic of proline and fluorinated proline analogs used to
793 generate *cis* and *trans* proline conformers at the position corresponding to P301 (red circle) in
794 peptide models. **d**. ThT aggregation reactions of the *cis*, *trans*, and neutral proline analogs
795 substituted into the R2R3 peptide fragment. ThT signals are an average of 6 independent
796 experiments with standard deviation shown.
797
798 **Figure 8. Molecular model of tau amyloid domain structural rearrangement and**
799 **subsequent aggregation.** Naïve tau monomer (left) exists with a propensity to form a relatively
800 collapsed structure, which buries the amyloid domain ^{306}VQIVYK^{311}. In the presence of disease-
801 associated mutations, proline isomerization events, or certain splice isoforms, the equilibrium is

802 shifted to disfavor local compact structure. This exposes the aggregation prone ^{306}VQIVYK^{311}

803 amyloid motif and enhances aggregation propensity, leading to subsequent tau pathology.

804

References

1. Cleveland, D. W., Hwo, S. Y. & Kirschner, M. W. Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *Journal of Molecular Biology* **116,** 227–247 (1977).
2. Eliezer, D. *et al.* Residual structure in the repeat domain of tau: Echoes of microtubule binding and paired helical filament formation. *Biochemistry* **44,** 1026–1036 (2005).
3. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **56,** 343 (2017).
4. Sanders, D. W. *et al.* Distinct Tau Prion Strains Propagate in Cells and Mice and Define Different Tauopathies. *Neuron* **82,** 1271–1288 (2014).
5. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–457 (2007).
6. Bergen, von, M. *et al.* Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local beta-structure. *J. Biol. Chem.* **276,** 48165–48174 (2001).
7. Mukrasch, M. D. *et al.* Sites of tau important for aggregation populate {beta}-structure and bind to microtubules and polyanions. *Journal of Biological Chemistry* **280,** 24978–24986 (2005).
8. Falcon, B., Zhang, W., Murzin, A. G., Murshudov, G. & Holly, J. Structures of filaments from Pick ' s disease reveal a novel tau protein fold. (2018).
9. Zhang, X. *et al.* RNA Stores Tau Reversibly in Complex Coacervates. *bioRxiv* 111245 (2017). doi:10.1101/111245
10. Ismail, T. & Kanapathipillai, M. Effect of cellular polyanion mimetics on tau peptide aggregation. *J. Pept. Sci.* **24,** e3125 (2018).
11. Kuret, J. *et al.* Evaluating triggers and enhancers of tau fibrillization. *Microsc. Res. Tech.* **67,** 141–155 (2005).
12. Zhao, J. *et al.* Glycan Determinants of Heparin-Tau Interaction. *Biophys. J.* **112,** 921–932 (2017).
13. Kellogg, E. H. *et al.* Near-atomic model of microtubule-tau interactions. **1780,** 1–9 (2018).
14. Mok, S.-A. *et al.* Mapping interactions with the chaperone network reveals factors that protect against tau aggregation. *Nat Struct Mol Biol* **25,** 384–393 (2018).
15. Baughman, H. E. R., Clouser, A. F., Klevit, R. E. & Nath, A. HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. *J. Biol. Chem.* **293,** 2687–2700 (2018).
16. Mirbaha, H. *et al.* Inert and seed-competent tau monomers suggest structural origins of aggregation. *eLife* **7,** (2018).
17. Sharma, A. M., Thomas, T. L., Woodard, D. R., Kashmer, O. M. & Diamond, M. I. Tau monomer encodes strains. *eLife* **7,** (2018).
18. Elbaum-Garfinkle, S. & Rhoades, E. Identification of an aggregation-prone structure of tau. *J. Am. Chem. Soc.* **134,** 16607–16613 (2012).
19. Chirita, C. N., Congdon, E. E., Yin, H. & Kuret, J. Triggers of full-length tau aggregation: a role for partially folded intermediates. *Biochemistry* **44,** 5862–5872 (2005).
20. Walker, S., Ullman, O. & Stultz, C. M. Using intramolecular disulfide bonds in tau protein to deduce structural features of aggregation-resistant conformations. *J. Biol. Chem.* **287,** 9591–9600 (2012).
21. Kadavath, H. *et al.* Folding of the Tau Protein on Microtubules. *Angew. Chem. Int. Ed. Engl.* **54,** 10347–10351 (2015).
22. Mukrasch, M. D. *et al.* Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* **129,** 5235–5243 (2007).
23. Mukrasch, M. D. *et al.* Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol* **7,** e34 (2009).
24. Mutations. (2017).
25. Holmes, B. B. *et al.* Proteopathic tau seeding predicts tauopathy in vivo. *Proc Natl Acad Sci USA* **111,** E4376–E4385 (2014).

