

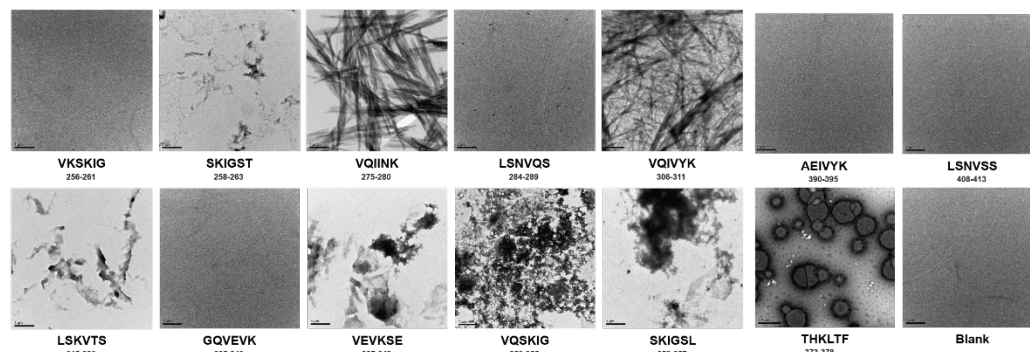
Supplementary information

Network of hotspots interactions cluster tau amyloid folds

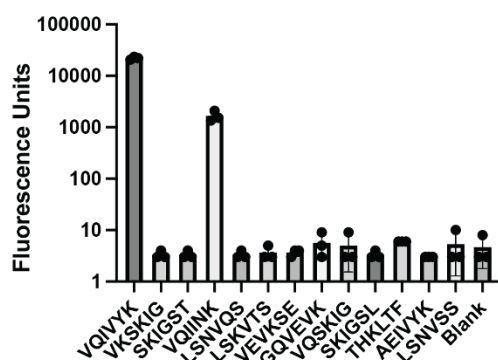
Vishruth Mullapudi^{1,4}, Jaime Vaquer-Alicea^{1,4}, Vaibhav Bommareddy¹, Anthony R. Vega¹, Bryan D. Ryder^{1,2}, Charles L. White III¹, Marc. I. Diamond¹, Lukasz A. Joachimiak^{1,3,#}

