Supplementary Information

Structural and Thermodynamic Classification of Amyloid Polymorphs

Jack P. Connor[1], Sheena E. Radford[1*], and David J. Brockwell[1*]

[1]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

[*]Correspondence: s.e.radford@leeds.ac.uk; d.j.brockwell@leeds.ac.uk

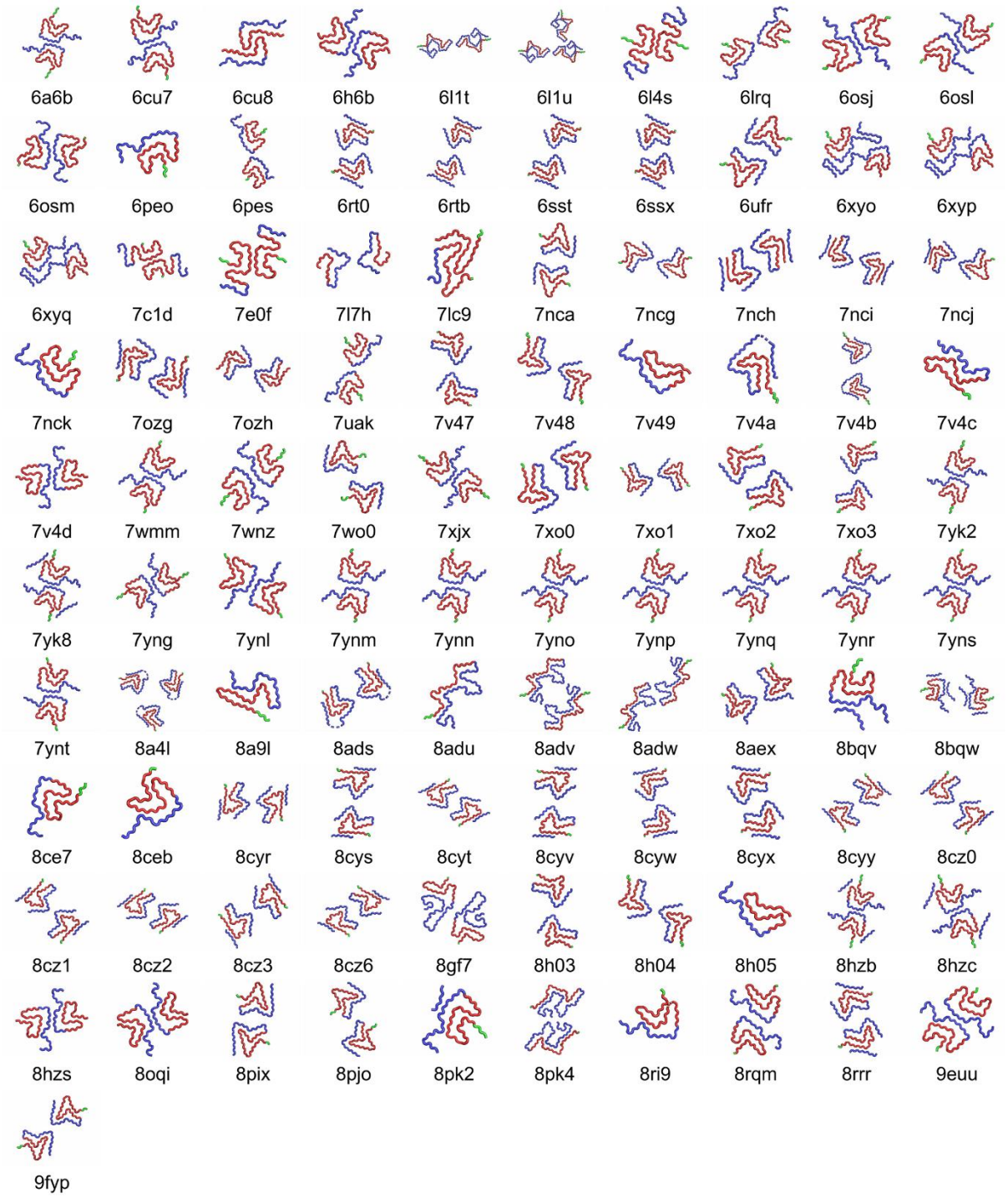


Figure S1: All 101 published cryo-EM α-synuclein amyloid fibril structures listed on Amyloid Atlas[1] (as of September 2024), related to STAR Methods. The domains are coloured as follows; N-terminal domain in blue, NAC region in red, and C-terminal domain in green. Regions shown contain residues that are present in the ordered fibril core detected by cryo-EM.