26. Yoshiyama, Y. *et al.* Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron* **53,** 337–351 (2007).
27. Fischer, D. *et al.* Structural and microtubule binding properties of tau mutants of frontotemporal dementias. *Biochemistry* **46,** 2574–2582 (2007).
28. Karamanos, T. K., Kalverda, A. P., Thompson, G. S. & Radford, S. E. Mechanisms of amyloid formation revealed by solution NMR. *Prog Nucl Magn Reson Spectrosc* **88-89,** 86–104 (2015).
29. Goedert, M., Wischik, C. M., Crowther, R. A., Walker, J. E. & Klug, A. Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau. *Proc Natl Acad Sci USA* **85,** 4051–4055 (1988).
30. Williams, D. R. *Tauopathies: Classification and clinical update on neurodegenerative diseases associated with microtubule-associated protein tau*. *Internal Medicine Journal* **36,** 652–660 (2006).
31. Schoch, K. M. M. *et al.* Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse Model. *Neuron* **90,** 941–947 (2016).
32. Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393,** 702–705 (1998).
33. Stöhr, J. *et al.* A 31-residue peptide induces aggregation of tau's microtubule-binding region in cells. *Nature Chemistry* **9,** 874–881 (2017).
34. Rizzu, P. *et al.* High Prevalence of Mutations in the Microtubule-Associated Protein Tau in a Population Study of Frontotemporal Dementia in the Netherlands. *The American Journal of Human Genetics* **64,** 414–421 (1999).
35. Mocanu, M.-M. *et al.* The potential for beta-structure in the repeat domain of tau protein determines aggregation, synaptic decay, neuronal loss, and coassembly with endogenous Tau in inducible mouse models of tauopathy. *J. Neurosci.* **28,** 737–748 (2008).
36. Furman, J. L., Holmes, B. B. & Diamond, M. I. Sensitive Detection of Proteopathic Seeding Activity with FRET Flow Cytometry. *J Vis Exp* e53205–e53205 (2015). doi:10.3791/53205
37. Leitner, A. *et al.* The Molecular Architecture of the Eukaryotic Chaperonin TRiC/CCT. *Structure/Folding and Design* **20,** 814–825 (2012).
38. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The Structural Basis of Substrate Recognition by the Eukaryotic Chaperonin TRiC/CCT. *Cell* **159,** 1042–1055 (2014).
39. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nat Meth* **5,** 315–318 (2008).
40. Ovchinnikov, S., Park, H., Kim, D. E., DiMaio, F. & Baker, D. Protein structure prediction using Rosetta in CASP12. *Proteins* **86 Suppl 1,** 113–121 (2018).
41. Lange, O. F. *et al.* Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 10873–10878 (2012).
42. Mylonas, E. *et al.* Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry* **47,** 10345–10353 (2008).
43. Lin, M. M. Leveraging symmetry to predict self-assembly of multiple polymers. *Chemical Physics Letters* **683,** 347–351 (2017).
44. Kaufman, S. K. *et al.* Tau Prion Strains Dictate Patterns of Cell Pathology, Progression Rate, and Regional Vulnerability In Vivo. *Neuron* **92,** 796–812 (2016).
45. Kier, B. L., Shu, I., Eidenschink, L. A. & Andersen, N. H. Stabilizing capping motif for beta-hairpins and sheets. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 10466–10471 (2010).
46. Vogel, M., Bukau, B. & Mayer, M. P. Allosteric regulation of Hsp70 chaperones by a proline switch. *Molecular Cell* **21,** 359–367 (2006).
47. Gustafson, C. L. *et al.* A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Molecular Cell* **66,** 447–457.e7 (2017).
48. Pastorino, L. *et al.* The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. *Nature* **440,** 528–534 (2006).
49. Torbeev, V. Y. & Hilvert, D. Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate 2-microglobulin amyloid assembly. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 20051–20056 (2013).

50. Lim, J. *et al.* Pin1 has opposite effects on wild-type and P301L tau stability and tauopathy. *Journal of Clinical Investigation* **118,** 1877–1889 (2008).
51. Mukrasch, M. D. *et al.* The 'jaws' of the tau-microtubule interaction. *Journal of Biological Chemistry* **282,** 12230–12239 (2007).
52. Meng, W., Lyle, N., Luan, B., Raleigh, D. P. & Pappu, R. V. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 2123–2128 (2013).
53. Ding, F., Jha, R. K. & Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure/Folding and Design* **13,** 1047–1054 (2005).
54. Kontsekova, E., Zilka, N., Kovacech, B., Skrabana, R. & Novak, M. Identification of structural determinants on tau protein essential for its pathological function: novel therapeutic target for tau immunotherapy in Alzheimer's disease. *Alzheimers Res Ther* **6,** 45 (2014).
55. Eichner, T. & Radford, S. E. A Generic Mechanism of B2-Microglobulin Amyloid Assembly at Neutral pH Involving a Specific Proline Switch. *Journal of Molecular Biology* **386,** 1312–1326 (2009).
56. Reumers, J., Maurer-Stroh, S., Schymkowitz, J. & Rousseau, F. D. Protein sequences encode safeguards against aggregation. *Human Mutation* **30,** 431–437 (2009).
57. Frost, B., Ollesch, J., Wille, H. & Diamond, M. I. Conformational diversity of wild-type Tau fibrils specified by templated conformation change. *Journal of Biological Chemistry* **284,** 3546–3551 (2009).
58. O'Nuallain, B. *et al.* Kinetics and Thermodynamics of Amyloid Assembly Using a High-Performance Liquid Chromatography-Based Sedimentation Assay. *Meth. Enzymol.* **413,** 34–74 (2006).
59. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Meth* **9,** 901–903 (2012).
60. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* **44,** 213–223 (2009).
61. Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327,** 1014–1018 (2010).
62. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters* **100,** (2008).
63. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958 (2010).
64. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).
65. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).
66. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20,** 786–798 (1999).
67. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).
68. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185,** 604–613 (2014).
69. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

