Native tau structure is disrupted by disease-associated mutations that promote aggregation

Kenneth W. Drombosky[1], Dailu Chen[1], DaNae Woodard[1], Levent Sari[3,4], Milo 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 75235, United States

[2] Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235, United States

[3] Green Center for Molecular, Computational, and Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75235, United States

[4] Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75235, United States

* 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 *MAPT* cause dominantly inherited tauopathies, but a biophysical mechanism driving its 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 crosslink mass spectrometry, intramolecular FRET, *in vitro* peptide systems, simulations, and cell models, we provide evidence that the aggregation prone ^{306}VQIVYK^{311} motif forms metastable compact structures with upstream sequences, which modulates aggregation propensity. Disease-associated mutations, isomerization of critical prolines, or alternative splicing are all sufficient to destabilize local structure and trigger spontaneous aggregation. These local structural rearrangements provide a biophysical framework supporting a model in which perturbations initiate early events in sporadic and genetic tau pathogenesis.

INTRODUCTION

Tauopathies belong to a class of over 20 neurodegenerative diseases in which the protein tau aggregates in neurons and glia. Tau aggregation correlates strongly with the degree of dementia and neurodegeneration, especially in Alzheimer's Disease. The structural mechanisms by which disease-associated mutations, alternative splicing, or other events promote aggregation and pathology is unknown. Understanding the molecular mechanisms of tau aggregation is critical for identifying therapeutic targets and diagnostic strategies.

The N-terminal ~200 and C-terminal ~80 residues of tau are largely disordered, rendering this system refractory to high-resolution structural biology methods[1]. In contrast, 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 strains[4]. Tau RD contains an amyloid motif (^{306}VQIVYK^{311}) (**Figure 1 A**), which is the basis of conversion between the soluble and insoluble states; ^{306}VQIVYK^{311}-mediated self-assembly is sufficient and necessary to form amyloid *in vitro*[5] and generate pathology *in vivo*[6], but its detailed mechanism of action is poorly understood.

Recent data from our group has indicated a switch occurs in soluble monomeric tau, which 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[7]. M_s has been purified from multiple sources: recombinant fibrils, Alzheimer's disease brain lysate, or brief exposure to heparin[7]. Recent solution nuclear magnetic resonance (NMR) methods mapped the heparin binding site to repeat 2, but how this interaction modulates tau conformation remains unclear[8]. Cryo-EM structures suggest an extended conformation of tau when bound to tubulin[9]. Building on this observation, double electron-electron resonance experiments indicated an expansion of this region upon heparin binding[10]. Other work mapping the recruitment of molecular chaperones with tau indicates many chaperones, including Hsp70, Hsp40, and Hsp90, localize around ^{306}VQIVYK^{311} [11]. Furthermore, expansion of tau RD appears to promote chaperone binding to the amyloid motif, suggesting that local conformational changes help recruit factors to suppress aggregation[12]. In contrast, intramolecular disulfide bridges in tau RD are predicted to form a local loop-like structure spanning repeats 2 and 3 that are incompatible with the formation of amyloid[13].

68 Further solution NMR data is available for a fragment of tau RD in complex microtubules[14]
69 where portions of the protein appear to form local contacts with upstream flanking sequences.
70 Structural models guided by experimental restraints from crosslinking mass spectrometry are consistent
71 with independent NMR data[14]: ^{306}VQIVYK^{311} has potential to form local structural elements.

72 We hypothesized that disease-associated mutations may contribute to tau's molecular
73 rearrangement from inert to an early seed-competent form. Of the missense mutations genetically
74 linked to tau pathology in humans, many disproportionally occur within tau RD and specifically cluster
75 near ^{306}VQIVYK^{311} [15] (**Figure 1 A, B**). Disease-associated mutations found near tau's amyloid motif,
76 such as P301L or P301S, have no definitive biophysical mechanism of action, but are nevertheless
77 widely used in cell and animal models[16,17]. These observations suggest that the local environment of
78 tau's amyloid motif influences aggregation propensity. A mechanistic understanding of how mutations in
79 tau drive pathology bears directly on how we might better develop conformation-specific diagnostics
80 and therapies.

81 As with disease-associated mutations, alternative splicing also impacts the primary sequence N-
82 terminal to ^{306}VQIVYK^{311}. Tau is expressed in the adult brain primarily as two major splice isoforms: 3-
83 repeat and 4-repeat[18]. In the truncated 3-repeat isoform, the second of four imperfectly repeated
84 segments in tau RD is absent. Expression of the 4-repeat isoform correlates with the deposition of
85 aggregated tau tangles in many tauopathies[19]; non-coding mutations that increase preferential splicing
86 or expression of the 4-repeat isoform are genetically linked to some tauopathies[19–21]. It is not obvious
87 why the incorporation or absence of the second repeat correlates with disease, as the primary
88 sequences, although imperfectly repeated, are relatively conserved.

89 Finally, previous reports have focused on studying intra-repeat interactions with the assumption
90 that each repeat functions independently within tau RD[22]. These reports have shown a relationship
91 between the length of an intra-repeat fragment, its propensity to spontaneously aggregate, and its
92 seeding capacity in cells[22]. However, inter-repeat interactions may also influence aggregation given that
93 both alternative splicing and many disease-associated mutations cluster around the repeat interfaces.
94 We hypothesize that wild-type tau aggregates less efficiently because flanking sequences shield
95 ^{306}VQIVYK^{311}, while disease-causing mutations, alternative splicing, or other factors destabilize local

96 structures around ^{306}VQIVYK^{311} allowing the transition of tau monomer to an aggregation-prone

97 conformation.

98

99 RESULTS

100 ***In Silico* modeling indicates tau RD local structure.** To initiate our studies, we first used
101 molecular modeling to gain insight into how putative hairpin structures in tau RD might behave. We built
102 predictive models using CS-ROSETTA, which implied metastable local β-hairpin structures across
103 repeat interfaces. We illustrate one of these interface fragments using a model built with CS-ROSETTA
104 for the R2-R3 ^{306}VQIVYK^{311}-containing interface (**Figure 1 C**). We focused on the ^{306}VQIVYK^{311}-
105 containing interface over others in tau RD for several reasons. The ^{306}VQIVYK^{311}-containing interface
106 has the highest frequency of proximal disease-associated mutations, particularly P301L and P301S,
107 which are commonly used as the genetic basis of disease modelling (**Figure 1 A**). Secondly, other
108 potential amyloid regions, such as ^{275}VQIINK^{280}, are nevertheless capable of aggregation
109 (**Supplemental Figure 1**), but are absent in recent cryo-EM structures of tau aggregates[3,23].

110 A standard fragment selection protocol in *ab initio* ROSETTA yielded similar local β-hairpin
111 structures of the same sequence, further suggesting that the amino acid composition near the amyloid
112 motif is compatible with beta-turns (**Figure 1 D**). The P-G-G-G sequences preceding ^{306}VQIVYK^{311}
113 found in tau RD are compatible with the formation of β-hairpins and are typically found in β-turn
114 secondary structures. Simulating 5,000 structures using each protocol for each of the four potential tau
115 hairpins led to structures consistent with published Rg measurements[24]. Ensemble-wide analysis of
116 each dataset showed an approximate ~25-59% propensity to form hairpin-like structures across repeat
117 interfaces centered on the P-G-G-G sequences (**Supplemental Figure 2**). Mapping known missense
118 mutations onto the *ab initio* β-hairpin structure (**Figure 1 E**), we hypothesized that this cluster of
119 disease-associated mutations could act through a single mechanism of action. Mutations that
120 destabilize this β-hairpin secondary structure may expose the amyloid motif ^{306}VQIVYK^{311}, allowing for
121 aggregation. This model is compatible with recent cryo-EM findings that indicate a disengagement of
122 ^{306}VQIVYK^{311}'s N-terminal flanking sequence in a fibril structure[3].

123 **P301L promotes extended forms of tau.** *In silico* modelling corroborated recent biochemical
124 findings[7] and suggested a minimal sequence necessary to form a collapsed structure around
125 ^{306}VQIVYK^{311}. To understand how these structures might self-assemble, we employed molecular

126 dynamics (MD) simulations of two tau fragments comprising the minimally structured fragment centered
127 around the R2 to R3 interface: R2R3-WT and R2R3-P301L (**Table 1**). To enable sufficient sampling of
128 oligomer structures, we employed an unbiased algorithm based on a recently-developed symmetry-
129 constraint approach[25]. The trimer conformations obtained in simulations are depicted on a root-mean-
130 square deviation (RMSD) matrix for both R2R3-WT (**Figure 2 A**) and for the R2R3-P301L mutant
131 (**Figure 2 B**). For R2R3-WT, we observed a dominant population of trimeric conformations composed of
132 primarily hairpins, while the P301L disease-associated mutation stabilized an extended form. The
133 energy basin for a R2R3-WT peptide hairpin was predicted to be 5 to 6 kJ/mol lower than an
134 extended/fibril state, whereas the R2R3-P301L peptide is 3 kJ/mol lower in a fibril state than a hairpin
135 (**Figure 2 C**). Additionally, the free energy surface suggested an energy barrier of approximately 5
136 kJ/mol to convert R2R3-WT from hairpin to fibril. That same barrier was less than 1 kJ/mole for R2R3-
137 P301L, predicting a faster rate of kinetic conversion between the two conformational states. Thus, MD
138 predicted that a P301L mutation would promote amyloid assembly by destabilizing monomeric hairpin
139 structures.

140 **Tau amyloid motif has metastable compact structure.** Cross-linking mass spectrometry (XL-
141 MS) defines contact points in proteins and protein complexes and can guide the determination of
142 structure for large protein complexes or transient protein-protein interactions[26,27]. Tau RD cross-linked
143 samples are confirmed by SDS-PAGE (**Supplemental Figure 3**) and monomers are purified by Ultra
144 Performance Liquid Chromatography (UPLC). Cross-linked samples are trypsin digested and analyzed
145 by mass spectrometry to yield intramolecular protein contact information (**Methods**). In each dataset,
146 the cross-links reported represent consensus data across five independent samples with a low false
147 discovery rate (**Methods**). XL-MS of recombinant tau RD revealed higher order contacts at 37 °C
148 (**Figure 3 D**) that decrease in frequency at 75 °C (**Figure 3 E**), consistent with heat denaturation. We
149 found a bimodal distribution of short and long-range contacts within multiple repeat domains, which are
150 consistent with compact local metastable structures.

151 In contrast, XL-MS of recombinant tau RD with a P301L missense mutation (tau RD-P301L)
152 revealed an increased susceptibility to heat denaturation. At 37 °C, the crosslinks found in tau RD-
153 P301L (**Figure 3 G**) are similar than wild type at the same temperature (**Supplemental Figure 4)**.

154 However, at 75 °C there were fewer crosslinks within tau RD-P301L compared to wildtype (**Figure 3 H**),
155 particularly within the N-terminal sector, which harbors the P301L perturbation.