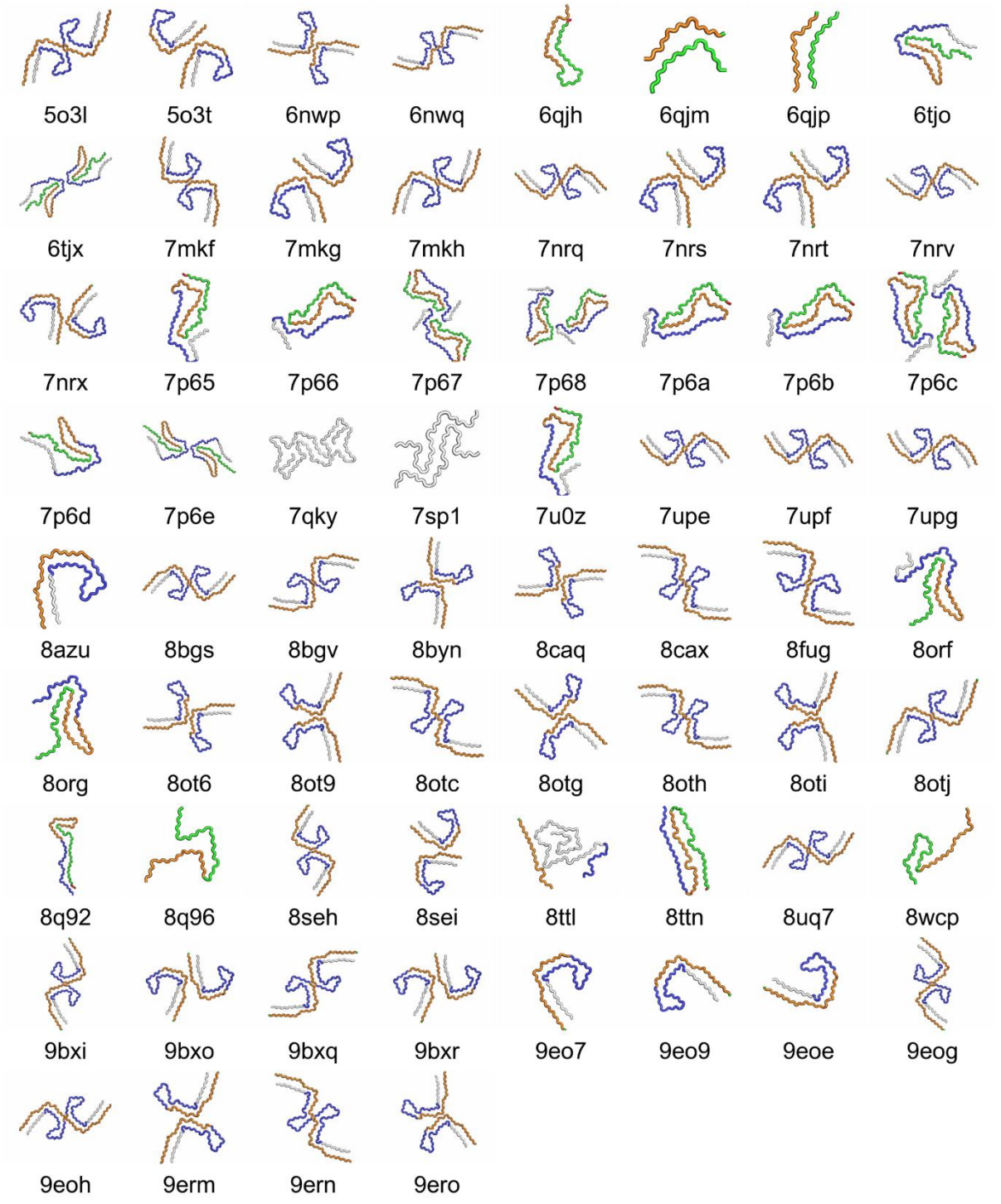


Figure S2: All 68 published cryoEM 4R or 3R+4R tau amyloid fibril structures listed on Amyloid Atlas[1] (as of September 2024), related to STAR Methods. The regions are coloured as follows; R1 in red, R2 in green, R3 in orange, R4 in blue and the remaining residues in white. Regions shown contain residues that are present in the ordered fibril core detected by cryo-EM.