Figure 1

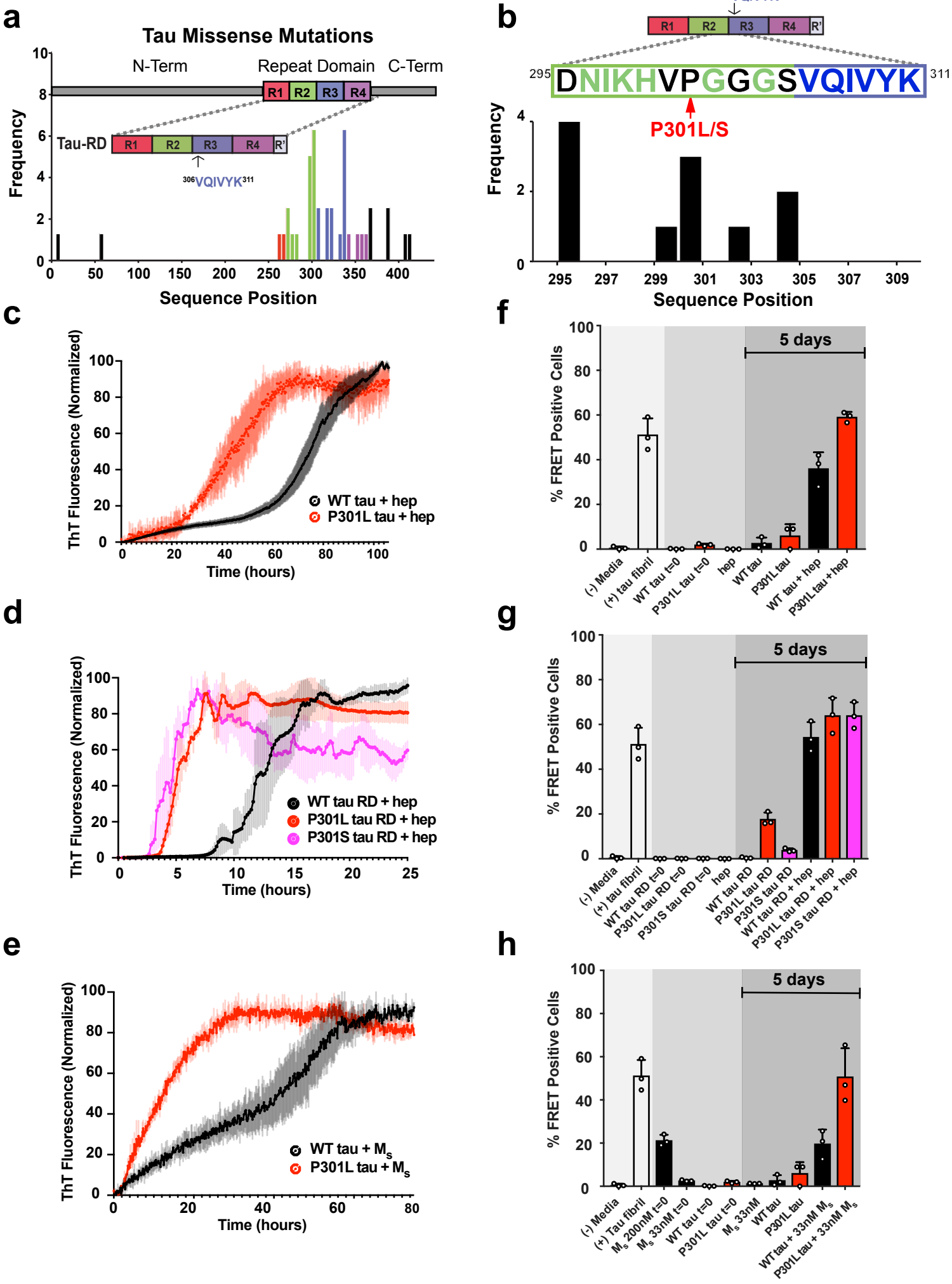

a
Tau Missense Mutations
N-Term
Repeat Domain
C-Term
R1
R2
R3
R4
Tau-RD
R1
R2
R3
R4
R'
306VQIVYK311
Frequency
Sequence Position
b
306VQIVYK311
R1
R2
R3
R4
R'
295
DNIKHVPGGGSVQIVYK
311
P301L/S
Frequency
Sequence Position
c
ThT Fluorescence (Normalized)
Time (hours)
WT tau + hep
P301L tau + hep
d
ThT Fluorescence (Normalized)
Time (hours)
WT tau RD + hep
P301L tau RD + hep
P301S tau RD + hep
e
ThT Fluorescence (Normalized)
Time (hours)
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 200nM t=0
Ms 33nM t=0
WT tau t=0
P301L tau t=0
Ms 33nM
WT tau
P301L tau
WT tau + 33nM Ms
P301L tau + 33nM Ms