156 To further examine the stability of the amyloid motif in tau, we labeled recombinant tau RD
157 monomer with complementary Fluorescence Resonance Energy Transfer (FRET) fluorophores
158 (**Methods**, **Table 1**) at the two endogenous cysteines in tau located at positions C291 and C322, which
159 serendipitously flank the amyloid motif ^{306}VQIVYK^{311} around the predicted β-hairpin structure (**Figure 3**
160 **A-C**). In a closed conformation, the two fluorophores should have relatively high FRET, while a
161 disordered, extended, or denatured chain should yield relatively low FRET signal. To test our system,
162 we first examined two previously identified monomer populations of FL WT tau: inert monomer (M_i) and
163 seed-competent monomer (M_s). M_i requires cofactors, such as heparin, to spontaneously aggregate *in*
164 *vitro*, while the M_s monomer, derived from tau fibrils or Alzheimer patient brain material, readily self-
165 assembles to form amyloid[7]. M_i and M_s were labeled at C291 and C322 and their steady-state FRET
166 intensities measured. M_i, which is predicted by XL-MS to be in a more closed conformation, showed
167 higher FRET intensities (E = 0.82 ± 0.08, 52.9 Å ± 4.5) compared to M_s (E = 0.46 ± 0.11, 67.7 Å ± 4.4)
168 (**Supplemental Figure 5**). Denaturing M_i and M_s at 95 °C overnight yielded similar FRET levels (68.9 Å
169 ± 1.8, 65.1 Å ± 1.2 respectively). Surprisingly, this region in M_s is denatured at room temperature,
170 suggesting an extended conformation of the R2R3 interface for M_s, versus a relatively compact in M_i.
171 Using wild type recombinant tau RD, a time course heat denaturation of tau monomer (M_i) between 65
172 °C to 95 °C gradually decreased FRET efficiency, consistent with denaturation of relatively stable local
173 structure (**Figure 3 F**). The predicted end-to-end distance of a denatured random chain 33-mer is
174 approximately 66 Å[28]. Consistent with the previous experiment, we observe M_i to be collapsed at room
175 temperature, and expand to a theoretical fully denatured structure with high temperature (**Figure 3 F**).

176 We next examined tau RD monomer, M_i, in the context of a P301L mutation, which sits between
177 the two FRET probes immediately upstream ^{306}VQIVYK^{311} (**Figure 3 B**). A P301L missense mutation is
178 associated with neurodegeneration in model systems[17,29] and is genetically linked to dementias in
179 human patients[30], though its biophysical mechanism of action is poorly understood. We found no
180 detectible difference between monomeric wildtype tau RD and tau RD-P301L FRET energies at room
181 temperature. Consistent with XL-MS data, tau RD-P301L rapidly loses FRET signal and local structure

182 with heat denaturation (**Figure 3 I**), suggesting that tau RD-P301L lacks the thermostability associated
183 with a wildtype tau RD.

184 **Tau amyloid formation is governed by flanking residues**. In tau RD, ^{306}VQIVYK^{311} is
185 sufficient and necessary to form an amyloid[5,6]. In solution, ^{306}VQIVYK^{311} hexapeptides aggregate
186 spontaneously and rapidly as measured by Thioflavin T (ThT) fluorescence intensity ($t_{1/2}$ < 1 hr,
187 **Supplemental Figure 1**) whereas the N-terminal sequence ^{295}DNIKHV^{300} does not aggregate.
188 Combining these sequences yields the minimal structural element spanning repeat 2 into repeat 3
189 (^{295}DNIKHVPGGGSVQIVYK^{311}, R2R3) predicted by *in silico* modelling (**Table 1, Figure 4 A**). This
190 sequence is also where numerous disease-associated mutations cluster. Indeed, WT peptide
191 fragments representing this motif did not aggregate readily, with no ThT detected up to 96 hrs (**Figure 4**
192 **C**). By contrast, disease-associated mutations (**Figure 4 B**) substituted into the R2R3 peptide fragment
193 were sufficient to generate spontaneous amyloid formation: R2R3-P301S ($t_{1/2}$ = 4.1 ± 1.3 hours), R2R3-
194 P301L ($t_{1/2}$ = 7.2 ± 0.2 hours), R2R3-N296Δ ($t_{1/2}$ = 31.9 ± 0.2 hours), R2R3-G303V ($t_{1/2}$ = 32.1 ± 0.7
195 hours), R2R3-S305N ($t_{1/2}$ = 41.2 ± 0.2 hours), and R2R3-V300I ($t_{1/2}$ = 77.8 ± 1.3 hours**, Figure 4 C**).
196 Each of these peptides was confirmed to form amyloid-like fibril morphologies by transmission electron
197 microscopy, except for the wildtype R2R3 peptide where no large structures were found (**Figure 5 B-**
198 **H**).

199 We next assessed whether these mutations perturb local structure. We performed circular
200 dichroism on both the R2R3-WT and R2R3-P301L peptides (**Figure 4 C**). Compared to R2R3-WT,
201 R2R3-P301L monomer had diminished 198nm signal consistent with more random coil and less β-
202 sheet. Deconvolution of CD spectra using BeStSel (beta structure selection) revealed predominantly
203 anti-parallel signal (**Table 3, Supplemental Figure 6**)[31]. Both peptides contained significant random
204 coil structure, which is expected for metastable structures, intrinsically disordered proteins, and short
205 peptides that cannot gain stability from tertiary structure. The compact structures identified are
206 consistent with intrinsically disordered proteins having residual structural preferences.

207 To test the structural compatibility of aggregates formed by *in vitro* tau models, we employed tau
208 biosensor HEK293 cells that stably express tau RD (P301S) fused to cyan and yellow fluorescent
209 proteins. These cells sensitively report a FRET signal (tau RD-CFP/tau RD-YFP) only when exposed to

210 tau amyloid seeds, and are unresponsive to aggregates formed by other proteins, such as huntingtin or
211 α-synuclein[32]. The tau biosensor cells responded to all disease-associated peptides that aggregated
212 spontaneously *in vitro*, but not to the wild-type R2R3 peptide (which did not aggregate *in vitro*) (**Figure**
213 **5 A**). Qualitatively, biosensor cells retained their diffuse tau localization when untreated or exposed to a
214 wild-type R2R3 peptide, but formed puncta when cultured with aggregated mutant peptides (**Figure 5 I-**
215 **P**). Interestingly, the biosensor cells responded to disease-associated mutant peptides with varying
216 degrees of sensitivity and created dissimilar aggregate morphologies. This is consistent with amyloid
217 structures that act as distinct templates, as we have previously observed to form the basis of tau prion-
218 like strains[4,33].

219 **Tau splice variants reveal differential aggregation propensity**. Tau is expressed in the adult
220 brain as 6 major splice isoform types that include either 3 or 4 repeated segments within RD (**Figure 6**
221 **A**). 3R lacks the second of four repeats. 4R tau correlates strongly with aggregation in most
222 tauopathies[19] and mutations that increase splicing of the 4R isoform correlate with disease[19,21]. We
223 examined whether this splice isoform affects the propensity of ^{306}VQIVYK^{311}-mediated aggregation due
224 to the R1 to R3 junction having a different sequence composition from R2 to R3. We constructed a
225 series of peptides to encompass the R1 to R3 repeat interface (R1R3, **Table 1**). This wild type
226 sequence peptide mimicking a 3R splice isoform (R1R3) did not spontaneously aggregate. Surprisingly,
227 an R1R3 peptide with a corresponding P301L mutation also did not aggregate (**Figure 6**). We
228 hypothesized that differences in amino acid composition between the two splice isoforms led to the
229 sensitivity to aggregate in the presence of disease-associated mutations.

230 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
231 amino acid positions. To test the stabilizing effect of the R1 leading sequence, we constructed 16
232 peptides with a P301L disease-associated mutation to represent every combinatorial sequence
233 between the two leading strands (**Figure 6 B**). In this way, we could identify which amino acid(s)
234 governed R1's stronger inhibitory effects. We identified a general trend where R2R3-P301L aggregates
235 in hours with zero or one R1 substitutions. With 2 R1 substitutions, R2R3-P301L aggregation was
236 delayed roughly an order of magnitude to tens of hours. With 3 R1 substitutions, R2R3-P301L
237 aggregation was further delayed to hundreds of hours. With all four R1 substitutions (R1R3-P301L), no

238 ThT signal was observed within a week (**Figure 6 B**). Thus, all four amino acids contributed to the
239 ability of the R1 leading sequence to delay ^{306}VQIVYK^{311}-mediated spontaneous aggregation in a 3R
240 splice isoform. This may explain the differential aggregation propensities of tau isoforms in human
241 pathology.

242 **Stabilizing β-hairpin structure blocks P301L-mediated aggregation.** Our model predicted
243 sequestration of the ^{306}VQIVYK^{311} motif in tau *via* local β-structure. To further test this, we hypothesized
244 that artificially enhancing the local β-hairpin structure by stabilizing the termini would promote a more
245 inert, closed conformation. Consequently, we modified the R2R3-P301L peptide with a tryptophan
246 zipper (Trp-R2R3-P301L-Trp, **Table 1**), which stabilizes a β-hairpin structure approximately -2.5 to -7
247 kJ/mol[34]. We confirmed a structural change using circular dichroism of Trp-R2R3-P301L-Trp, where we
248 found an increase in β-sheet signal at 198 nm compared to R2R3-P301L (**Supplemental Figure 6,**
249 **Table 3**). Subtracting the R2R3-P301L signal from the Trp-R2R3-P301L-Trp revealed a curve with
250 significantly increased signal at 198 nm, indicative of β-sheet formation. Consistent with our model, Trp-
251 R2R3-P301L-Trp does not spontaneously aggregate *in vitro* (**Figure 7 B**).

252 To ensure that this effect wasn't a result of adding bulky tryptophan residues, we constructed
253 control peptides that contain only the N-term (Trp-R2R3-P301L) or the C-term (R2R3-P301L-Trp)
254 portion of the tryptophan zipper sequence (**Figure 7 A**). Both half-sequence controls spontaneously
255 aggregated, implying that a tryptophan in either position is insufficient to block aggregation (**Figure 7**
256 **B**). Only a fully intact tryptophan zipper that stabilizes a β-hairpin conformation ameliorates aggregation
257 propensity. Alternative methods to stabilize a β-hairpin architecture, such as introducing isoelectric
258 interactions, also delayed aggregation: peptides containing additional aspartic acids on the N-terminus
259 and lysines on the C-terminus (R2R3-IEZip, **Table 1**) retarded R2R3-P301L aggregation over an order
260 of magnitude ($t_{1/2}$ = 8 hours to $t_{1/2}$ = 83 hours, **Supplemental Figure 7**).

261 To test this effect in cells, we generated biosensor HEK293 cells expressing tau RD with a
262 comparable tryptophan zipper sequence. These biosensors had a significantly diminished capacity to
263 be seeded; R2R3-P301S peptide aggregates triggered aggregation in 14.5 ± 1.6% of tau biosensor
264 cells, but only 0.3 ± 0.12% of the tryptophan zipper stabilized biosensor cells (**Supplemental Figure 8**).