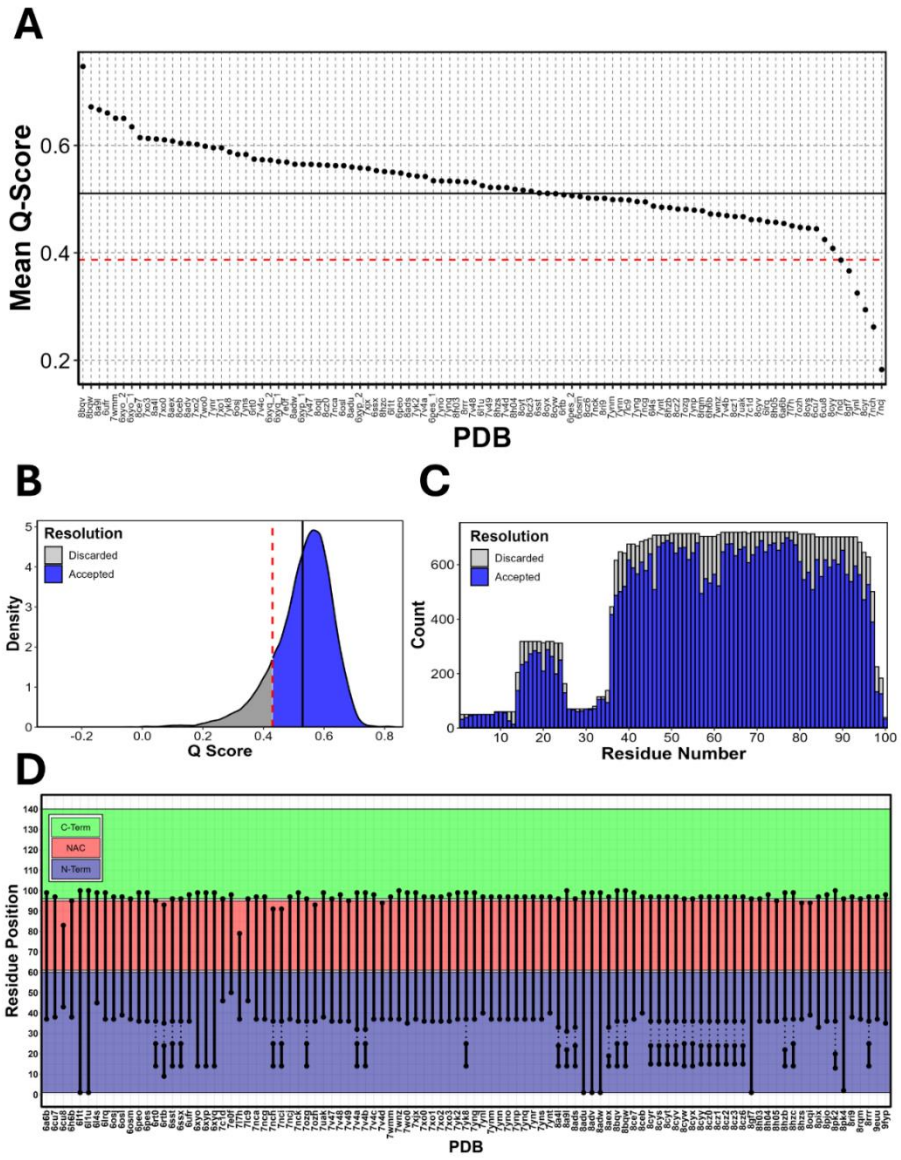


Figure S3: Q-Score filtering to remove poorly resolved structures and individual residues in α-synuclein amyloid structures, related to STAR Methods. **A)** The average Q-Score for each cyro-EM α-synuclein structure. The mean Q-score of all structures and the mean minus one standard deviation are shown by the black and red lines, respectively. Structures with a Q-score lower than the mean - 1SD were removed from subsequent analysis. **B)** A density plot showing the per residue Q-score for the remaining structures. Again, the overall mean Q-score and mean - 1SD are shown by black and red lines. Residues with a Q-score lower than the mean - 1SD were removed from subsequent analysis. **C)** A bar plot showing the number of times each residue position occurs across all structures. The bars are coloured by the number of times the given residue position is above or below the Q-score threshold. **D)** Shown are the residues resolved in the fibril core of the solved amyloid structure for each PDB. Dotted lines indicate regions of unresolved residues within the core. Note that α-synuclein has 140 residues. Residues 101-140 are not observed at high resolution in any cryo-EM maps.