Figure 2

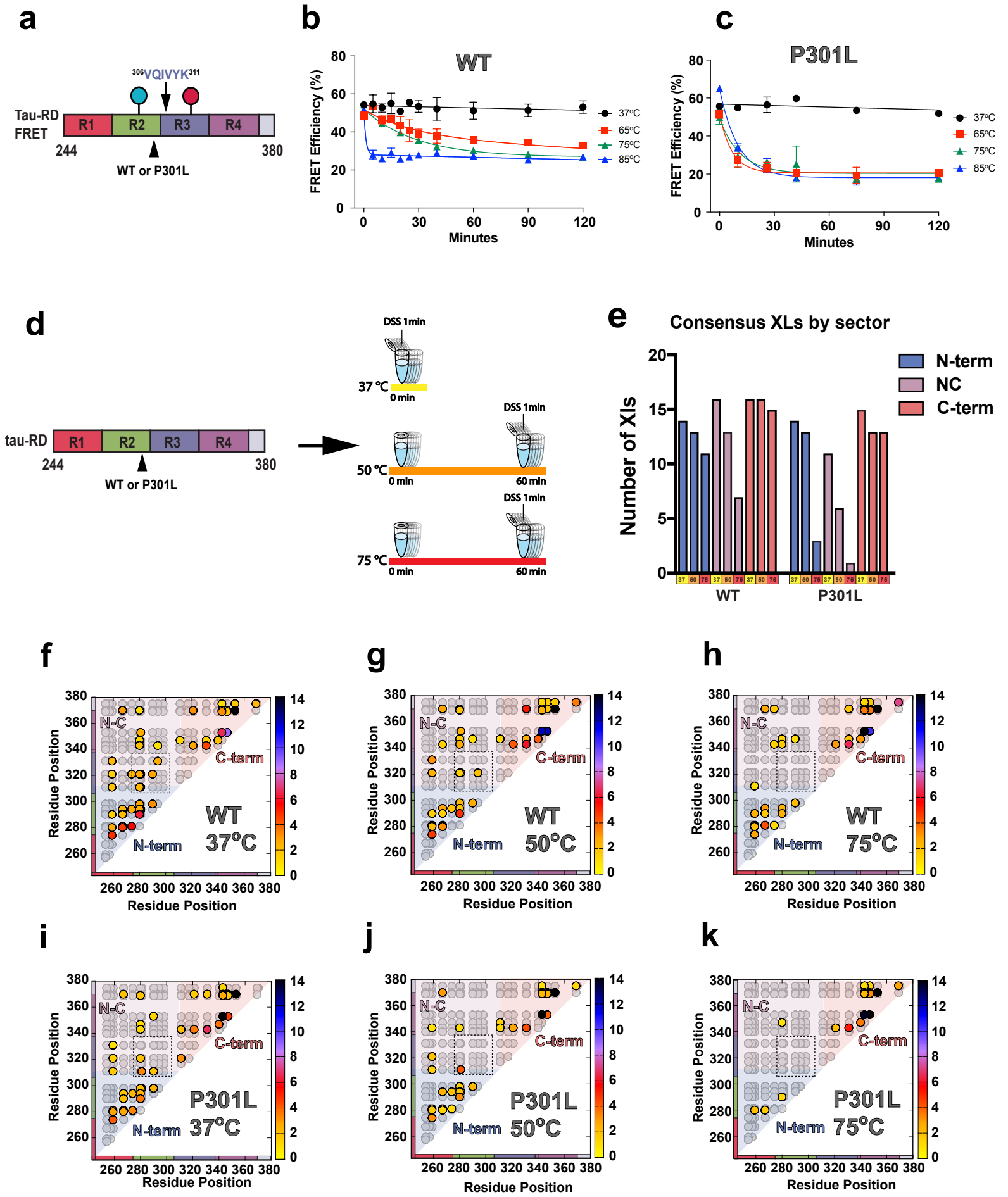

a
306VQIVYK311
Tau-RD FRET
R1
R2
R3
R4
244
380
WT or P301L
b
WT
FRET Efficiency (%)
Minutes
37ºC
65ºC
75ºC
85ºC
c
P301L
FRET Efficiency (%)
Minutes
37ºC
65ºC
75ºC
85ºC
d
tau-RD
R1
R2
R3
R4
244
380
WT or P301L
DSS 1min
37 ºC
0 min
DSS 1min
50 ºC
0 min
60 min
DSS 1min
75 ºC
0 min
60 min
e
Consensus XLs by sector
Number of XIs
N-term
NC
C-term
WT
P301L
f
WT 37ºC
g
WT 50ºC
h
WT 75ºC
i
P301L 37ºC
j
P301L 50ºC
k
P301L 75ºC
N-C
C-term
N-term
Residue Position