265 **Proline 301 *cis-trans* isomerization modulates aggregation.** Many proteins in the cell utilize
266 proline isomerization as a molecular switch, such as heat shock protein activation[35] or cell cycle
267 regulation[36]. In some proteins, proline isomerization directly induces or mitigates aggregation into
268 amyloid[37–39]. Proline isomerization events in tau have been proposed to play a role in aggregation and
269 disease[38], but P301 isomerization has not been linked to tau aggregation and pathology. Serine or
270 leucine substitutions at P301 proximal to ^{306}VQIVYK^{311} drastically alter aggregation propensity. We
271 hypothesized that P301 plays a crucial role inducing a β-turn in a P-G-G-G motif, which mediates a
272 collapsed structure. We tested whether isomerization of P301 could influence spontaneous amyloid
273 formation. We constructed a series of R2R3 peptides with proline analogs that preferentially populate
274 either: (1) a *cis* rotamer (2S,4S)-fluoroproline; (2) a *trans* rotamer (2S,4R)-fluoroproline; or (3) a proline
275 analog that easily interconverts between *cis* and *trans* (4,4)-difluoroproline (**Table 1, Figure 7 C**). Only
276 R2R3-Trans spontaneously aggregated (**Figure 7 D**), indicating the potential for proline isomerization
277 events in tau pathogenesis.
278
279

DISCUSSION

Here we establish the molecular and functional basis for how a series of prominent tau mutations drive spontaneous aggregation. Using intramolecular FRET and XL-MS, we identified local structure within the inter-repeat junctions of tau RD. Local compact structure in the R2-R3 junction, which encompasses the amyloidogenic ^{306}VQIVYK^{311} motif, is perturbed by disease-associated mutations, such a P301L. Using peptide model systems, we demonstrate that splice isoforms of tau are functionally dissimilar in their capacity to spontaneously aggregate with missense mutations. Furthermore, these peptide models indicate a potential role of proline isomerization in tau aggregation at the critical P301 position. Conversely, stabilizing local structure with tryptophan zippers or other means drastically limits the aggregation *in vitro* and *in vivo*. Taken together, our work provides a biophysical framework for how pathogenic mutations, splice isoforms, and other perturbations can drive early events in tau aggregation.

Our study has suggested that local structure encompassing the amyloid motif ^{306}VQIVYK^{311} regulates aggregation of the protein. Modeling of tau RD by *ab initio* or CS-ROSETTA indicated that repeat interfaces encode local structure that is compatible with a β-hairpin. Importantly, a cluster of disease-associated mutations localized to the repeat2-repeat3 interface suggested a possible mechanism. To explore the energetic landscape of local compact structure proximal to ^{306}VQIVYK^{311}, we employed molecular dynamics simulations. A WT peptide encoding the ^{306}VQIVYK^{311}-containing repeat 2/3 interface preferentially populated a hairpin conformation with a relatively low frequency of extended fibril-like structures. Conversely, a disease-associated mutation, P301L, dramatically shifted the equilibrium away from monomeric hairpins to fibril-like ensembles. To test the extent of structure formed in tau RD, we utilized orthogonal biophysical approaches. First, XL-MS revealed contacts spanning the local repeats, and long-range contacts within tau RD. Heat denaturation experiments indicated a marked drop in long-range contacts, consistent with the unfolding of secondary and tertiary structures. Concurrently, intramolecular FRET studies of tau RD experimentally followed unfolding of the ^{306}VQIVYK^{311} repeat 2/3 interface in real time by measuring distances between fluorescent probes that flank ^{306}VQIVYK^{311}. Under native conditions the local structure was collapsed, but expanded to random coil with heat denaturation. Conversely, tau RD with a P301L substitution was less

308 thermostable and rapidly lost FRET signal. This suggests that P301L is more susceptible to
309 conformational changes that expose the ^{306}VQIVYK^{311} amyloid motif. No structural differences were
310 seen between wildtype and tau RD-P301L at non-denaturing temperatures. While these differences are
311 subtle, we observe that P301L-mediated structural rearrangement only manifest under moderate stress
312 conditions (i.e. heat). This may explain the elusiveness of a biophysical basis of this cluster of
313 pathogenic mutations.

314 We further examined whether these structural perturbations influenced aggregation propensity.
315 WT tau fragments containing ^{306}VQIVYK^{311} did not aggregate spontaneously. Single point substitutions
316 of 6 disease-associated mutations immediately N-terminal to ^{306}VQIVYK^{311} consistently induced
317 spontaneous aggregation propensity *in vitro*. The fragments which aggregated *in vitro* formed classical
318 amyloid-like fibrils by EM and seeded soluble tau RD aggregation in tau biosensor cells. Alternative
319 splicing around the ^{306}VQIVYK^{311} sequence modulated spontaneous aggregation, and suggested why
320 3R tau aggregates less readily than 4R. To further probe the effect of local structure on aggregation
321 propensity, we stabilized a β-hairpin structure using a tryptophan zipper or isoelectric forces. After
322 deconvolution of the CD signal by BeStSel, we identified that essentially all the β-sheet signal from the
323 peptides tested was anti-parallel, consistent with β-hairpin structure. The magnitude of these changes
324 are modest as detected by CD; however, the functional consequences are significant.

325 Disease-associated mutations found near tau's amyloid motif, such as P301L or P301S have no
326 definitive biophysical mechanism but are nevertheless widely used in cell and animal models[16,17].
327 Mutations are generally categorized into splice or missense mutants. Two of the pro-aggregation
328 missense mutations studied (N296Δ, S305N) are reported in the literature as splice mutants with no
329 notable effects on protein aggregation or microtubule binding[40,41]. Using our peptide model system, we
330 observed that these missense mutations spontaneously aggregate *in vitro* and seed tau RD in cell
331 models. We therefore provide a rationale for the toxic-gain-of-function for several mutations and expand
332 on the previously reported splice effects of N296Δ and S305N. This peptide model system compliments
333 *in vivo* data by disentangling the toxic effects of mutations exerted through either mRNA or protein-
334 based mechanisms.

335 Studying the missense mutations in tau has generated valuable disease models[17,29]; however,
336 the majority of human tauopathies have no observed genetic mutation in *tau*[30]. Critical proline residues
337 N-terminal to the amyloid motif can isomerize into *cis* or *trans* rotamers spontaneously or through
338 unidentified cellular mechanisms. We observe that P301 *cis* and *trans* rotamers have distinct
339 aggregation propensities *in vitro*. In fact, the aggregation kinetics for a *trans* rotamer of P301 are on par
340 with some disease mutants (N296Δ, V300I). The concept of proline isomerization triggering
341 aggregation into amyloid is not entirely novel, as this is an accepted mechanism of β2-microglobulin
342 aggregation in kidney dialysis amyloidosis[42]. Other proline residues outside of tau repeat domain are
343 proposed to undergo proline isomerization[38]. We postulate a new mechanism whereby WT tau
344 aggregation could be controlled *in vivo*: specific prolyl isomerization events – possibly triggered by
345 cellular proline isomerases – could trigger spontaneously aggregation by modulating inter-repeat
346 structural elements.

347 Previous reports have been focused on studying intra-repeat interactions, with the assumption
348 that each repeat functions independently within tau RD. Peptide models have shown a relationship
349 between the length of a peptide fragment and the seeding capacity of tau[22]. Whereas previous work
350 defined the minimal sequence necessary to act as a fully functional seed, our model defines a minimal
351 sequence necessary for inhibiting aggregation. This work suggests that inter-repeat contacts play a
352 crucial role as structural elements that modulate aggregation propensity. The composition of these
353 inter-repeat sequences, governed by alternative splicing or missense mutations, directly impacts
354 stability of local structures and aggregation propensity. It is tempting to speculate that local structure
355 surrounding each of the four inter-repeat regions play independent roles in the exposure of amyloid
356 sequences. This modular nature of the tau-RD region may explain how these independent regions can
357 lead to different tau assemblies. A more comprehensive structure-function analysis of other sites may
358 help explain how each repeat contributes to the formation of different tau structures.

359 The expression levels of the two major isoform types of tau in the CNS – 3R and 4R – are
360 similar in the adult brain[18]. However, the 3R:4R ratio of aggregate deposits is disproportionally shifted
361 towards 4R in most tauopathies[19]. Mutations in tau that affect alternative splicing and generate excess
362 4R isoforms correlate with some genetic tauopathies[20,21]. The N-terminal flanking sequences leading

363 into ^{306}VQIVYK^{311} differ by four amino acids between the two isoforms. We find that these two isoforms
364 have drastically different aggregation propensities in the presence of disease-associated mutations ($t_{1/2}$
365 = 7 hours vs $t_{1/2}$ > 200 hours, respectively). Chimeras of R1R3 / R2R3 transition from aggregation-
366 resistant to aggregation-prone as they lose R1 N-terminal flanking character. The ability of an R1
367 leading strand to mitigate ^{306}VQIVYK^{311} aggregation may explain why 4-repeat tau correlates more
368 closely with pathology. Thus, inter-repeat contacts may explain aggregation propensities of tau isoforms
369 in disease. Encouraging data for a tau vaccine targeting a ^{300}HXPGGG^{304} sequence suggests it's
370 possible to utilize inter-repeat regions to select between pathogenic and non-pathogenic conformations
371 of tau[43].

372 Taking these observations together, we propose that sequences N-terminal to tau's amyloid
373 motif forms local contacts consistent with a β-hairpin-like compact structure, which shields the amyloid
374 motif and mitigates aggregation (**Figure 8**). This represents a simple model of tau aggregation that
375 unifies key observations throughout tau literature.

376 Algorithms that identify potential amyloid nucleating regions, such as TANGO, have indicated
377 that nearly 75% of aggregation nucleating regions in the human proteome use two or more
378 "gatekeeper" residues, with proline being the most common single gatekeeping residue[44]. These
379 gatekeeping residues are more likely than average to be the site of disease-associated missense
380 mutations. These observations tie in closely with our findings of gatekeeping residues near tau's
381 amyloid motif, and suggest that local flanking sequences and their structural contacts may play an
382 important role in mitigating aggregation propensity.

383 Finally, the identification and characterization of metastable compact structures near
384 ^{306}VQIVYK^{311} may itself prove to be a valuable therapeutic target. One might be able to shift the
385 structural rearrangement of tau amyloid motif from exposed (aggregation prone) to buried (inert) using
386 small molecules, antibodies, or cellular co-factors. Our results indicate that subtle changes in local
387 structure have immense functional ramifications; therefore, small molecules that shift this structural
388 equilibrium modestly may have significant benefits.

389

390 METHODS

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

394 **ThT Fluorescence Assays**. Peptides were disaggregated as previously described[45], lyophilized, and
395 resuspended in 2x PBS (273 mM NaCl, 5.4 mM KCl, 20 mM Na_2HPO_4, 3.6 mM KH_2PO_4, pH 7.4). 25
396 mM ThT was added to 200uL of 200 μM peptide in a 96-well clear bottom plate. ThT kinetic scans were
397 run every 5 minutes on a Tecan M1000 plate reader at 446 nm Ex (5 nm bandwidth), 482 nm Em (5 nm
398 bandwidth). Blank wells containing buffer and ThT were subtracted from experimental values.

399 **Tau intramolecular FRET.** Tau RD was expressed and purified as previously described[46]. 10-fold
400 excess TCEP was added to 100 μM Tau RD, the atmosphere was deoxygenated with N_2 gas, and
401 incubated at RT for 30 minutes. To label, C_5 maleimide Alexa-488 and C_2 maleimide Alexa-647 were
402 dissolved in DMSO and added in 5-fold excess to Tau RD, and incubated at 4° C overnight. Alexa-
403 labelled Tau RD was purified using an 8 kDa cut-off mini dialysis exchanged into 1x PBS overnight.
404 FRET efficiencies were calculated as a function of acceptor emission, given $E = (I_{AD}\varepsilon_{AA}-I_{AA}\varepsilon_{AD})/I_{AA}\varepsilon_{DD}$
405 where I_{AA} is Acceptor intensity following Acceptor excitation, I_{AD} is Acceptor intensity following Donor
406 excitation, and ε is the extinction coefficient of the fluorophores at given excitations and emissions.
407 Distances in angstroms were derived from FRET energies as a function of $Å = (R_{<0>})[(1-E)/E]^{1/6}$ $R_{<0>}$ is
408 a known value of 56 Å for the FRET pair Alexa-488 and Alexa-647, and E is the calculated energy
409 above. Alexa 488 Ex/Em was measured at 490 nm/520 nm, Alexa 647 Ex/Em 647 nm/670 nm with a 5
410 nm bandwidth for both.