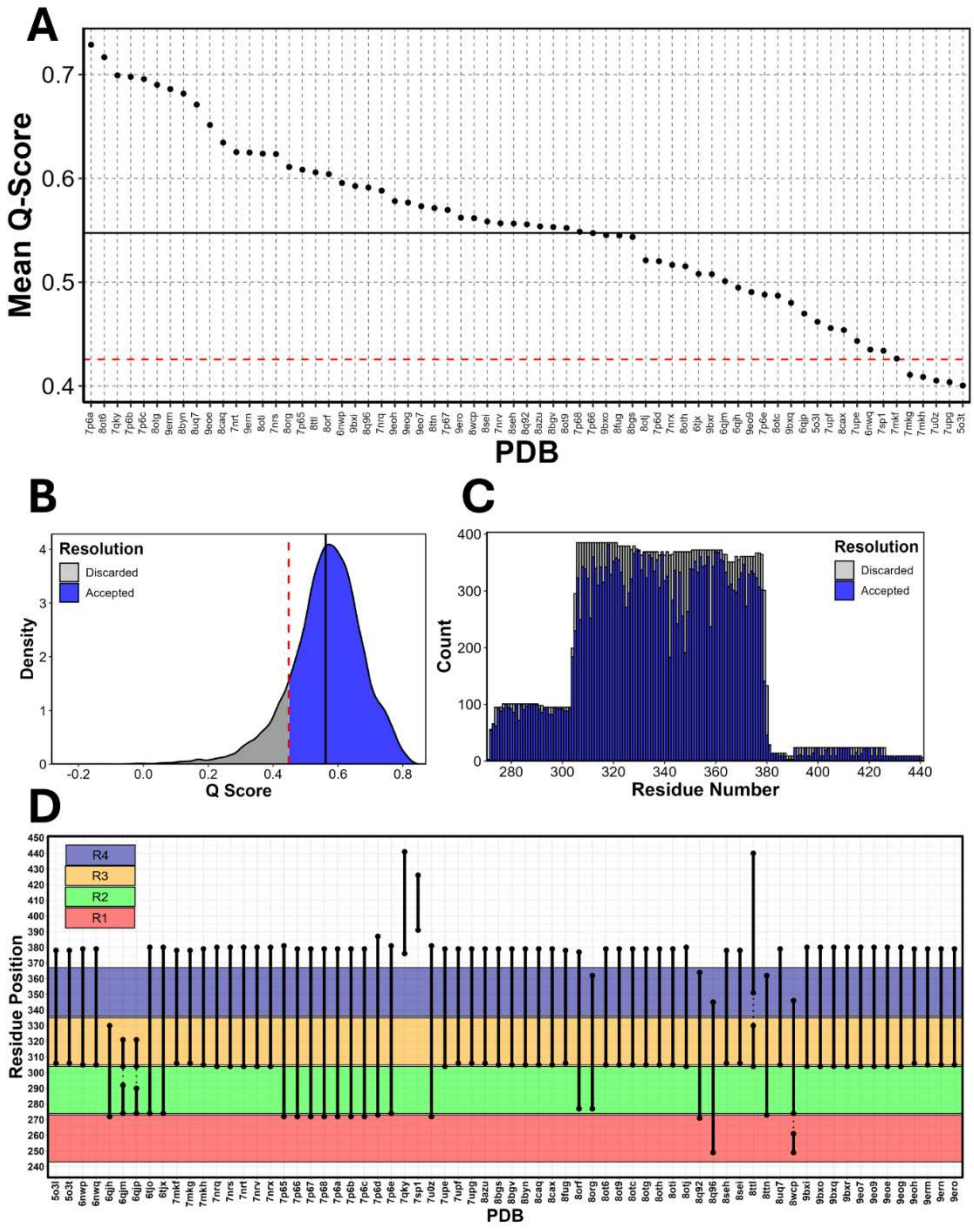


Figure S4: Q-Score filtering to remove poorly resolved structures and individual residues in 3R+4R and 4R tau amyloid structures, related to STAR Methods. **A)** The average Q-Score for each cyroEM tau structure. The mean Q-score of all structures and the mean minus one standard deviation are shown by the black and red lines, respectively. Structures with a Q-score lower than the mean - 1SD were removed from subsequent analysis. **B)** A density plot showing the per residue Q-score for the remaining structures. Again the overall mean Q-score and mean - 1SD are shown by black and red lines. Residues with a Q-score lower than the mean - 1SD were removed from subsequent analysis. **C)** A bar plot showing the number of times each residue position occurs across all structures. The bars are coloured by the number of times the given residue position is above or below the Q-score threshold. **D)** Shown are the residues resolved in the fibril core of the solved amyloid structure for each PDB. Dotted lines indicate regions of unresolved residues within the core. Residues spanning the four repeats (R1-R4) are individually coloured. Note that 4R and 3R tau can be a maximum of 441 and 410 residues respectively, with the R2

domain absent in 3R tau. Residues 1-248 are not observed at high resolution in any of the cryoEM maps.