Figure 3

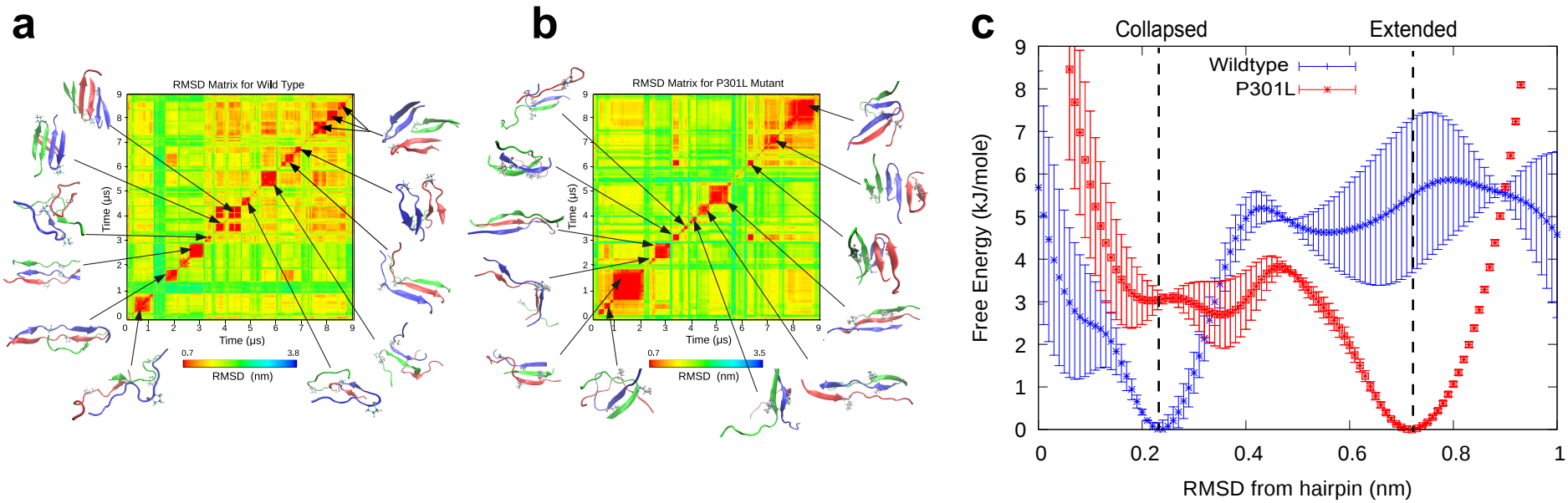

a
RMSD Matrix for Wild Type
Time (μs)
Time (μs)
0.7
3.8
RMSD (nm)
b
RMSD Matrix for P301L Mutant
Time (μs)
Time (μs)
0.7
3.5
RMSD (nm)
c
Collapsed
Extended
Wildtype
P301L
Free Energy (kJ/mole)
RMSD from hairpin (nm)

Figure 4

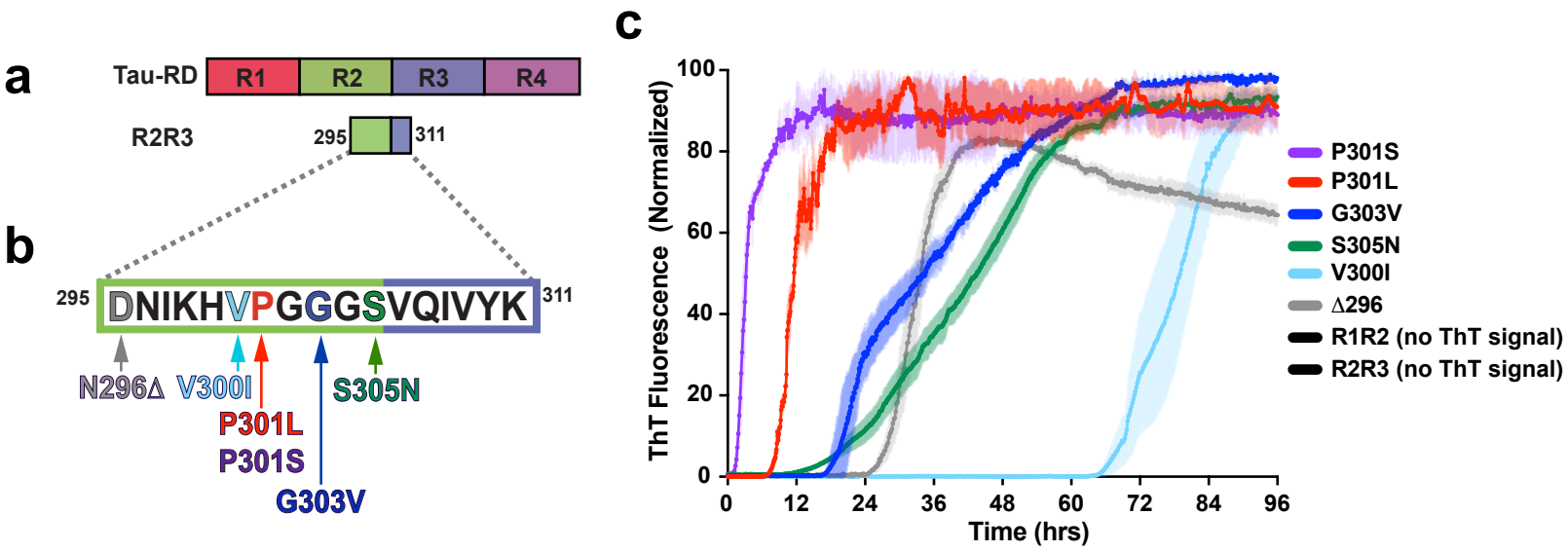

a
Tau-RD
R1
R2
R3
R4
R2R3
295
311
b
295
DNIKHVPGGGSVQIVYK
311
N296Δ
V300I
P301L
P301S
G303V
S305N
c
ThT Fluorescence (Normalized)
100
80
60
40
20
0
0
12
24
36
48
60
72
84
96
Time (hrs)
P301S
P301L
G303V
S305N
V300I
Δ296
R1R2 (no ThT signal)
R2R3 (no ThT signal)