Supplementary Figures and Legends

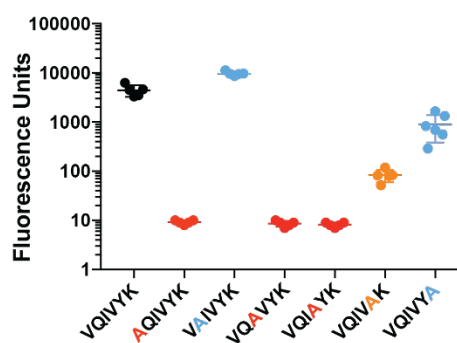
a



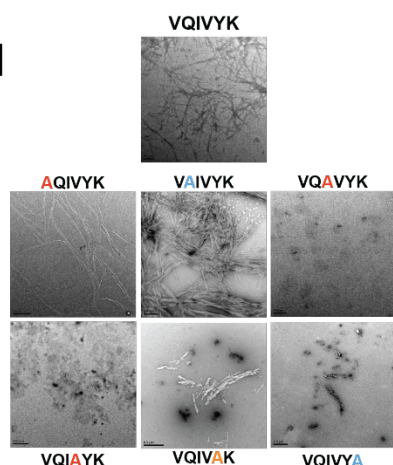
b



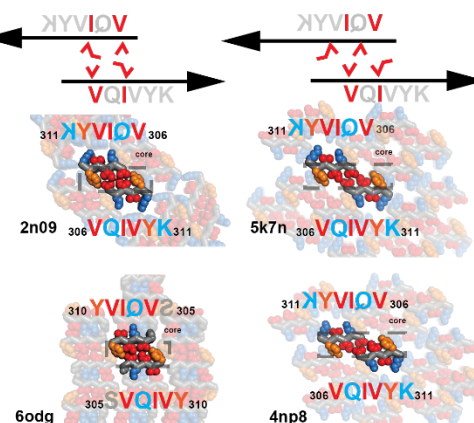
c



d



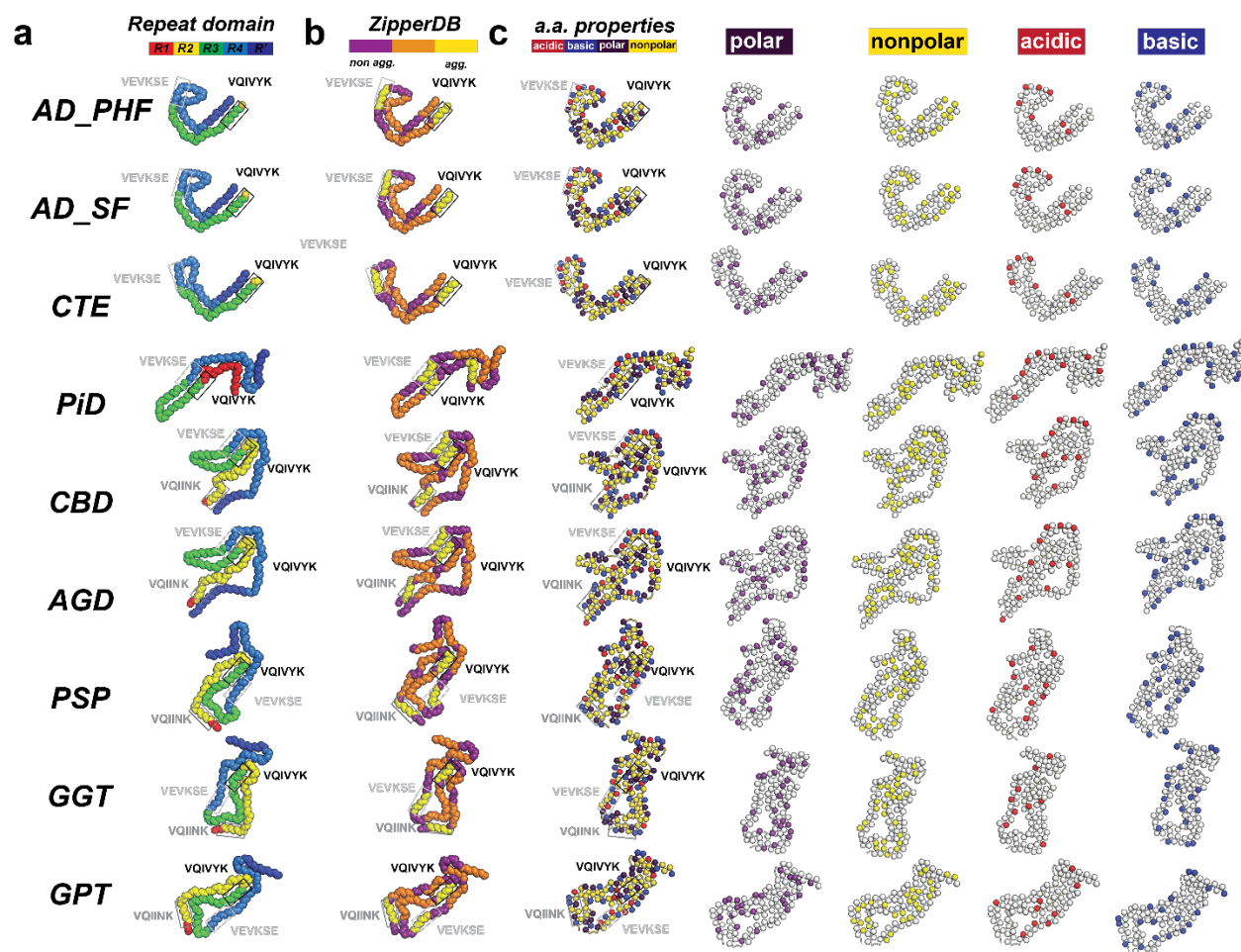
e



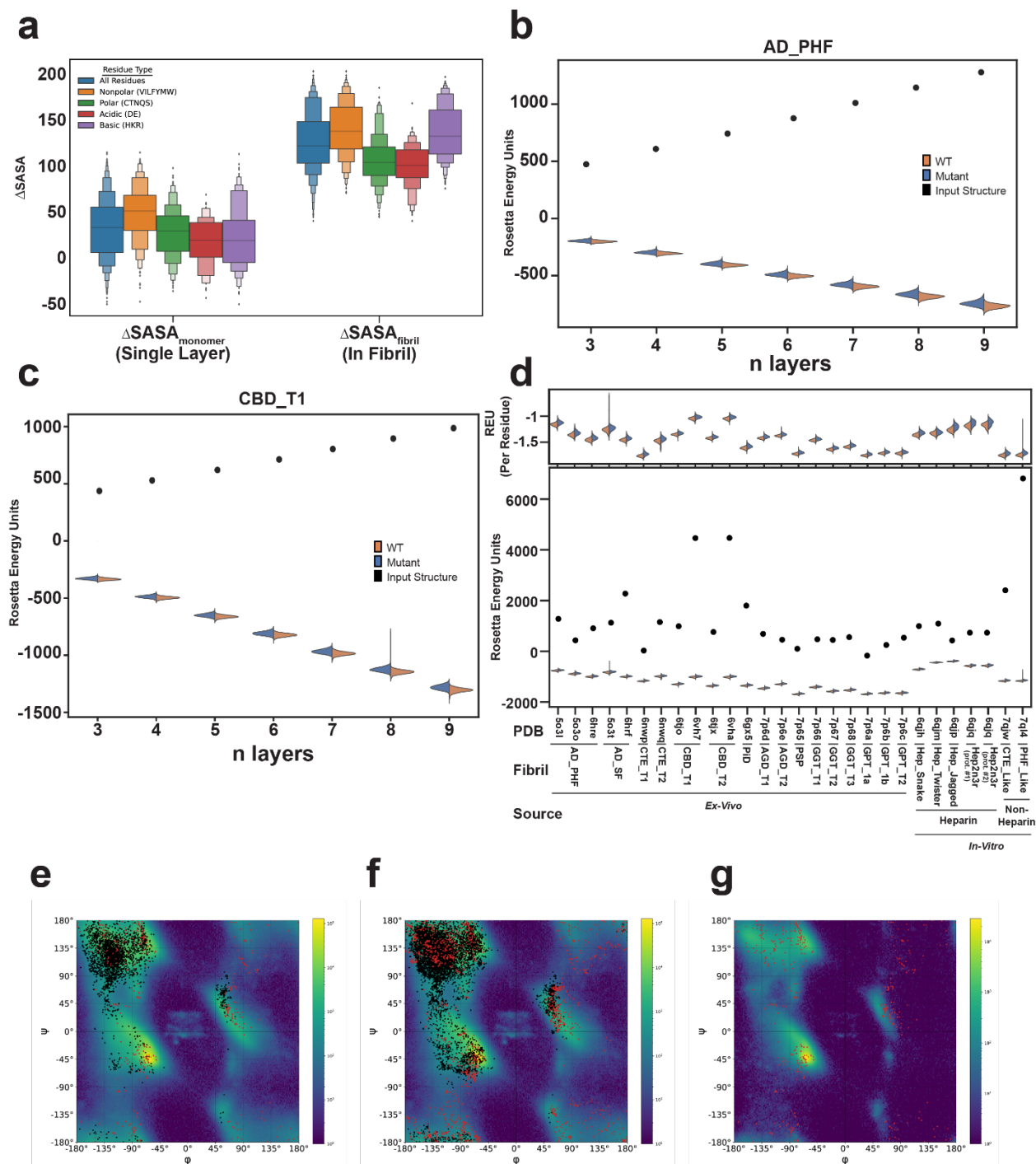
Supplementary Figure 1. Amyloid properties that determine assembly. (a)

Representative TEM images of amyloid motifs from tau predicted by ZipperDB. Images were acquired from ThT fluorescence aggregation endpoints from (b). Scale bar is 1 μ m. (b) ThT fluorescence end points for predicted amyloid motifs from tau. Aggregation experiments were performed in triplicate and shown as averages with standard deviation. (c) ThT fluorescence aggregation end points for ³⁰⁶VQIVYK³¹¹ and alanine mutants at each position. Aggregation experiments were performed as five replicates and are shown as averages with standard deviation. Peptides that yielded high ThT are

shown in blue, peptides with no ThT are colored in red and an intermediate mutant is colored in light blue. **(d)** TEM images of VQIVYK and its alanine mutants. Scale bar is 1 μ m. **(e)** Mapping aggregation properties from the alanine mutants onto four available structures of VQIVYK (A-C; PDB IDs: 2n09, 5k7n and 4np4) or SVQIVY (D; PDB id 6odg). Symmetry related lattice of the peptide is shown in spacefill representation and the residues in the core “dimer” are stabilized key residues important for aggregation. The residues are colored by their aggregation signal from **(c)**. The structures are summarized by the relative register of the monomers in the core dimer, in register (left) or off set (right).