411 **Model generation of tau RD using ROSETTA.** The backbone NH, N, CA, CB and C=O chemical shift
412 assignments for the tau fragment from 211-324 (generously provided by Juan Lopez and Guy Lippens)
413 were used in CS-Rosetta to generate fragment libraries for subsequent model refinement. First,
414 chemical shift parameters were used to predict backbone torsional angles using TALOS to generate a
415 CS-guided fragment library representing the conformations of the protein[47]. For the *ab initio* ROSETTA
416 calculations, the tau RD sequence was used to generate 3-mer and 9-mer fragments derived from the

417 protein data bank. The Rosetta energy function was used to assemble and iteratively refine 5000
418 structural models using each set of fragments[48,49]. Radius of gyration for each model was computed by
419 a python script. Ensemble wide calculation of cα-cα end to end distances between residues 264-280,
420 295-311, 327-343 and 359-375 were carried out using a python script. All simulations were done on
421 UTSW's biohpc computing cluster. All plots were generated with gnuplot. Images were created using
422 Pymol.

423 **Circular Dichroism**. Far-UV CD measurements were performed on a JASCO J-815 spectropolarimeter
424 using a 1-mm path length cuvette. CD samples were prepared in parallel to ThT assays and diluted to
425 0.1 mg/ml in 20 μM NaF. CD spectra were analyzed using BeStSel and CONTIN.

426 **Tau Biosensor Cells**. Biosensor cells were plated into 96-well plates at 20,000 cells per well. 10 μg of
427 aggregated peptide material was added to 0.5 uL lipofectamine and Opti-MEM to a total volume of 10
428 μL, incubated at room temperature for 30 minutes, and added directly to cell media. After 3 days, cells
429 were harvested with 0.05% trypsin, then resuspended in Flow buffer (1x HBSS, 1% FBS, 1 mM EDTA,
430 1x DPBS) and analyzed by flow cytometry.

431 **Flow Cytometry**. A BD LSRFortessa was used to perform FRET flow cytometry. To measure CFP and
432 FRET, cells were excited with the 405 nm laser, and fluorescence was captured with a 405/50 nm and
433 525/50 nm filter, respectively. To measure YFP, cells were excited with a 488 laser and fluorescence
434 was captured with a 525/50 nm filter. To quantify FRET, we used a gating strategy where CFP bleed-
435 through into the YFP and FRET channels was compensated using FlowJo analysis software. We then
436 created a plot of FRET vs. CFP and introduced a triangular gate to assess the number of FRET-positive
437 cells, as previously described[16]. For each experiment, 20,000 cells per replicate were analyzed. Data
438 analysis was performed using FlowJo v10 software (Treestar).

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

444 **Crosslinking, sample processing and LC-MS/MS analysis.** Preparation of tau RD was crosslinked
445 at a total protein concentration of 1.0 mg/mL using 100 µg of starting material. The crosslinking buffer
446 was 1X PBS and 1mM DTT. For each temperature (37°C, 75°C, 85°C and 95°C) five replicates
447 samples were prepared and equilibrated at the appropriate temperature for 30 minutes. The
448 crosslinking reaction was initiated by adding disuccinimidyl suberate (DSS) stock solution (25 mM DSS-
449 d_0 and –d_{12}, Creative Molecules) in DMF to a final concentration of 1 mM. Samples were further
450 incubated at 37°C, 75°C, 85°C or 95°C for 1 min with 350 RPM shaking. Excess reagent was quenched
451 by addition of ammonium hydrogen carbonate to 50 mM and incubation at 37°C for 30 min, and
452 subsequently flash frozen at -80°C and evaporated to dryness by lyophilization. Proteins were
453 resuspended in 8M urea, reduced with 2.5mM TCEP (37°C, 30 min) and alkylated with 5mM
454 iodoacetamide (30min, room temperature, protected from light). The sample solutions were diluted to 1
455 M urea with 50 mM ammonium hydrogen carbonate and trypsin (Promega) was added at an enzyme-
456 to-substrate ratio of 1:50. Proteolysis was carried out at 37°C overnight followed by acidification with
457 formic acid to 2% (v/v). Samples were then purified by solid-phase extraction using Sep-Pak tC18
458 cartridges (Waters) according to standard protocols. Samples were evaporated to dryness and
459 reconstituted in water/acetonitrile/formic acid (95:5:0.1, v/v/v) to a final concentration of approximately
460 0.5 µg/µl. 2µL each were injected for duplicate LC-MS/MS analyses on an Eksigent 1D-NanoLC-Ultra
461 HPLC system coupled to a Thermo Orbitrap Fusion Tribrid system. Peptides were separated on self-
462 packed New Objective PicoFrit columns (11cm x 0.075mm I.D.) containing Magic C_{18} material
463 (Michrom, 3µm particle size, 200Å pore size) at a flow rate of 300nL/min using the following gradient. 0-
464 5min = 5 %B, 5-95min = 5-35 %B, 95-97min = 35-95 %B and 97-107min = 95 %B, where A =
465 (water/acetonitrile/formic acid, 97:3:0.1) and B = (acetonitrile/water/formic acid, 97:3:0.1). The mass
466 spectrometer was operated in data-dependent mode by selecting the five most abundant precursor ions
467 (m/z 350-1600, charge state 3+ and above) from a preview scan and subjecting them to collision-
468 induced dissociation (normalized collision energy = 35%, 30ms activation). Fragment ions were
469 detected at low resolution in the linear ion trap. Dynamic exclusion was enabled (repeat count 1,
470 exclusion duration 30sec).

471 **Analysis of mass spectrometry data.** Thermo .raw files were converted to the open .mzXML format
472 using msconvert (proteowizard.sourceforge.net) and analyzed using an in-house version of xQuest[50].
473 Spectral pairs with a precursor mass difference of 12.075321 Da were extracted and searched against
474 the respective FASTA databases containing Tau (TAU_HUMAN P10636-8) or with a P301L
475 substitution. xQuest settings were as follows: Maximum number of missed cleavages (excluding the
476 crosslinking site) = 2, peptide length = 5-50 aa, fixed modifications = carbamidomethyl-Cys (mass shift
477 = 57.021460 Da), mass shift of the light crosslinker = 138.068080 Da, mass shift of mono-links =
478 156.078644 and 155.096428 Da, MS^1 tolerance = 10 ppm, MS^2 tolerance = 0.2 Da for common ions
479 and 0.3 Da for crosslink ions, search in ion-tag mode. Post-search manual validation and filtering was
480 performed using the following criteria: xQuest score > 18, mass error between -2.2 and +3.8ppm, %TIC
481 > 10, and a minimum peptide length of six aa. In addition, at least four assigned fragment ions (or at
482 least three contiguous fragments) were required on each of the two peptides in a crosslink. False
483 discovery rates (FDR's) for the identified crosslinks were estimated using xprophet[51] and estimated to
484 be 1.4-3.3% (**Supplemental Figure 9, Supplemental Table 1**). At each temperature, the 5 replicate
485 datasets were compared and only crosslinks present in 5 of the 5 datasets were used to generate a
486 consensus dataset (**Supplemental Table 2**). Crosslink data was visualized using Xvis[52]. Histograms
487 illustrating the contact order for the consensus crosslink sets were generated using gnuplot.
488 **MD Simulations.** Well-Tempered Metadynamics[53] was employed to enable accelerated conformational
489 sampling and to construct the associated free energy surface. Metadynamics was performed on a two-
490 dimensional space of parallel-beta sheet content and anti-parallel sheet content. To increase search
491 efficiency in oligomeric space, we have incorporated conformational symmetry constraints, which have
492 been shown to enable sampling of multi-polymer landscapes[25]. The initial dodecahedron simulation box
493 was constructed from a trimer of a randomly unfolded structure of 295-311 by adding 7587 SPCE
494 explicit waters and 3 neutralizing Cl ions (one for each monomer). The AMBER99sb-ildn force-field[54]
495 was used for all simulations. After an initial 1009 steepest descent steps of converged energy
496 minimization, 10 ns of NVT and 20 ns of NPT (first 10 with Berendsen[55] and the last 10 with Parrinello-
497 Rahman[56] barostats) equilibrations were performed. The subsequent production level trajectories are
498 based on 5 fs time steps using hydrogen-only virtual sites[57]. Production level trajectories were obtained

499 for an NPT ensemble with Parrinello-Rahman barostat, and periodic boundary conditions with Particle
500 Mesh Ewald (PME)[58] summation for long-range electrostatics. The tuned well-tempered metadynamics
501 parameters are 10, 1.4 kJ/mole, and 0.3 for bias factor, Gaussian height, collective variable space
502 Gaussian widths, respectively. The Gaussian perturbations were included into MD every 2.5 ps using
503 the PLUMED package[59] as an external patch to Gromacs-5.0.4[60]. A total of 18 μs trajectories were
504 generated, 9 μs for wild type and 9 μs for the P301L mutant, over a total of 6 independent runs. All
505 simulations were done on UTSW's biohpc computing cluster.

506 **Statistics:** All statistics were calculated using GraphPad Prism 7.0. A minimum of 3 independent ThT
507 experiments were run for each condition. Plots were fitted to a non-linear sigmoidal curve, from which
508 $t_{1/2}$ values were derived. $t_{1/2}$ error represents a 95% CI. Flow cytometry cell aggregation was conducted
509 a minimum of 3 independent experiments, whose values are plotted. Error bars represent a 95% CI.

510

511 **Acknowledgements**

512 This work was supported by grants from the Tau Consortium and NIH grants awarded to
513 1R01NS071835 (M.I.D.), the Effie Marie Cain Endowed Scholarship (L.A.J.) and the Heising-Simons
514 Foundation (M.M.L.). We appreciate the help of the Structural Biology Laboratory and Proteomics Core
515 Facility at the University of Texas Southwestern Medical Center.

516

517 **Author Contributions**

518 K.D. M.I.D., and L.A.J. conceived and designed the overall study. K.D. performed *in vitro* aggregation
519 assays, peptide studies, flow cytometry, and cell models. D.C. performed crosslink mass spectrometry.
520 D.W. performed electron microscopy. L.S. and M.L. performed molecular dynamics simulations. K.D.
521 and L.A.J. wrote the manuscript, and all authors contributed to its improvement.