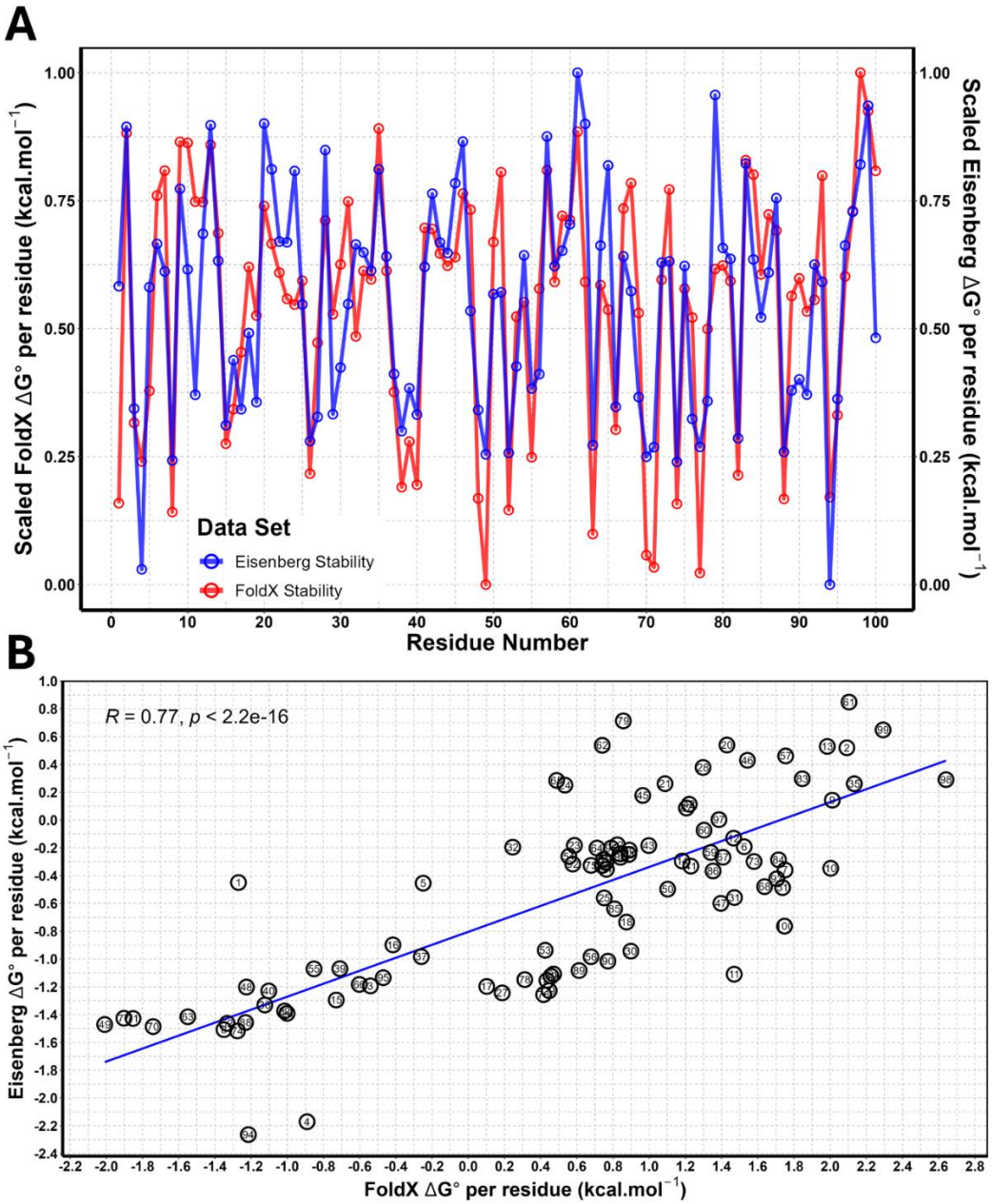


Figure S5: α-Synuclein amyloid: Comparing FoldX calculated ΔG° per residue with the Eisenberg group's solvation free energy calculations, related to STAR Methods. **A)** The average ΔG° per residue across all unique α-synuclein structures calculated using FoldX[2,3] (red) and Eisenberg's solvation free energy[4,5] (blue). Note data are scaled so that the minimum value is set to 0 and the maximum value is equal to 1. **B)** Scatter plot comparing the unscaled FoldX and Eisenberg calculated ΔG° per residue values. The Pearson correlation coefficient is shown in the top left. Numbers within each black circle denote the residue number in the 140 residue α-synuclein sequence.

