nature research

Corresponding author(s): LA Joachimiak

Last updated by author(s): Aug 10, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	ThT data were acquired on a Tecan Spark platereader (https://lifesciences.tecan.com/multimode-plate-reader). FRET analysis on tauRD mutants was acquired on a BD Fortessa flow cytometer. TEM images were acquired on a FEI Tecnai G2 Spirit Biotwin microscope. MD simulations were performed using GROMACS-5.04 (available http://www.gromacs.org). Tau fibril dREU calculations were carried out with ROSETTA (available at https://www.rosettacommons.org/).
Data analysis	All images of structures were produced in pymol v1.8.4.2. All plots were generated with GraphPad Prism 8. FRET data was analyzed using FlowJo v10 (available at https://www.flowjo.com/solutions/flowjo/downloads).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All ThT aggregation data, Flow cytometry data and XL-MS data are available as source data 1, 2 and 3, respectively. All other data are available from the authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not study populations either of animals or humans, thus population sample size is not applicable. In the case of Rosetta dREU calculations we ran 35 independent replicates for each designed mutant (and wild-type) minimization which we and others have shown converges on a solution (Barlow et al. 2018). Using the biohpc computer cluster at UTSW this represented 2 weeks of computer time.
Data exclusions	No data were excluded in the analyses
Replication	Figure 1 and Supplementary Figure 1. The ThT fluorescence aggregation assay was performed as technical triplicates and the data were plotted as averages with standard deviation. The data were fit to a non-linear regression model fitting in GraphPad Prism to estimate an average t1/2max with a standard deviation. TEM images were collected twice. The tau aggregation experiments in cells were carried out as biological triplicates and the FRET levels were reported as averages with standard deviation. Figure 2 and Supplementary Figure 2. The ThT aggregation assay was performed as technical triplicates and the data were fit to a non-linear regression model fitting in GraphPad Prism to estimate a t1/2max with a standard deviation. The t1/2max is plotted as averages with standard deviation. TEM images were collected twice. Figure 3 and Supplementary Figure 3. Molecular Dynamic simulations were performed as 5 independent replicate trajectories. The merged trajectories are shown as contact maps. Similarly, from the merged ensembles we calculated distance distributions to residue at position 320. The ensembles were clustered based on similarity and representative structures were selected for the two top 5 clusters. Distances between different residues were used to classify the ensemble below a distance threshold. Figure 4 and Supplementary Figure 4. The ThT fluorescence aggregation assay was performed as technical triplicates and the data were plotted as averages with standard deviation. The data were fit to a non-linear regression model fitting in GraphPad Prism to estimate an average t1/2max with a standard deviation. TEM images were collected twice. The tau aggregation experiments in cells were carried out as biological triplicates and the FRET levels were reported as averages with standard deviation. The crosslinking mass spectrometry experiment were performed in 5 technical replicates. The frequency of each high-scored crosslink identified was normalized to the total number of high-scored crosslinks and the average % with standard deviations were plotted in the bar plots. The contact maps show only the average %. Figure 5 and Supplementary Figure 5. The Rosetta design calculations at defined positions were substituted for each amino acid using a 9-layer fibirl assembly. Simulations were carried out using 35 replicates to ensure convergence of energies as previously determined (Barlow et al. 2018). Low energy structures of each mutant are shown as a matrix. Similarly the lowest energy structure is shown. The ThT aggregation assay was performed as technical triplicates. The data were fit to a non-linear regression model fitting in GraphPad Prism to estimate an average t1/2max with a standard deviation. TEM images were collected twice. The cell experiments were carried out as biological triplicates and the FRET levels were reported as averages with standard deviation. Figure 6. N/A
Randomization	Samples were not allocated into groups, so randomization is not applicable.
Blinding	Samples were not allocated into groups; blinding was not applied.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293T/17 [HEK293T/17] (ATCC CRL-1268)
Authentication	Cell lines were not authenticated
Mycoplasma contamination	Cells were confirmed to be free of mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used in this study

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	HEK293 cells stably expressing tauRD fused C-terminally to mEOS3.2 were expressed. A portion of the mEOS was converted with UV, the cells fixed with PFA and analyzed by FACS.
Instrument	BD Fortessa
Software	FlowJo v10
Cell population abundance	20,000 cells were analyzed for each condition in triplicate
Gating strategy	Gates were selected in this order: live cells, FSC singlets, SSC singles, FITC/mCherry positive cells, and FRET positive population

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.