522

523

524

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

Peptide 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
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 KKIETH 374

526 [1]Fluorinated proline analogs were used to preferentially populate *cis, trans,* or lowered E_a barrier
527 conformers at the positions indicated as **P**.
528 [2]Cysteines labeled for Tau RD internal FRET are market as **C**.
529

530 **Table 2: List of AlzForum Disease-Associated Mutations**

Peptide 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**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**N 368

531
532

533 **Table 3: Summary of Circular Dichroism Results**

Peptide Name	% α-helix	% anti-parallel	% parallel	% turn	% random
R1R2	0	39.5	0.9	14.5	45.1
R2R3	0	36.7	2.8	14.6	45.9
R2R3-P301L	0.3	34.0	2.6	15.4	47.7
Trp-R2R3-Trp	0	42.0	0	13.9	44.1
Trp-R2R3-P301L-Trp	0	42.5	0	13.6	43.9

534
535

536 **Figure Legends**
537
538 **Figure 1**. **Tauopathy mutations cluster to inter repeat regions**. A. Disease-associated mutation
539 frequency found in human tauopathies. Most mutations are found within the repeat domain (tau-RD)
540 (repeat 1 = red; repeat 2 = green; repeat 3 = blue; repeat 4 = purple). Amyloidogenic sequences
541 ^{275}VQIINK^{280} and ^{306}VQIVYK^{311} are shown in the inset cartoon. B. Detailed mutation frequencies found
542 near the ^{306}VQIVYK^{311} amyloid motif. C, D. Secondary structure models of the repeat2-repeat3 interface
543 (^{295}DNIKHVPGGGSVQIVYK^{311}) predicted from CS-ROSETTA (C) or *ab initio* (D) shown in cartoon
544 and colored by repeat domain. The amyloid motif ^{306}VQIVYK^{311} is colored blue, the leading repeat-2
545 sequence is green. E. Positions of disease-associated mutations (red sphe res) mapped onto the
546 ^{306}VQIVYK^{311}-containing *ab initio* β-hairpin structure.
547
548 **Figure 2. Wildtype and mutant peptides differentially populate hairpin and extended**
549 **conformations.** A. Conformations obtained for a peptide fragment wild type trimer of the sequence
550 ^{295}DNIKHVPGGGSVQIVYK^{311}. Two-dimensional root-mean-squared-differences (RMSD's) are
551 calculated between all pairs of conformations visited during MD simulations. Snapshots of trimeric
552 structures are depicted for select metastable basins, with each peptide monomer represented by a different
553 color. B. The same analysis as above, but for the P301L substituted trimer. C. The free energy surface as
554 a function of deviation from a canonical hairpin structure. Two distinct basins, corresponding to hairpin
555 and fibril/extended sub-ensembles, are found.
556
557 **Figure 3. Tau RD encodes global and local structure.** A-C. Cartoon schematic of tau RD used for FRET
558 and XL-MS studies colored according to repeat domain (repeat 1 = red; repeat 2 = green; repeat 3 = blue;
559 repeat 4 = purple). For FRET studies, endogenous cysteines are labelled with alexa-488 and alexa-647
560 (red and blue balls), which flank the ^{306}VQIVYK^{311} amyloid motif. D. Monomeric tau-RD with labeled
561 FRET pairs was denatured at varying temperatures. E-F. Tau RD samples were incubated at 37 °C or 75
562 °C for one hour. After cross-linking, trypsin fragmentation, and LC-MS/MS analysis, consensus cross-

563 link patterns (circles) are shown as contact maps. Short range crosslinks within the N-term (blue), C-term
564 (red) and long range contacts across N- to C-term (purple) are shown as sectors. Crosslinks are colored
565 according to mean frequency across replicates. G-I. Same as D-F above, except with tau RD that contains
566 a P301L disease-associated substitution.
567
568 **Figure 4. Tauopathy mutations drive aggregation propensity**. A. Schematic of tau-RD and the derived
569 peptides representing minimal structural elements around ^{306}VQIVYK^{311}. B. Primary amino acid sequence
570 mapped onto a cartoon of the predicted minimal structural element proximal to ^{306}VQIVYK^{311}. Mutations
571 are shown as spheres. C. Wild type and mutant peptides were disaggregated, resuspended at 200 μM, and
572 allowed to aggregate in the presence of ThT at room temperature. ThT signals are an average of at least 3
573 independent experiments and are colored according to mutation.
574
575 **Figure 5. Peptides form amyloid structures and seed *in vivo*.** A. After 96 hours of *in vitro* incubation,
576 peptides from previous ThT experiments (**Figure 4 C**) were transduced into tau biosensor cells *via*
577 lipofectamine (**Methods**). FRET signal from each condition (tau-RD-CFP/tau-RD-YFP) was measured
578 by flow cytometry used 3 biological triplicates of at least 10,000 cells per condition. Error bars represent
579 a 95% CI of each condition. Solid and dashed horizontal lines represent the mean and 95% error from
580 untreated biosensor cells, respectively, for ease of statistical comparison. B-H. Electron microscopy
581 images of each peptide from previous ThT experiments (**Figure 4 C**). The black bar represents 200 nm
582 distance in each image I-P. Qualitative fluorescence microscopy images of tau biosensor cells immediately
583 prior to Flow Cytometry experiments.
584
585 **Figure 6. Alternative splicing modulates aggregation propensity.** A. Cartoon schematic for tau 4R and
586 3R splice isoforms illustrate the difference in primary amino acid sequence leading into the amyloidogenic
587 ^{306}VQIVYK^{311} motif. B. A full combinatorial panel of R2R3-P301L and R1R3-P301L chimeras were
588 aggregated *in vitro*. ^{306}VQIVYK^{311} is shown in blue, amino acids common between the splice isoforms

589 are shown in black, amino acids unique to an R3 isoform are colored red, amino acids unique to an R4
590 isoform are colored green. The kinetic times to aggregate, represented as $t_{1/2}$ in hours with 95% CI, are
591 listed in the right-side column alongside its respective peptide.
592
593 **Figure 7. Enhancing β-hairpin structure rescues spontaneous aggregation phenotypes.** Cartoon
594 schematic representation of the tryptophan zipper motif (green bars) and controls used to stabilize a β-
595 hairpin structure in a R2R3-P301L peptide (**Table 1**). B. Aggregation reactions of the tryptophan zipper
596 peptide and controls measured by ThT fluorescence (200 μM, 25 °C). ThT signals are an average of at
597 least 3 independent experiments. C. Schematic of proline and fluorinated proline analogs used to generate
598 *cis* and *trans* proline conformers at the position corresponding to P301 (red ball) in peptide models. D.
599 ThT aggregation reactions of the *cis*, *trans*, and neutral proline analogs substituted into the R2R3 peptide
600 (200 μM). ThT signals are an average of at least 6 independent experiments.
601
602 **Figure 8. Molecular model of tau amyloid domain structural rearrangement and subsequent**
603 **aggregation.** Naïve tau monomer (left) exists with a propensity to form a relatively collapsed structure,
604 which buries the amyloid domain ^{306}VQIVYK^{311}. In the presence of disease-associated mutations, proline
605 isomerization events, or certain splice isoforms, the equilibrium is shifted to disfavor local compact
606 structure. This exposes the aggregation prone ^{306}VQIVYK^{311} amyloid motif and enhances aggregation
607 propensity, leading to subsequent tau pathology.
608

609 **References**

610 1. Cleveland, D. W., Hwo, S. Y. & Kirschner, M. W. Physical and chemical properties of purified tau
611 factor and the role of tau in microtubule assembly. *J. Mol. Biol.* **116,** 227–247 (1977).
612 2. Eliezer, D. *et al.* Residual structure in the repeat domain of tau: Echoes of microtubule binding
613 and paired helical filament formation. *Biochemistry* **44,** 1026–1036 (2005).
614 3. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*
615 **547,** 185–190 (2017).
616 4. Sanders, D. W. *et al.* Distinct tau prion strains propagate in cells and mice and define different
617 tauopathies. *Neuron* **82,** 1271–1288 (2014).
618 5. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers.
619 *Nature* **447,** 453–457 (2007).
620 6. Von Bergen, M. *et al.* Mutations of Tau Protein in Frontotemporal Dementia Promote
621 Aggregation of Paired Helical Filaments by Enhancing Local Beta-Structure. *J. Biol. Chem.* **276,**
622 48165–48174 (2001).
623 7. Mirbaha, H. *et al.* Inert and seed-competent tau monomers elucidate the structural origins of
624 aggregation. *bioRxiv* (2017).
625 8. Zhao, J. *et al.* Glycan Determinants of Heparin-Tau Interaction. *Biophys. J.* **112,** 921–932 (2017).
626 9. Kellogg, E. H. *et al.* Near-atomic model of microtubule-tau interactions. **1780,** 1–9 (2018).
627 10. Zhang, X. *et al.* RNA Stores Tau Reversibly in Complex Coacervates. *bioRxiv* 111245 (2017).
628 doi:10.1101/111245
629 11. Mok, S.-A. *et al.* Mapping interactions with the chaperone network reveals factors that protect
630 against tau aggregation. *Nat. Struct. Mol. Biol.* **25,** 384–393 (2018).
631 12. Baughman, H. E. R., Clouser, A. F., Klevit, R. E. & Nath, A. HspB1 and Hsc70 chaperones engage
632 distinct tau species and have different inhibitory effects on amyloid formation. *J. Biol. Chem.*
633 **293,** 2687–2700 (2018).
634 13. Walker, S., Ullman, O. & Stultz, C. M. Using intramolecular disulfide bonds in tau protein to
635 deduce structural features of aggregation-resistant conformations. *J. Biol. Chem.* **287,** 9591–
636 9600 (2012).
637 14. Kadavath, H. *et al.* Folding of the Tau Protein on Microtubules. *Angew. Chemie - Int. Ed.* **54,**
638 10347–10351 (2015).
639 15. Mutations. (2017). Available at: http://www.alzforum.org/mutations/. (Accessed: 27th
640 September 2017)
641 16. Holmes, B. B. *et al.* Proteopathic tau seeding predicts tauopathy in vivo.
642 doi:10.1073/pnas.1411649111
643 17. Yoshiyama, Y. *et al.* Synapse Loss and Microglial Activation Precede Tangles in a P301S
644 Tauopathy Mouse Model. *Neuron* **53,** 337–351 (2007).
645 18. Goedert, M., Wischik, C. M., Crowther, R. A., Walker, J. E. & Klug, A. Cloning and sequencing of
646 the cDNA encoding a core protein of the paired helical filament of Alzheimer disease:
647 identification as the microtubule-associated protein tau. *Proc. Natl. Acad. Sci.* **85,** 4051–4055
648 (1988).
649 19. Williams, D. R. Tauopathies: Classification and clinical update on neurodegenerative diseases
650 associated with microtubule-associated protein tau. *Internal Medicine Journal* **36,** 652–660
651 (2006).
652 20. Schoch, K. M. M. *et al.* Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse
653 Model. *Neuron* **90,** 941–947 (2016).
654 21. Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited
655 dementia FTDP-17. *Nature* **393,** 702–704 (1998).
656 22. Stöhr, J. *et al.* A 31-residue peptide induces aggregation of tau's microtubule-binding region in