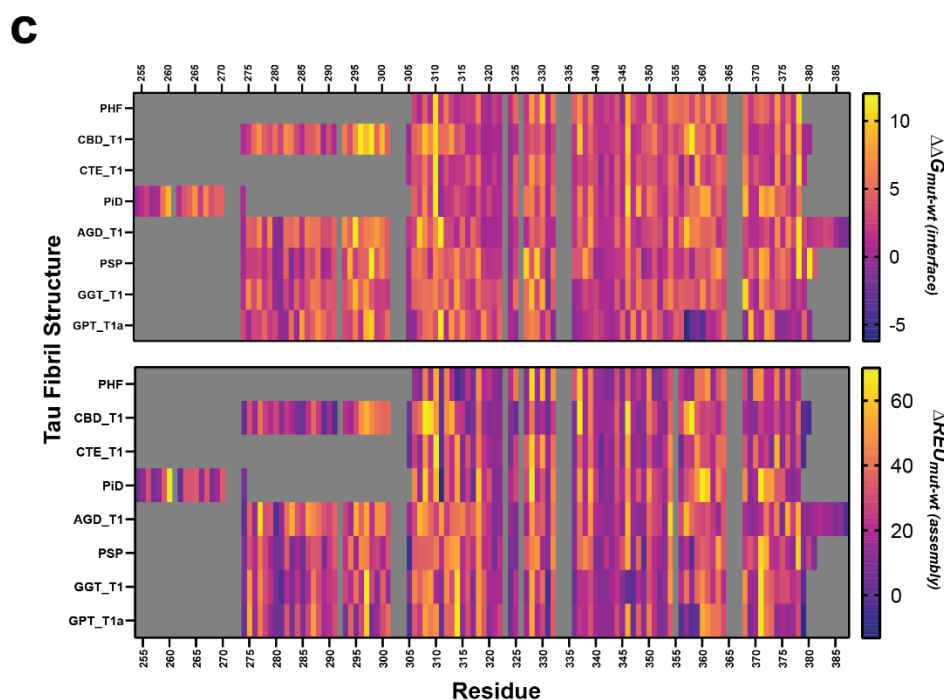
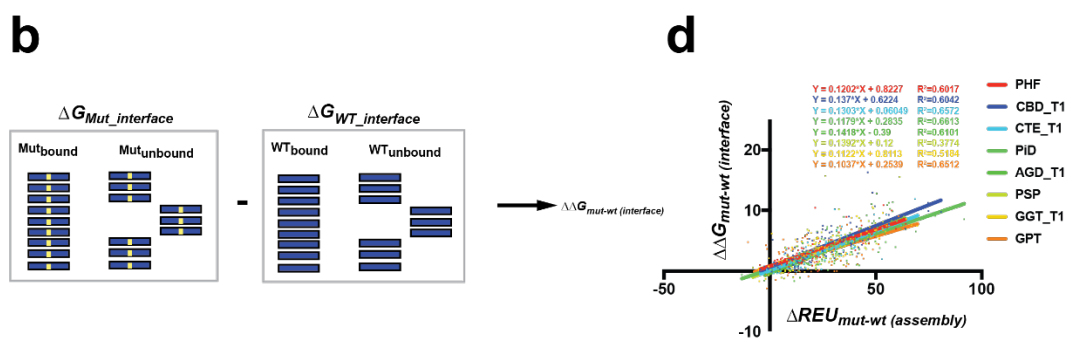
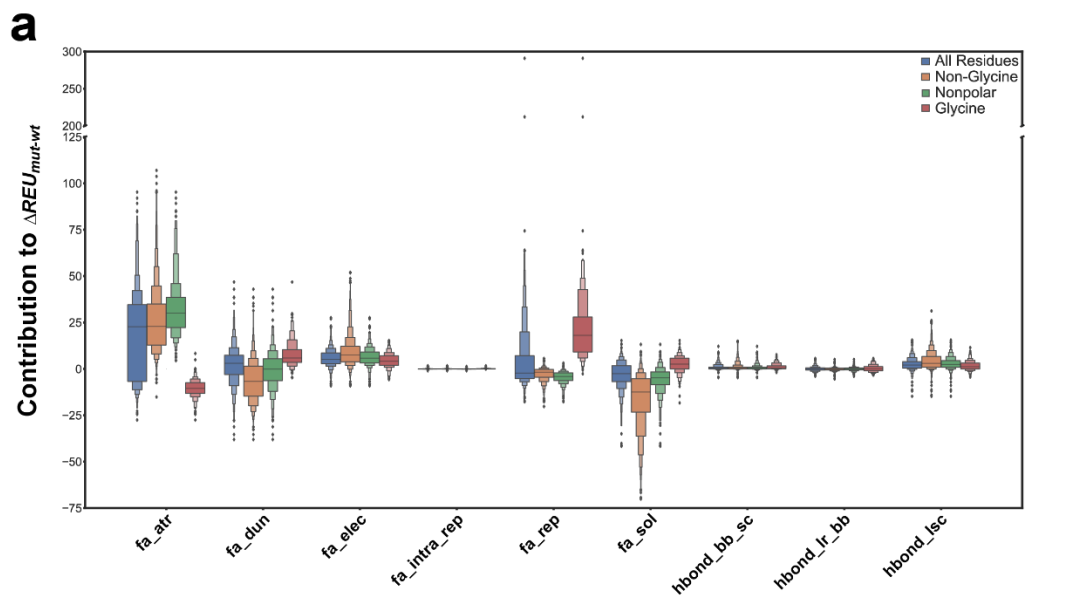