Figure 5

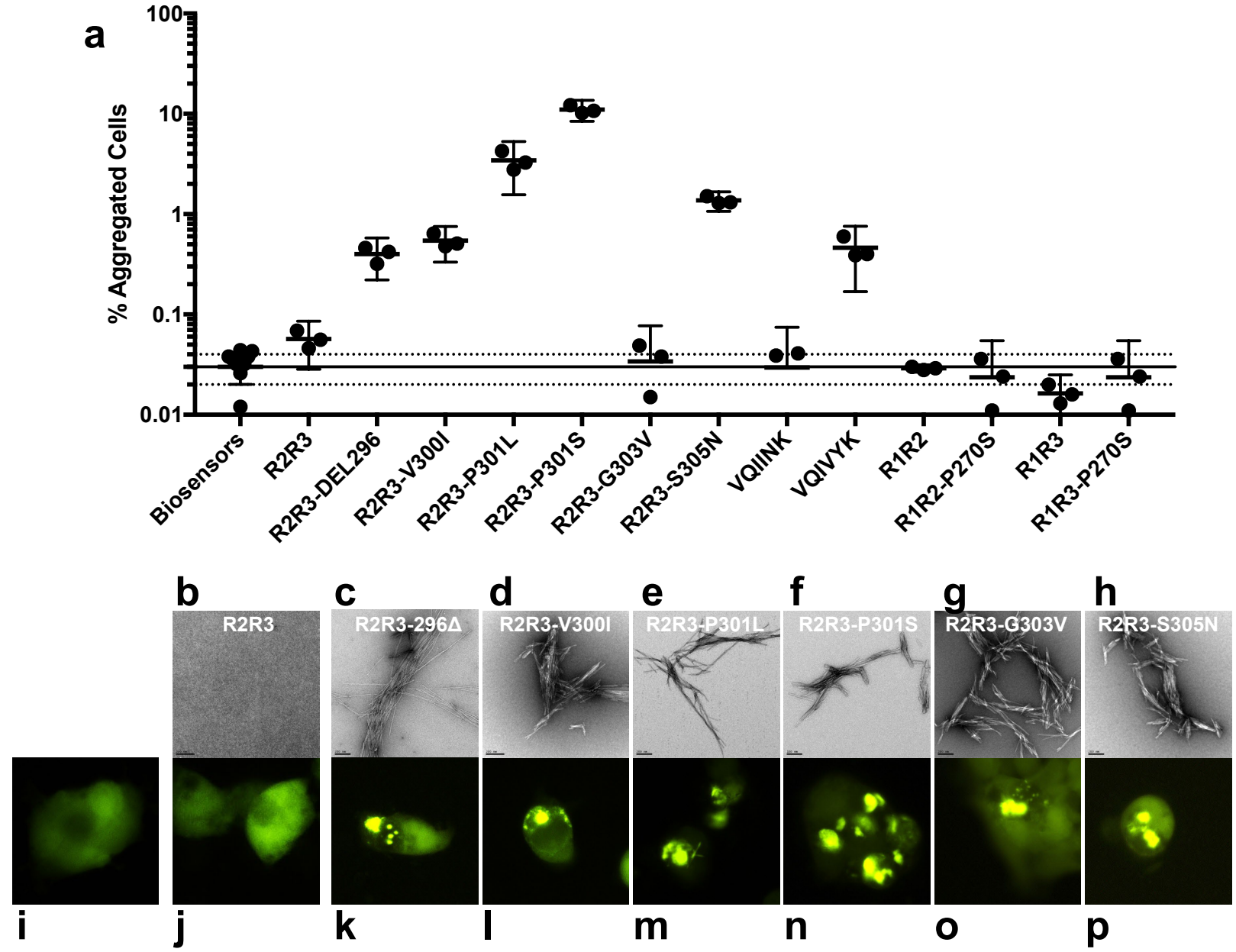

a
% Aggregated Cells
100
10
1
0.1
0.01
Biosensors
R2R3
R2R3-DEL296
R2R3-V300I
R2R3-P301L
R2R3-P301S
R2R3-G303V
R2R3-S305N
VQIINK
VQIVYK
R1R2
R1R2-P270S
R1R3
R1R3-P270S
b
c
d
e
f
g
h
R2R3
R2R3-296Δ
R2R3-V300I
R2R3-P301L
R2R3-P301S
R2R3-G303V
R2R3-S305N
i
j
k
l
m
n
o
p