657 cells. *Nat. Chem.* **9,** 874–881 (2017).
658 23. Falcon, B., Zhang, W., Murzin, A. G., Murshudov, G. & Holly, J. Structures of filaments from Pick '
659 s disease reveal a novel tau protein fold. (2018).
660 24. Mylonas, E. *et al.* Domain conformation of tau protein studied by solution small-angle X-ray
661 scattering. *Biochemistry* **47,** 10345–10353 (2008).
662 25. Lin, M. M. Leveraging symmetry to predict self-assembly of multiple polymers. *Chem. Phys. Lett.*
663 **683,** 347–351 (2017).
664 26. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of
665 substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159,** 1042–1055 (2014).
666 27. Tam, S. *et al.* The chaperonin TRiC blocks a huntingtin sequence element that promotes the
667 conformational switch to aggregation. *Nat. Struct. Mol. Biol.* **16,** 1279–1285 (2009).
668 28. Flory, P. J. & Volkenstein, M. Statistical mechanics of chain molecules. *Biopolymers* **8,** 699–700
669 (1969).
670 29. Mocanu, M.-M. *et al.* The potential for beta-structure in the repeat domain of tau protein
671 determines aggregation, synaptic decay, neuronal loss, and coassembly with endogenous Tau in
672 inducible mouse models of tauopathy. *J. Neurosci.* **28,** 737–748 (2008).
673 30. Rizzu, P. *et al.* High Prevalence of Mutations in the Microtubule-Associated Protein Tau in a
674 Population Study of Frontotemporal Dementia in the Netherlands. *Am. J. Hum. Genet.* **64,** 414–
675 421 (1999).
676 31. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular
677 dichroism spectroscopy. *PNAS* **112,** E3095–E3103 (2015).
678 32. Furman, J. L., Holmes, B. B. & Diamond, M. I. Sensitive Detection of Proteopathic Seeding
679 Activity with FRET Flow Cytometry. *J. Vis. Exp.* (2015). doi:10.3791/53205
680 33. Kaufman, S. K. *et al.* Tau Prion Strains Dictate Patterns of Cell Pathology, Progression Rate, and
681 Regional Vulnerability In Vivo. *Neuron* 1–17 (2016). doi:10.1016/j.neuron.2016.09.055
682 34. Kier, B. L., Shu, I., Eidenschink, L. a & Andersen, N. H. Stabilizing capping motif for beta-hairpins
683 and sheets. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10466–10471 (2010).
684 35. Vogel, M., Bukau, B. & Mayer, M. P. Allosteric regulation of Hsp70 chaperones by a proline
685 switch. *Mol. Cell* **21,** 359–367 (2006).
686 36. Gustafson, C. L. *et al.* A Slow Conformational Switch in the BMAL1 Transactivation Domain
687 Modulates Circadian Rhythms. *Mol. Cell* **66,** 447–457.e7 (2017).
688 37. Torbeev, V. Y. & Hilvert, D. Both the cis-trans equilibrium and isomerization dynamics of a single
689 proline amide modulate 2-microglobulin amyloid assembly. *Proc. Natl. Acad. Sci.* **110,** 20051–
690 20056 (2013).
691 38. Pastorino, L. *et al.* The prolyl isomerase Pin1 regulates amyloid precursor protein processing and
692 amyloid-beta production. *Nature* **440,** 528–34 (2006).
693 39. Lim, J. *et al.* Pin1 has opposite effects on wild-type and P301L tau stability and tauopathy. *J. Clin.*
694 *Invest.* **118,** 1877–1889 (2008).
695 40. Hasegawa, M., Smith, M. J., Iijima, M., Tabira, T. & Goedert, M. FTDP-17 mutations N279K and
696 S305N in tau produce increased splicing of exon 10. *FEBS Lett.* **443,** 93–96 (1999).
697 41. Pastor, P. *et al.* Familial atypical progressive supranuclear palsy associated with homozigosity for
698 the delN296 mutation in the tau gene. *Ann. Neurol.* **49,** 263–267 (2001).
699 42. Eichner, T. & Radford, S. E. A Generic Mechanism of B2-Microglobulin Amyloid Assembly at
700 Neutral pH Involving a Specific Proline Switch. *J. Mol. Biol.* **386,** 1312–1326 (2009).
701 43. Kontsekova, E., Zilka, N., Kovacech, B., Skrabana, R. & Novak, M. Identification of structural
702 determinants on tau protein essential for its pathological function: novel therapeutic target for
703 tau immunotherapy in Alzheimer's disease. *Alzheimers. Res. Ther.* **6,** 45 (2014).
704 44. Reumers, J., Maurer-Stroh, S., Schymkowitz, J. & Rousseau, F. Protein sequences encode

safeguards against aggregation. *Hum. Mutat.* **30,** 431–437 (2009).
45. O'Nuallain, B. *et al.* Kinetics and Thermodynamics of Amyloid Assembly Using a High-Performance Liquid Chromatography-Based Sedimentation Assay. *Methods Enzymol.* **413,** 34–74 (2006).
46. Mirbaha, H., Holmes, B. B., Sanders, D. W., Bieschke, J. & Diamond, M. I. Tau trimers are the minimal propagation unit spontaneously internalized to seed intracellular aggregation. *J. Biol. Chem.* (2015). doi:10.1074/jbc.M115.652693
47. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
48. 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.* **109,** 10873–10878 (2012).
49. Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science (80-.).* **327,** 1014–1018 (2010).
50. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5,** 315–318 (2008).
51. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9,** 901–903 (2012).
52. Grimm, M., Zimniak, T., Kahraman, A. & Herzog, F. XVis: A web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res.* **43,** W362–W369 (2015).
53. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100,** (2008).
54. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78,** 1950–1958 (2010).
55. 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).
56. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
57. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20,** 786–798 (1999).
58. 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).
59. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).
60. 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

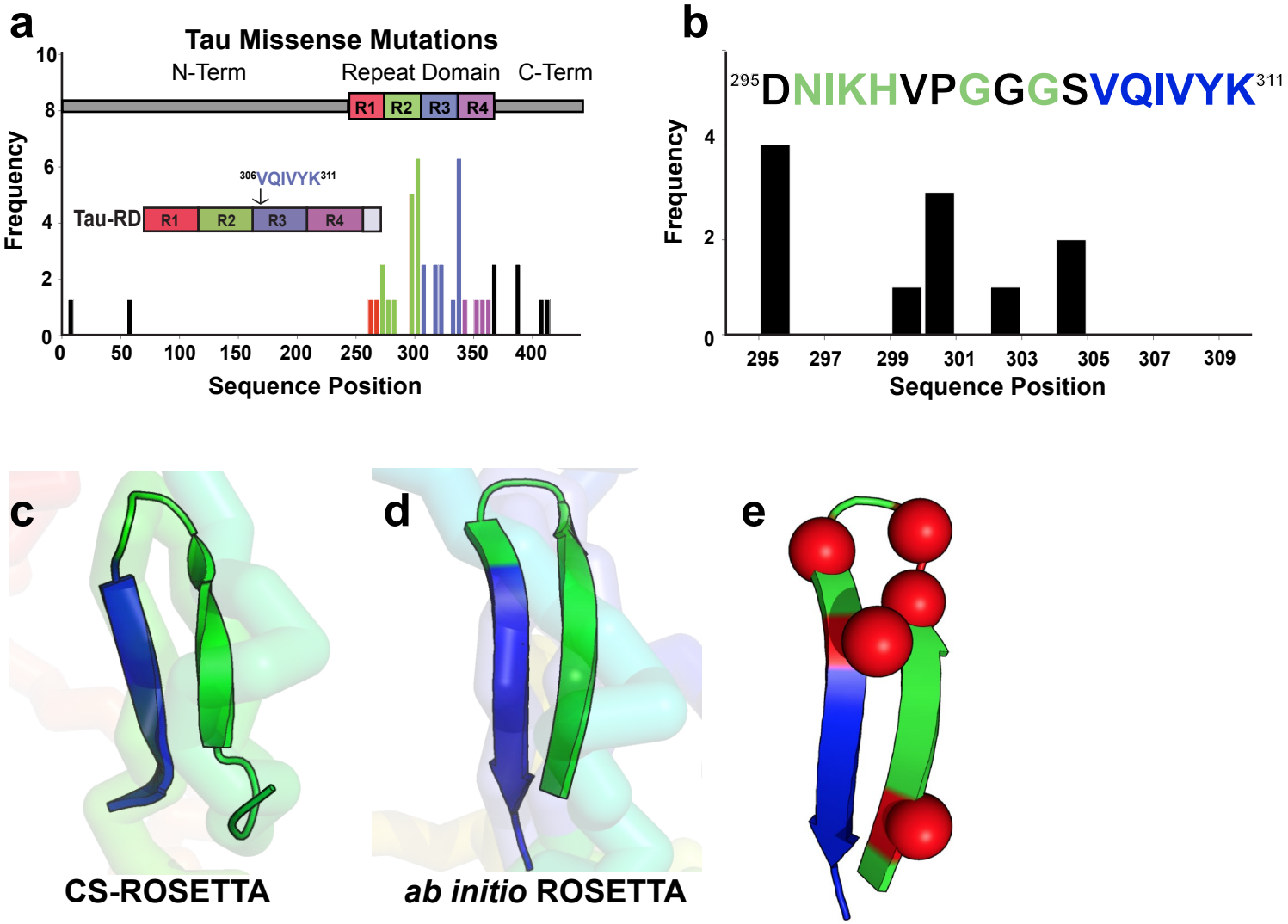

a
Tau Missense Mutations
N-Term
Repeat Domain
C-Term
R1
R2
R3
R4
306VQIVYK311
Tau-RD
Frequency
Sequence Position
b
295DNIKHVPGGGSVQIVYK311
Frequency
Sequence Position
c
CS-ROSETTA
d
ab initio ROSETTA
e

Figure 2

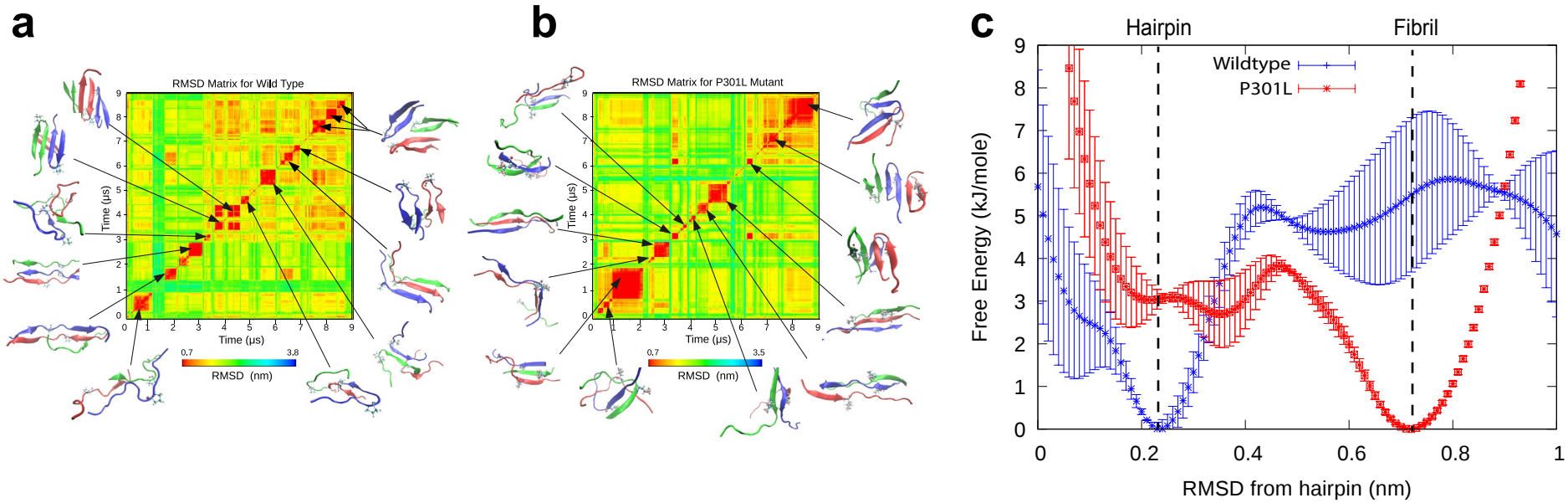

a
RMSD Matrix for Wild Type
Time (μs)
Time (μs)
0.7
3.8
RMSD (nm)
b
RMSD Matrix for P301L Mutant
Time (μs)
Time (μs)
0.7
3.5
RMSD (nm)
c
Hairpin
Fibril
Wildtype
P301L
Free Energy (kJ/mole)
RMSD from hairpin (nm)