Supplementary Figure 2. Features of the tauopathy fibril structures. (a) Fibril structures are colored by repeat domain in red, yellow, green, blue and dark blue for repeat domains R1, R2, R3, R4, and R' respectively. Structures are shown in spacefill representation for the backbone. (b) Fibrils are colored by aggregation propensity of the different motifs colored in yellow, orange and magenta for high, medium, and low aggregation propensity. Structures are shown in spacefill for the backbone. (c) Fibrils are colored by amino acid type, nonpolar, polar, acidic, and basic are colored yellow, purple, red and blue respectively. Structures are shown using a ribbon representation and the c- β atom is shown in spheres. The $^{306}\text{VQIVYK}^{311}$, $^{275}\text{VQIINK}^{280}$, and $^{337}\text{VEVKSE}^{342}$ motifs are highlighted in a box in each structure. PDB ids: 5o3l, 5o3t, 6gx5, 6nwp, 6tjo, 7p6d, 7p65, 7p66 and 7p6a.



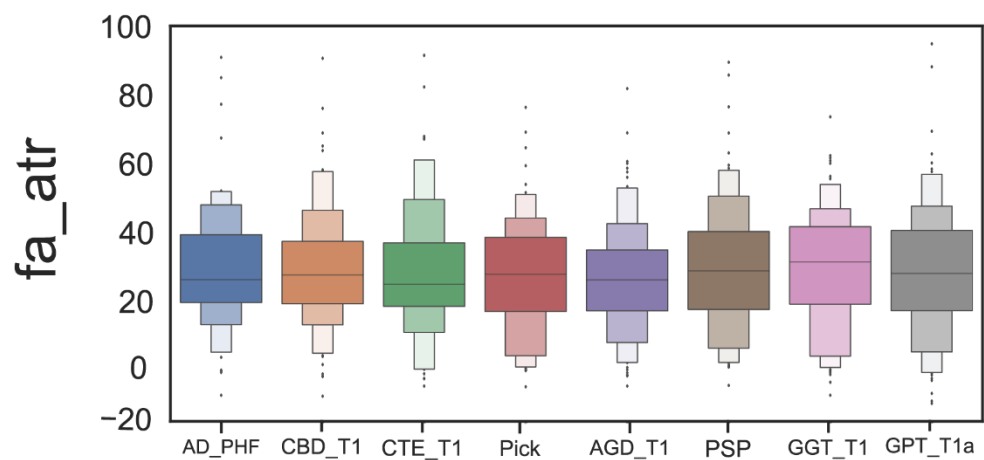
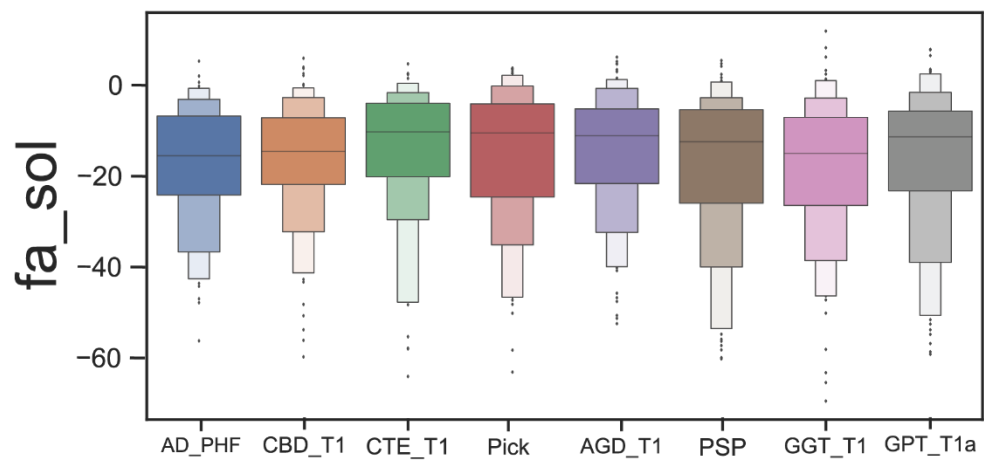
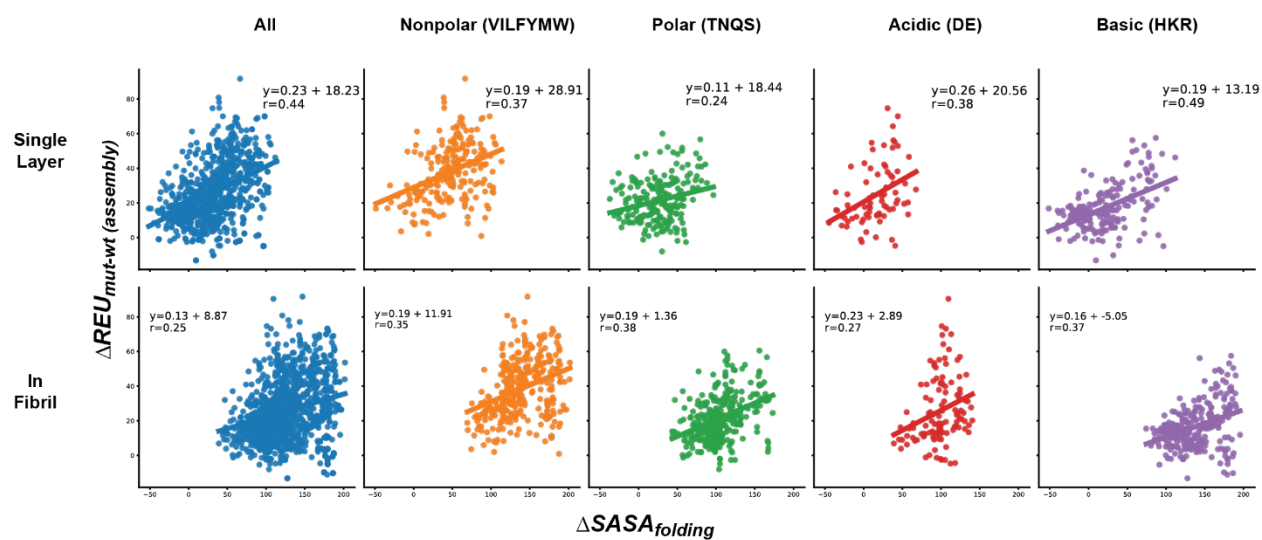
Supplementary Figure 3. Evaluating stability of fibril assemblies. (a) Quantification of the change in Solvent Accessible Surface Area (SASA) from the unfolded monomer to the folded monomer ($\Delta\text{SASA}_{\text{folding}}^{\text{monomer}}$) (left panel) or to the folded monomer in the fibril ($\Delta\text{SASA}_{\text{folding}}^{\text{in fibril}}$) (right panel) calculated for all, nonpolar, polar, basic, and acidic residues colored in blue, orange, green, red, and purple. Data is shown as a boxen or 'letter-value' plot for ΔSASA values for nine tauopathy protofilament structures. Evaluation of total energies (REU) for AD-PHFs (PDB id: 5o3l) (b) and CBD (PDB id: 6tjo) (c) using