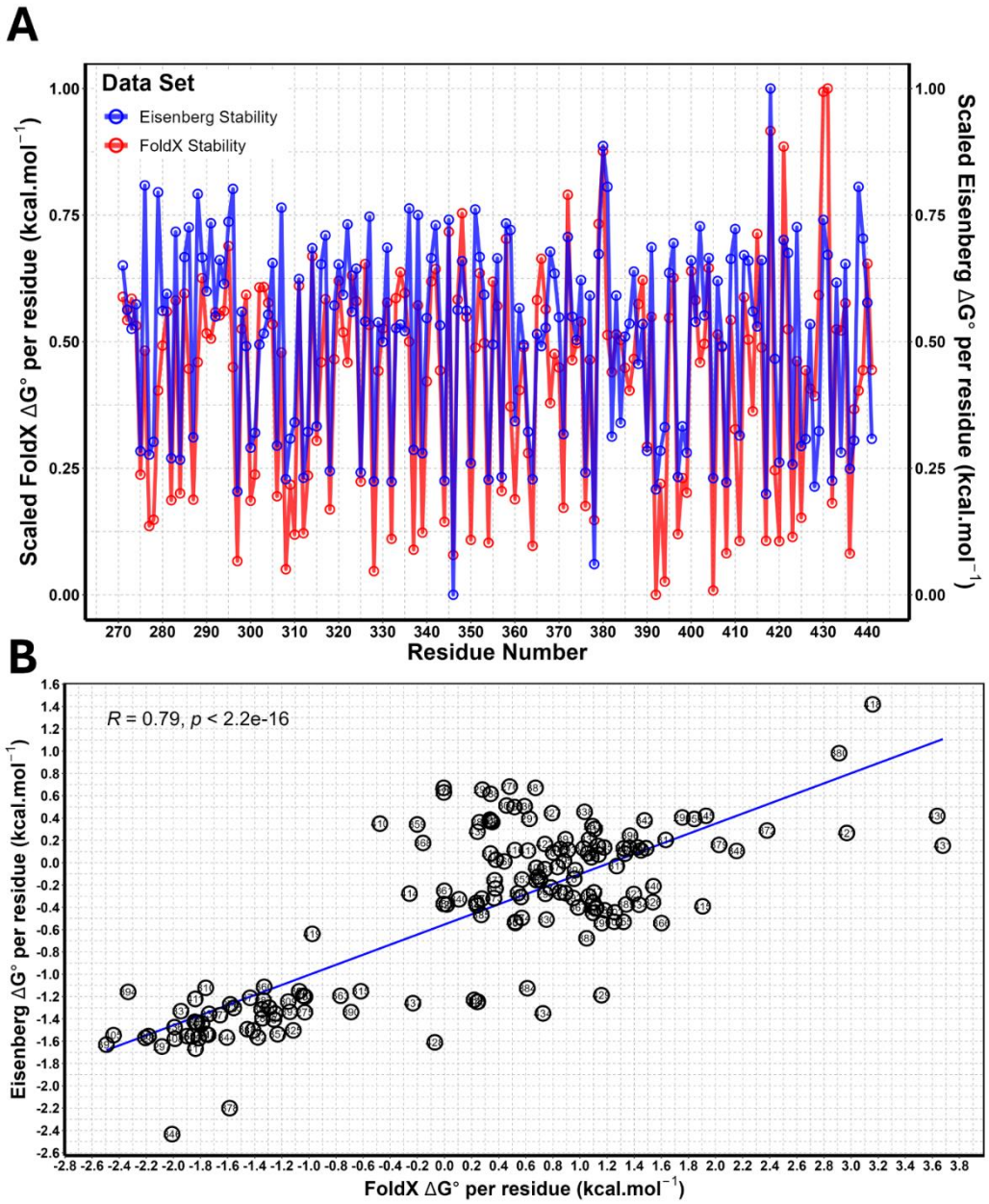


Figure S6: Tau amyloid: Comparing FoldX calculated ΔG° per residue with the Eisenberg group's solvation free energy calculations, related to STAR Methods. A) The average ΔG° per residue across all unique 3R+4R and 4R tau amyloid structures calculated using FoldX[2,3] (red) and Eisenberg's solvation free energy[4,5] (blue). Note data are scaled so that the minimum value is set to 0 and the maximum value is equal to 1. **B)** Scatter plot comparing the unscaled FoldX and Eisenberg calculated ΔG° per residue values. The Pearson correlation coefficient is shown in the top left. Numbers within each black circle denote the residue number in the 410/411 residue 3R/4R tau sequences.