Figure 3

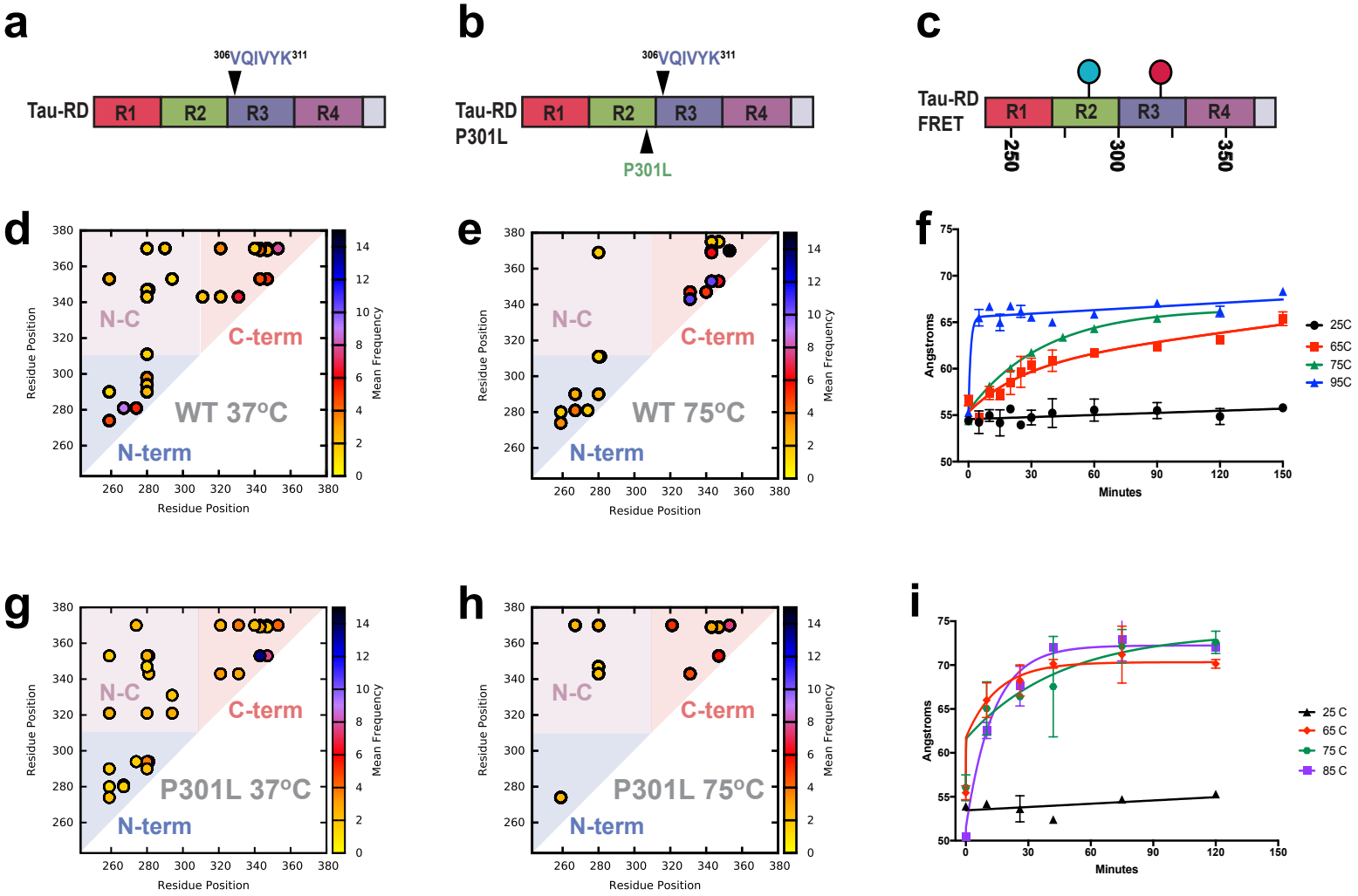
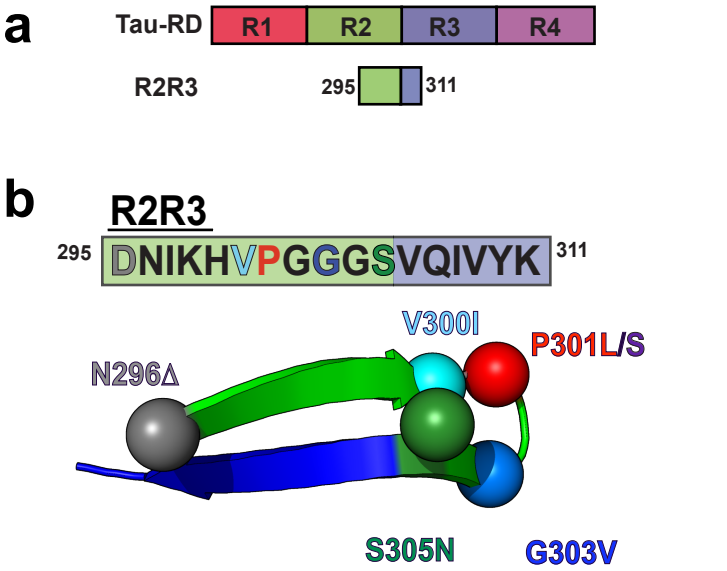

a
306VQIVYK311
Tau-RD
R1
R2
R3
R4
b
306VQIVYK311
Tau-RD
P301L
R1
R2
R3
R4
P301L
c
Tau-RD
FRET
R1
R2
R3
R4
250
300
350
d
N-C
C-term
WT 37ºC
N-term
Residue Position
Residue Position
Mean Frequency
e
N-C
C-term
WT 75ºC
N-term
Residue Position
Residue Position
Mean Frequency
f
Angstroms
Minutes
25C
65C
75C
95C
g
N-C
C-term
P301L 37ºC
N-term
Residue Position
Residue Position
Mean Frequency
h
N-C
C-term
P301L 75ºC
N-term
Residue Position
Residue Position
Mean Frequency
i
Angstroms
Minutes
25 C
65 C
75 C
85 C

Figure 4


a
Tau-RD
R1
R2
R3
R4
R2R3
295
311
b
R2R3
295
DNIKHVPGGGSVQIVYK
311
V300I
P301L/S
N296Δ
S305N
G303V

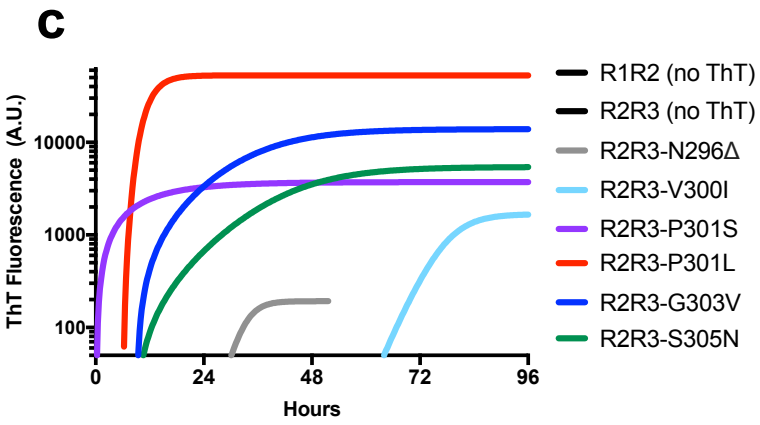

c
ThT Fluorescence (A.U.)
10000
1000
100
0
24
48
72
96
Hours
R1R2 (no ThT)
R2R3 (no ThT)
R2R3-N296Δ
R2R3-V300I
R2R3-P301S
R2R3-P301L
R2R3-G303V
R2R3-S305N

Figure 5

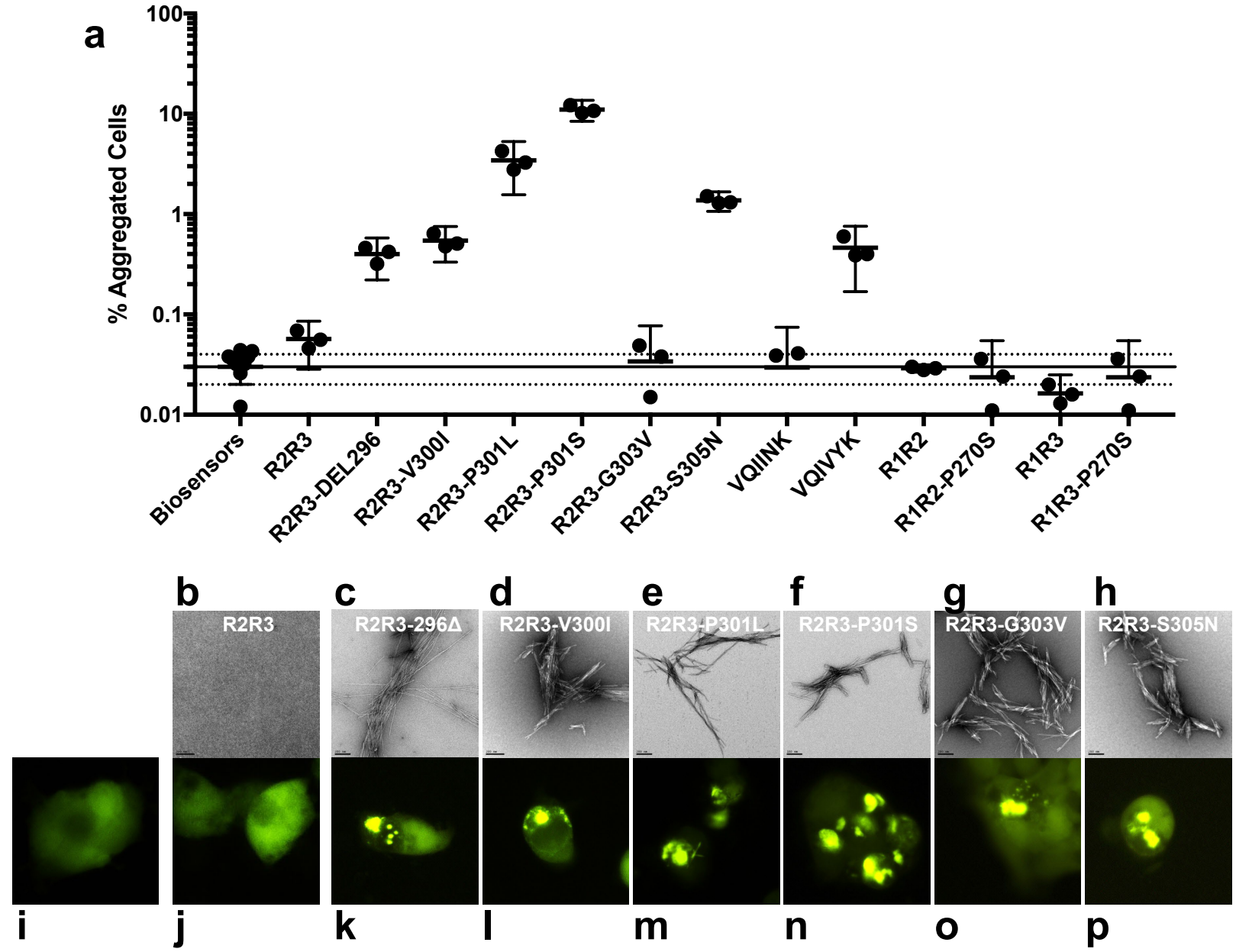

a
% Aggregated Cells
100
10
1
0.1
0.01
Biosensors
R2R3
R2R3-DEL296
R2R3-V300I
R2R3-P301L
R2R3-P301S
R2R3-G303V
R2R3-S305N
VQIINK
VQIVYK
R1R2
R1R2-P270S
R1R3
R1R3-P270S
b
c
d
e
f
g
h
R2R3
R2R3-296Δ
R2R3-V300I
R2R3-P301L
R2R3-P301S
R2R3-G303V
R2R3-S305N
i
j
k
l
m
n
o
p