different numbers of layers from three to nine for all structures produced during the alanine scan. Distribution of energies for WT and alanine mutants are shown in orange and blue, respectively. Energies for native starting structures are shown as black points. **(d)** Total energy distributions for minimized wild-type and alanine mutant nine-mer (bottom) using our protocol for all available tauopathy and four heparin tau fibrils, along with the energies of the input, un-minimized structures (points). The same distributions are shown normalized to the number of residues in the repeating monomer unit (top). The plot is colored as in **(b, c)**. PDB IDs: 7p6d, 7p6e, 6vha, 6vh7, 6tjo, 6tjx, 5o3l, 6hre, 7qjw, 6nwp, 6nwq, 6hrf, 7ql4, 5o3o, 5o3t, 7p65, 7p66, 7p67, 7p68, 6gx5, 7p6c, 7p6a, 7p6b. **(e)** Overlay of ϕ/ψ torsional distributions for X-ray structures of globular proteins with resolution between 1.5 Å and 2.5 Å (110,677 structures) to ϕ/ψ torsional angles for all residues (black) and glycine residues (red) in tau fibril structures (35 structures) or **(f)** all fibril structures (96 structures) determined using helical reconstruction using RELION. **(g)** Overlay of alanine ϕ - ψ torsional distributions (background) for X-ray structures of globular proteins with resolution between 1.5 Å and 2.5 Å with glycine ϕ/ψ torsional angles (red) in tau fibrils determined using helical reconstruction in RELION. Δ SASA distributions are shown as letter-value plots with the center 2 boxes showing 50% of the data with each smaller box contain half of the remaining data. REU energy distributions are shown across 35 replicates at each position in each fibril and plotted as violin plots.