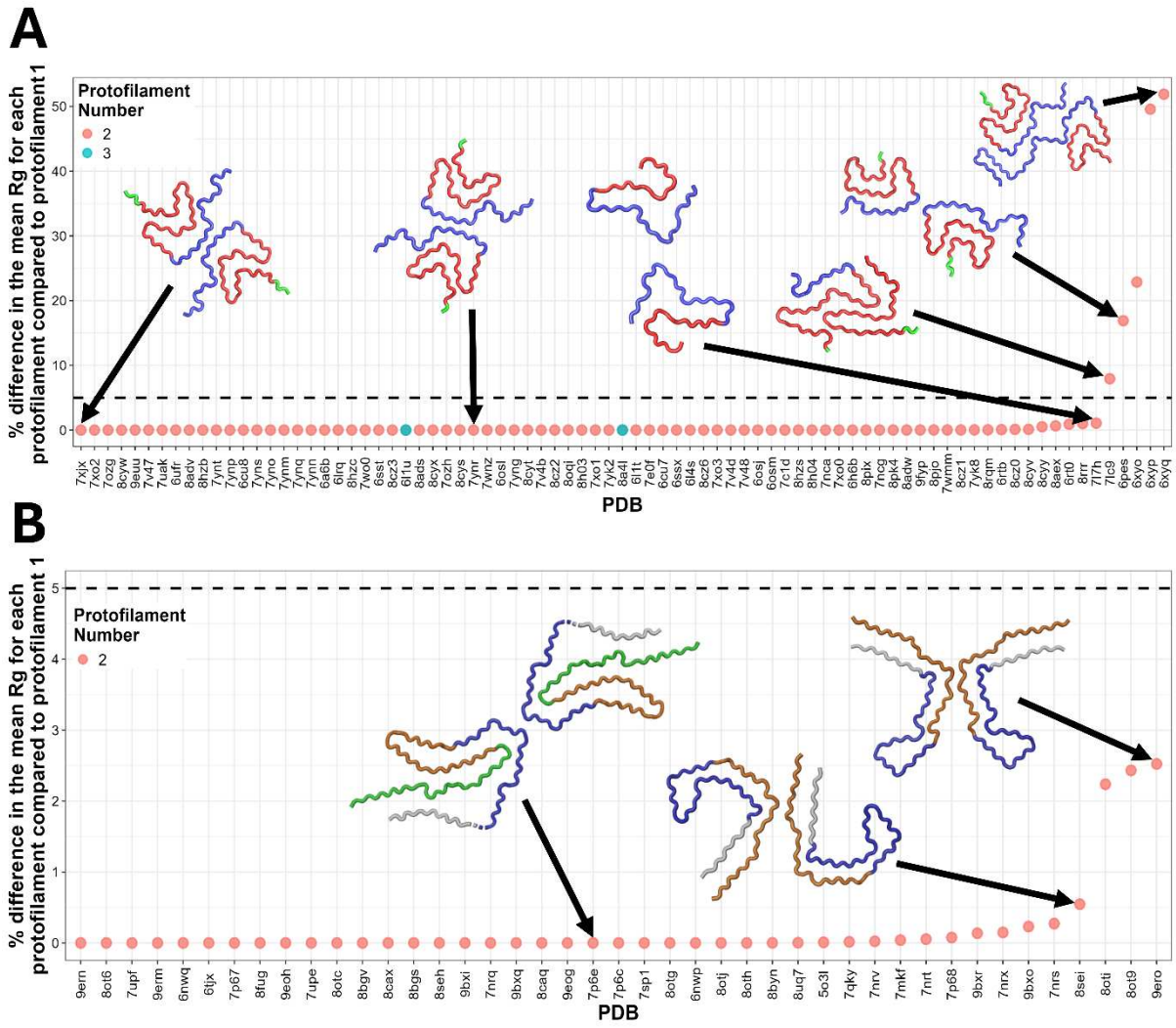


Figure S7: Intra-PDB differences between the Rg of protofilaments for α-synulcein (A) and tau (B), related to STAR Methods. Shown are the percentage differences between the Rg of protofilaments. The threshold for two protofilaments to be considered distinct was set to 5%, indicated by the black dashed line. Example asymetric unit structures for α-synuclein and Tau fibrils are shown.