Figure 6

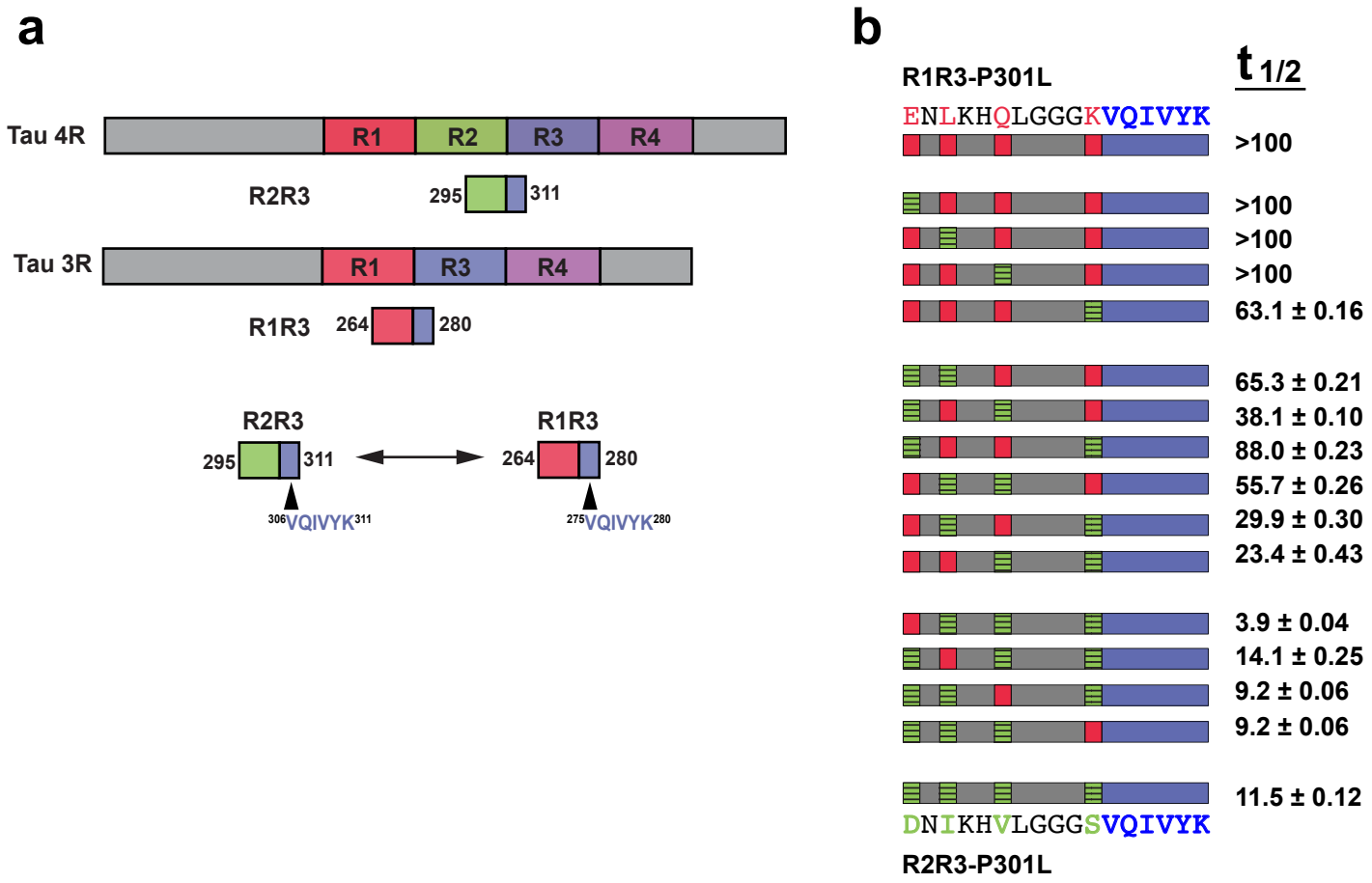

a
Tau 4R
R1
R2
R3
R4
R2R3
295
311
Tau 3R
R1
R3
R4
R1R3
264
280
R2R3
295
311
306VQIVYK311
R1R3
264
280
275VQIVYK280
b
R1R3-P301L
t 1/2
ENLKHQLGGGKVQIVYK
>100
>100
>100
>100
63.1 ± 0.16
65.3 ± 0.21
38.1 ± 0.10
88.0 ± 0.23
55.7 ± 0.26
29.9 ± 0.30
23.4 ± 0.43
3.9 ± 0.04
14.1 ± 0.25
9.2 ± 0.06
9.2 ± 0.06
11.5 ± 0.12
DNIKHVLGGGSVQIVYK
R2R3-P301L