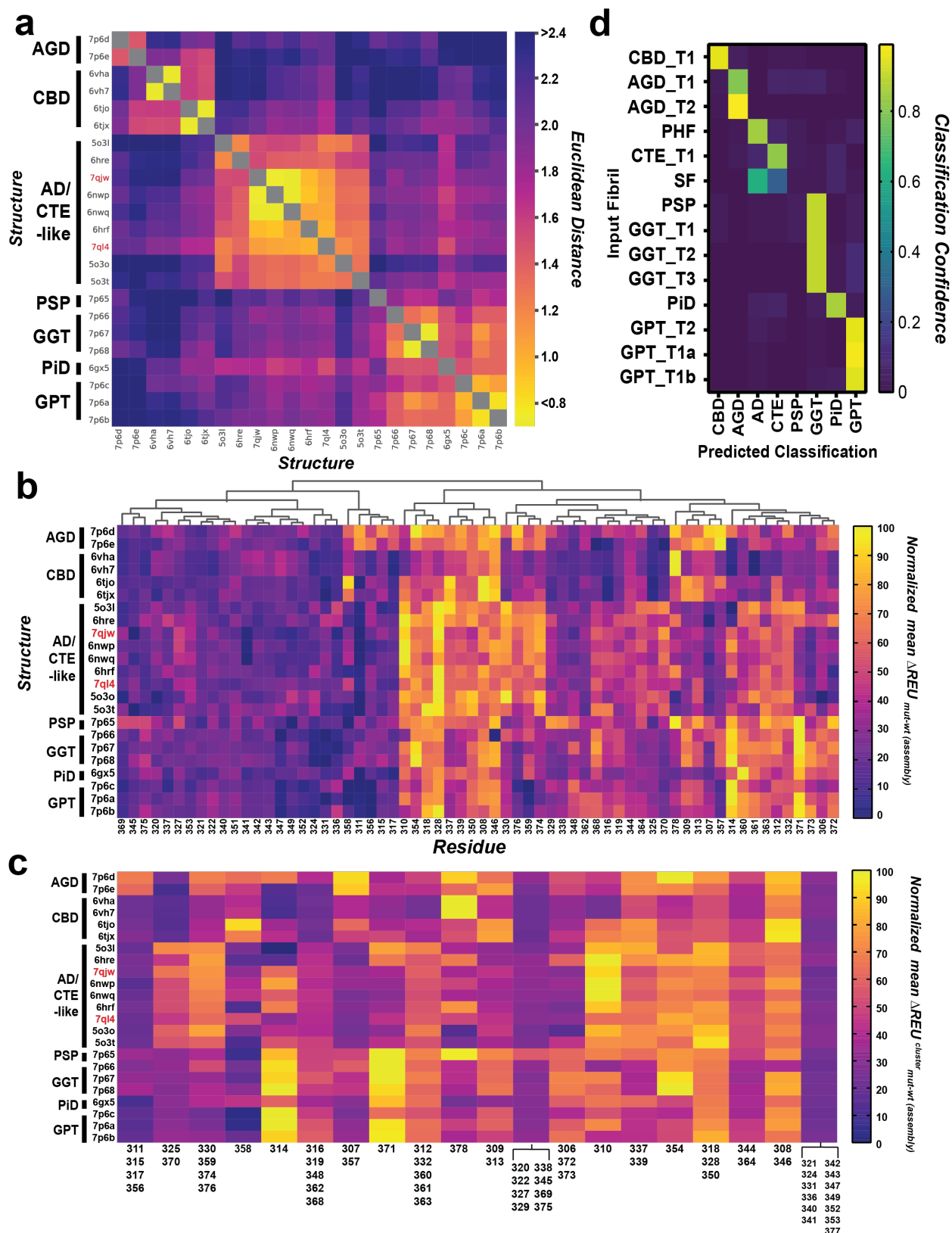
Supplementary Figure 4. Evaluation of per residue energetics in fibrils. (a)

Breakdown of energy terms from Rosetta energy calculations evaluating their overall contribution to the assembly stability across nine different tauopathy fibril structures. Distributions for all, non-glycine, nonpolar and glycine are colored in blue, orange, green and red, respectively. Energy term distributions are shown across 35 replicates across positions clustered by amino acid properties. Energy distributions are shown as letter-value plots with the center 2 boxes showing 50% of the data with each smaller box contain half of the remaining data. **(b)** Schematic for the Flex ddG-based protocol for determining the contribution of residues to the energetics of inter-layer interfaces in protein fibrils. **(c)** Correlations between the *in silico* estimation of assembly stability upon mutation (x-axis) and the *in silico* “interface” inter-layer contribution of the assembly stability upon mutation (y-axis) for tauopathy fibrils. PDB ids: 5o3l, 6gx5, 6nwp, 6tjo, 7p6d, 7p65, 7p66 and 7p6a. **(d)** Heatmap comparison of the energetic change in response to substitution to alanine of the interface within a tau fibril (top) or of the total structural energy (bottom) as predicted by Rosetta. Scale is colored in the plasma color scheme from yellow (most important) to purple (least important).