Figure 6

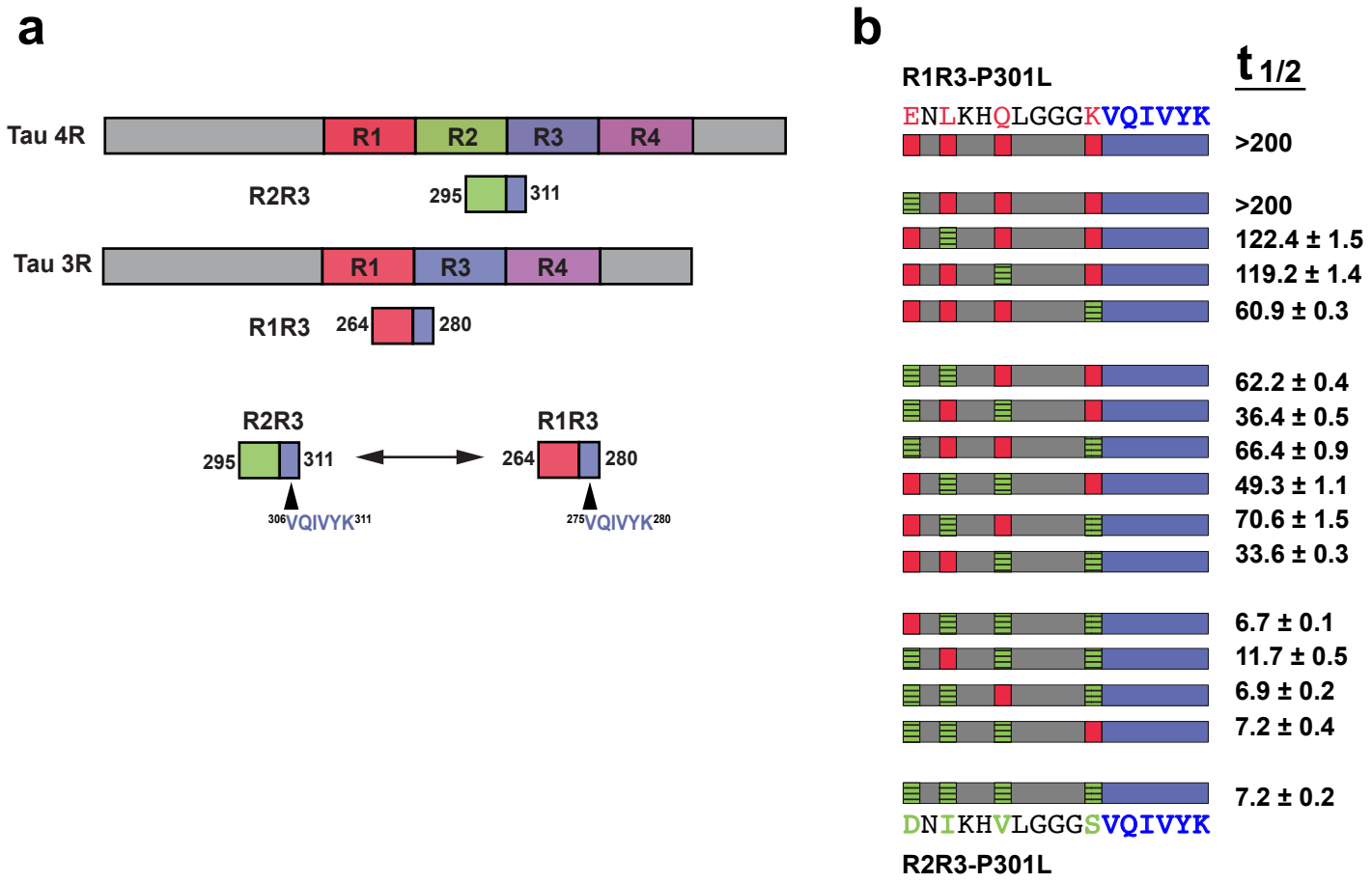

a
Tau 4R
R1
R2
R3
R4
R2R3
295
311
Tau 3R
R1
R3
R4
R1R3
264
280
R2R3
295
311
306VQIVYK311
R1R3
264
280
275VQIVYK280
b
R1R3-P301L
t 1/2
ENLKHQLGGGKVQIVYK
>200
>200
122.4 ± 1.5
119.2 ± 1.4
60.9 ± 0.3
62.2 ± 0.4
36.4 ± 0.5
66.4 ± 0.9
49.3 ± 1.1
70.6 ± 1.5
33.6 ± 0.3
6.7 ± 0.1
11.7 ± 0.5
6.9 ± 0.2
7.2 ± 0.4
7.2 ± 0.2
DNIKHVLGGGSVQIVYK
R2R3-P301L

Figure 7

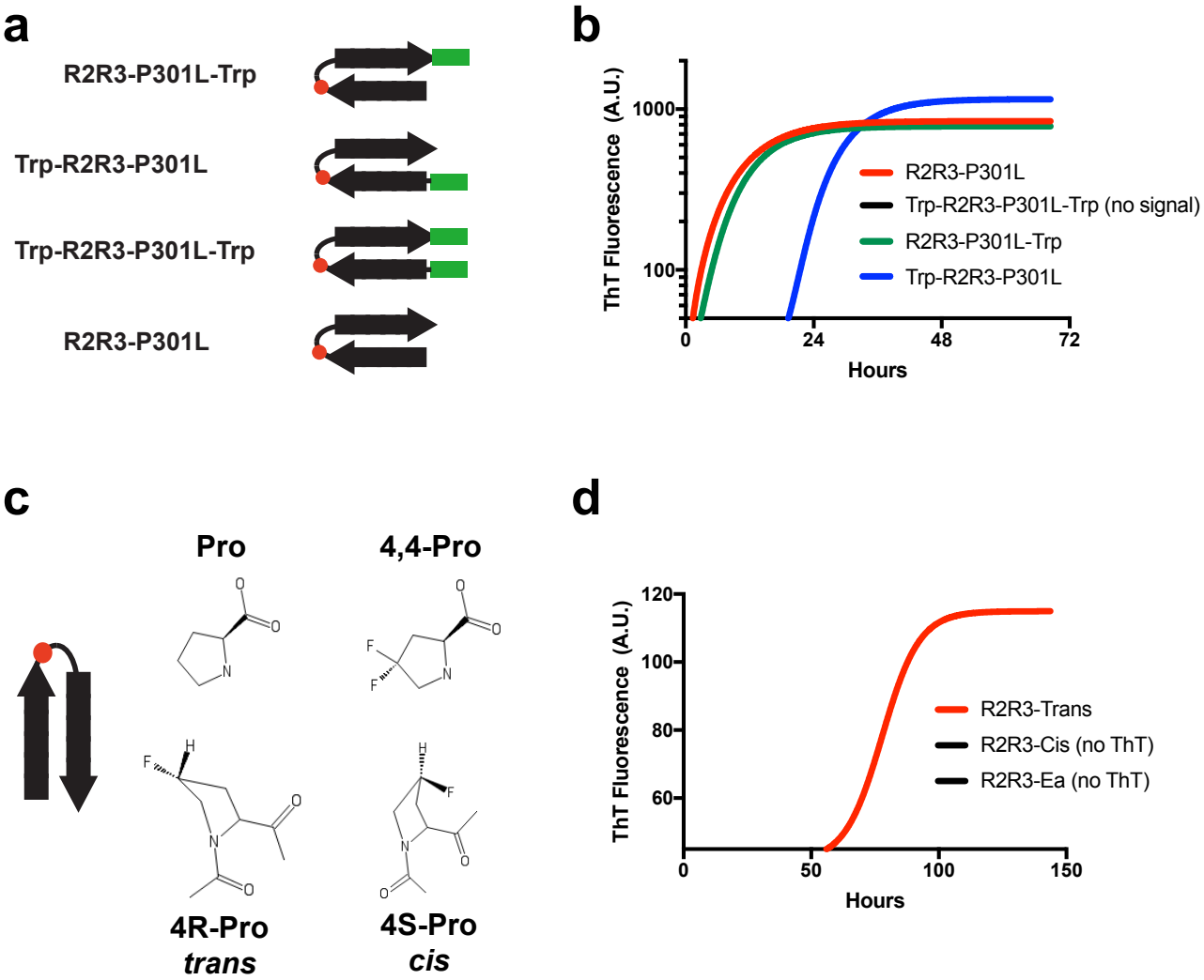
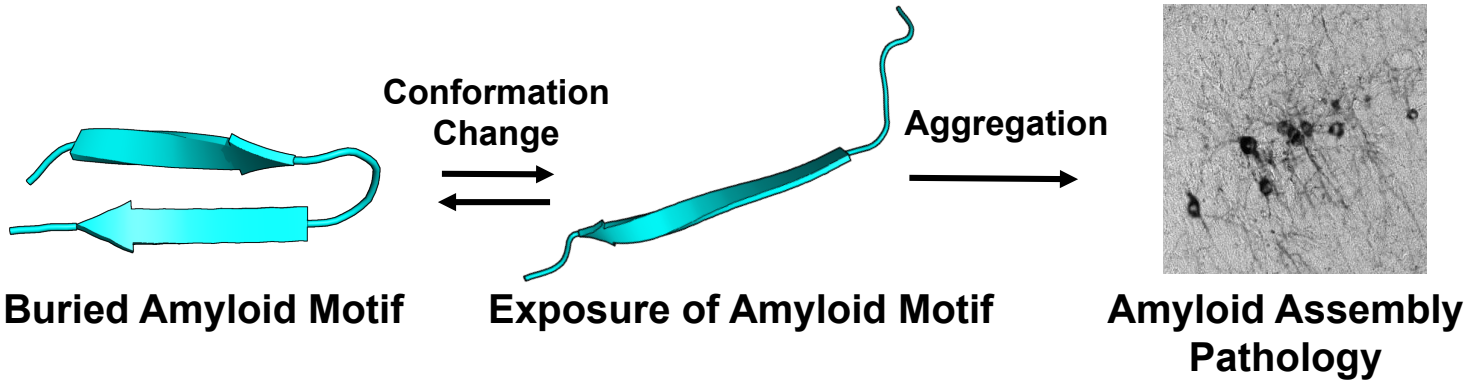

a
R2R3-P301L-Trp
Trp-R2R3-P301L
Trp-R2R3-P301L-Trp
R2R3-P301L
b
ThT Fluorescence (A.U.)
1000
100
0
24
48
72
Hours
R2R3-P301L
Trp-R2R3-P301L-Trp (no signal)
R2R3-P301L-Trp
Trp-R2R3-P301L
c
Pro
4,4-Pro
4R-Pro
trans
4S-Pro
cis
d
ThT Fluorescence (A.U.)
120
100
80
60
0
50
100
150
Hours
R2R3-Trans
R2R3-Cis (no ThT)
R2R3-Ea (no ThT)

Figure 8


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