Figure 7

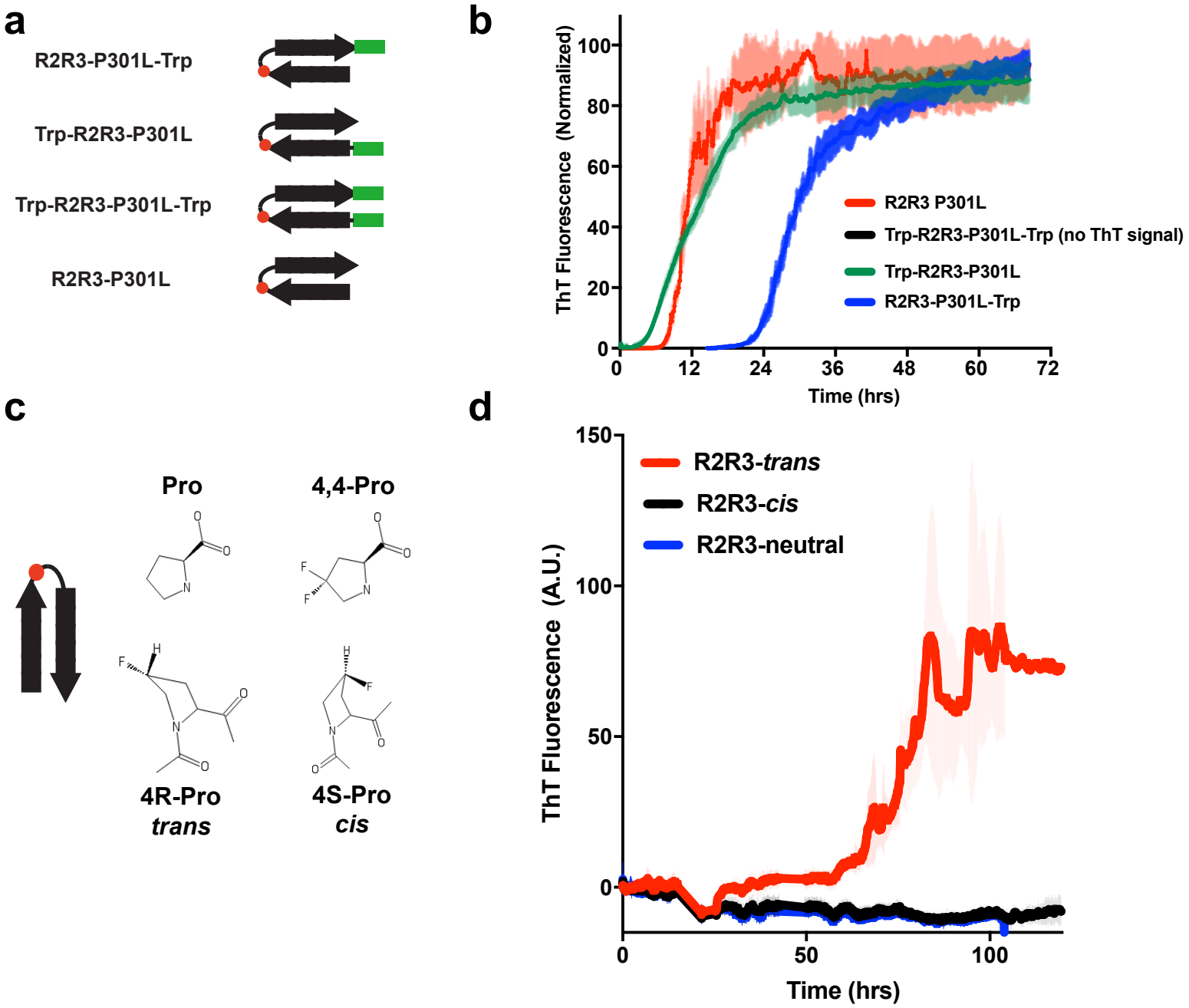
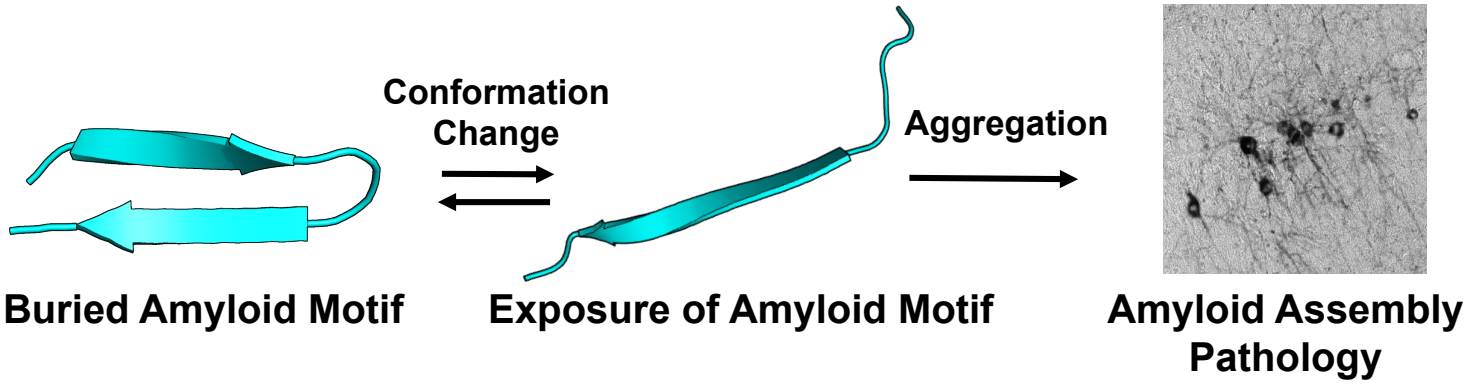

a
R2R3-P301L-Trp
Trp-R2R3-P301L
Trp-R2R3-P301L-Trp
R2R3-P301L
b
ThT Fluorescence (Normalized)
Time (hrs)
R2R3 P301L
Trp-R2R3-P301L-Trp (no ThT signal)
Trp-R2R3-P301L
R2R3-P301L-Trp
c
Pro
4,4-Pro
4R-Pro
trans
4S-Pro
cis
d
ThT Fluorescence (A.U.)
Time (hrs)
R2R3-trans
R2R3-cis
R2R3-neutral

Figure 8


Conformation
Change
Aggregation
Buried Amyloid Motif
Exposure of Amyloid Motif
Amyloid Assembly
Pathology