a**b****c**

Supplementary Figure 5. Energy contributions and solvent accessibility to stabilizing fibril conformations. Comparison of the contribution of the full atom Lennard-Jones attractive potential energy (fa_atr) **(a)** and the full atom Lazaridis-Karplus solvation energy (fa_sol) **(b)** as calculated by Rosetta to the *in silico* predicted energy for the assembly of selected tauopathy fibril structures. Contributions are plotted in Rosetta Energy Units and are shown as letter-value plots with the center 2 boxes showing 50% of the data with each smaller box contain half of the remaining data. **(c)** Correlations between the Δ SASA of folding and the *in silico* change in assembly energy upon mutation (y-axis) for all, nonpolar, polar, acidic, and basic residues, either in the context of a single monomer folded into its fibril conformation without (top) or within (bottom) the context of the fibril. PDB ids: 5o3l, 6gx5, 6nwp, 6tjo, 7p6d, 7p65, 7p66 and 7p6a.

conformations on ³⁵⁸DNITHV³⁶³. Residues are colored in the plasma color scheme from yellow (important) to purple (unimportant) by their effect on *in silico* assembly stability when mutated to alanine. Structures are shown as ribbons and key residues are shown in stick representation. Endpoint Thioflavin T fluorescence values for 200 μ M ³⁰⁶VQIVYK³¹¹ peptide aggregation (left), 200 μ M ³⁰⁶VQIVYK³¹¹ with 50, 100 or 200 μ M competitor (middle) and 50, 200 and 200 μ M competitor alone (right). Competitor peptides ³⁵⁰VQSKIG³⁵⁵ (c) and ³³⁷VEVKSE³⁴² (d) were co-aggregated with ³⁰⁶VQIVYK³¹¹. Experiments were performed as six technical replicates and are reported as averages with standard deviation. TEM images of each endpoint sample are shown on the right. Scale bars are shown as 0.2 μ m. (e) Aggregation of 50 μ M ³³⁷VEVKSE³⁴², VEVKSE alanine mutants (AEVKSE and VEAKSE) and GSPSGS control peptide with and without the presence of 200 μ M ³⁰⁶VQIVYK³¹¹ peptide. Experiment was performed as six technical replicates. Bars report average and standard deviation. (f) Gating strategy for flow cytometry readout of in-cell incorporation assay on tau alanine mutants.



Supplementary Figure 7. ML-based classification of features based on residue energetics. (a) Similarity matrix of ex vivo fibrils based on *in silico*

$\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$ measurements. Cells colored with the plasma color scheme by Euclidean distance between vectors of fibril pairs' $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$ measurements from yellow (<0.8 REU, most similar) to blue (>2.4, most dissimilar). **(b)** Dendrogram generated by hierarchical clustering via Ward's method of fibril residues, displayed alongside heatmap of *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$. Dendrogram groups residues that covary by *in silico* $\Delta\text{REU}_{\text{wt-mut}}^{\text{assembly}}$ across structures together. Residues are colored in the heatmap with the plasma color scheme from yellow (highest *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$) to purple (lowest *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$). **(c)** Heatmap of residue clusters generated via feature agglomeration, recursively combining the most similar residues/residue clusters together. Clusters are colored by the arithmetic mean value of residues comprising the cluster from yellow (highest mean *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$) to purple (lowest mean *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$). **(d)** Classification confidence matrix of the Random Forest Classifier when predicting classifications using the clustered mean *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$ values derived from the *in silico* alanine scan. When given an input fibril structure's clustered mean *in silico* $\Delta\text{REU}_{\text{mut-wt}}^{\text{assembly}}$ values, the squares represent the class probabilities the random forest classifier assigns to the input, colored with the plasma color scheme from yellow (1.0 probability, the classifier model is highly certain of a class assignment to the input) to purple (0 probability, the classifier model does not predict the input to belong to a given class). PDB IDs: 7p6d, 7p6e, 6vha, 6vh7, 6tjo, 6tjx, 5o3l, 6hre, 7qjw, 6nwp, 6nwq, 6hrf, 7ql4, 5o3o, 5o3t, 7p65, 7p66, 7p67, 7p68, 6gx5, 7p6c, 7p6a, 7p6b.