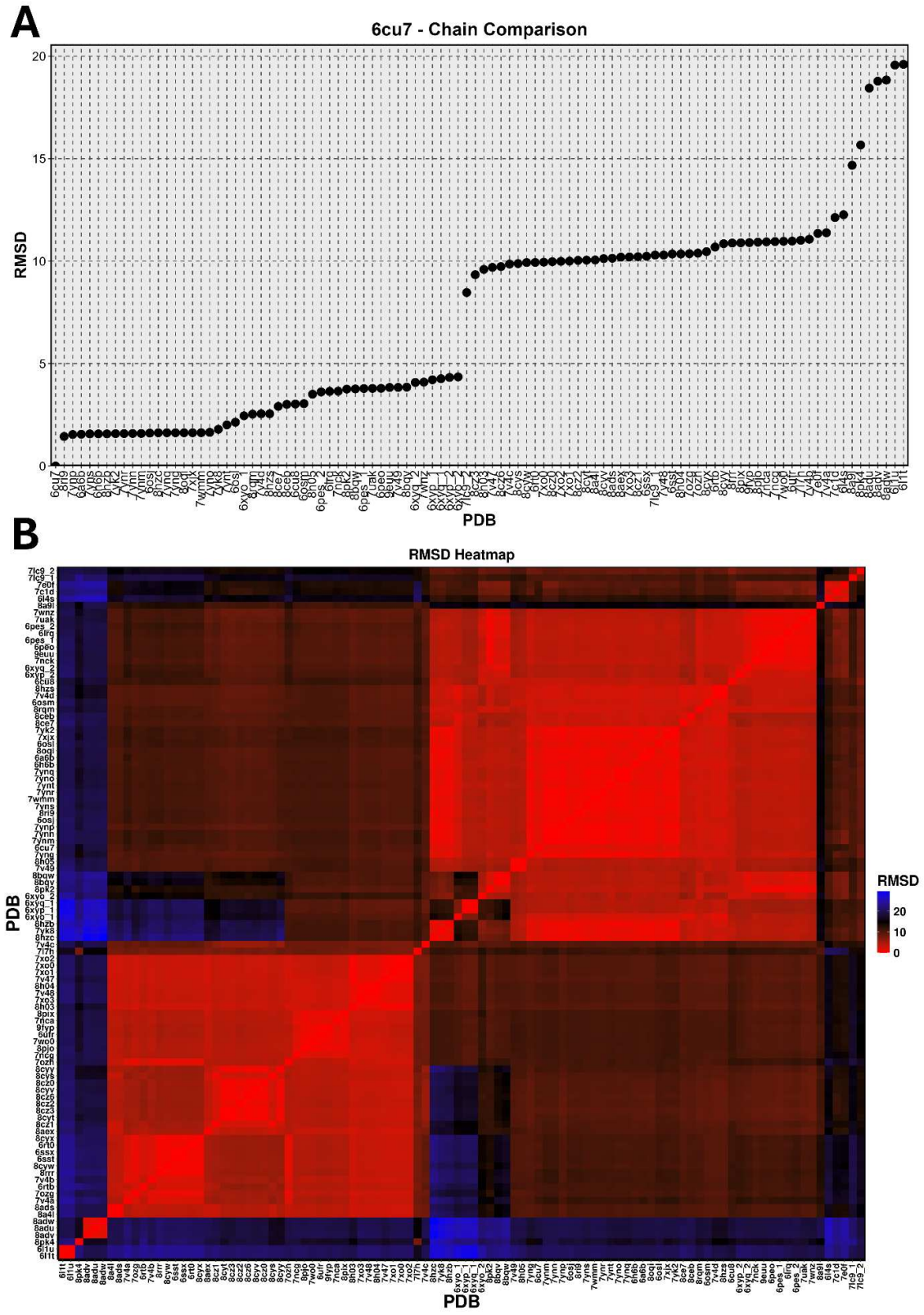


Figure S8: Using RMSD to measure the similarity of different α-synuclein amyloid structures, related to STAR Methods. A) RMSD scores (Å) of each α-synuclein amyloid fold compared with the structure 6cu7[6]. **B)** Heat map showing the pairwise RMSD comparisons across all α-synuclein amyloid folds.

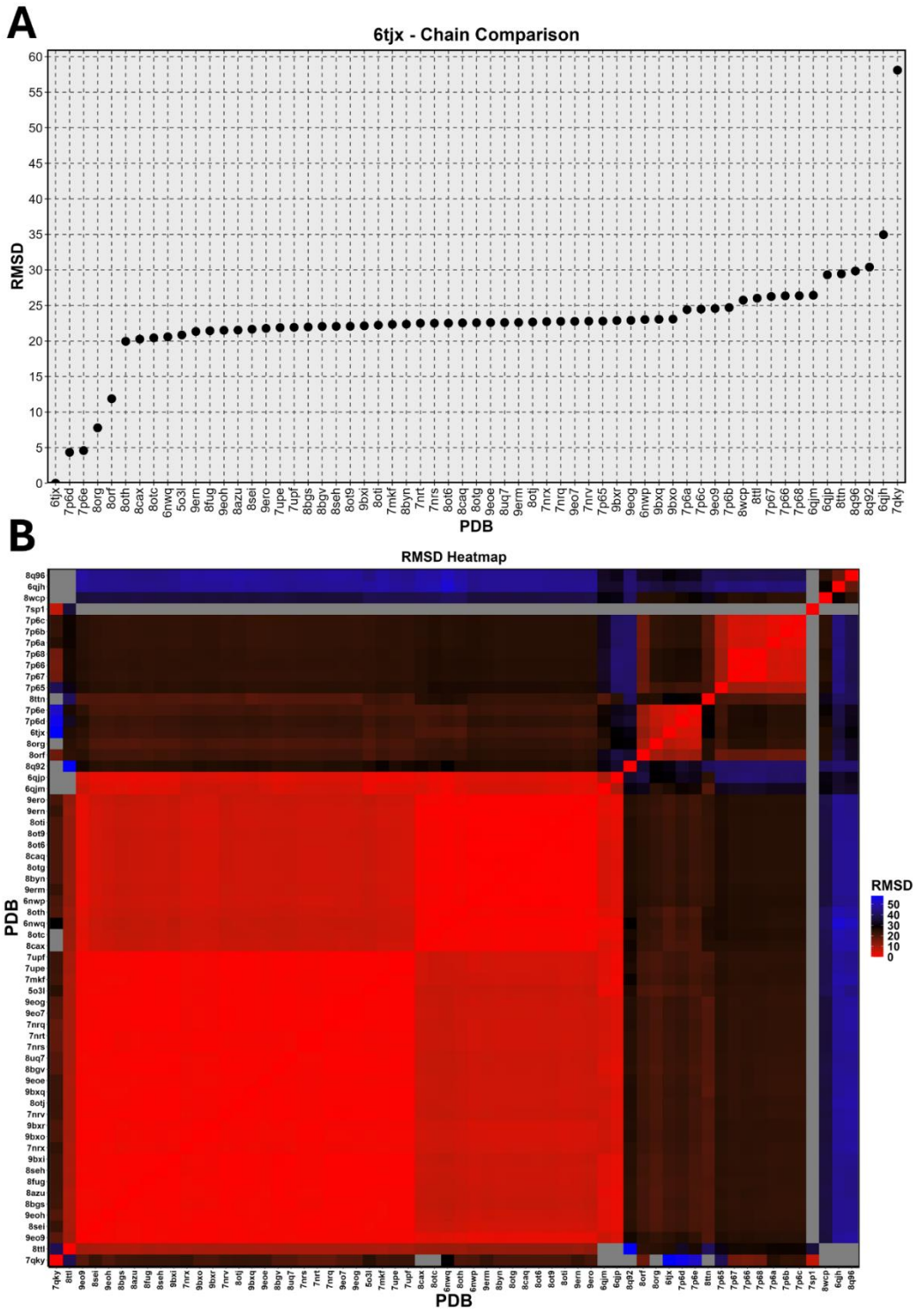


Figure S9: Using RMSD to measure the similarity of different 3R+4R and 4R tau amyloid structures, related to STAR Methods. **A)** RMSD scores (Å) of each tau amyloid fold compared with the structure 6tjx[7]. **B)** Heat map showing the pairwise RMSD comparisons across all tau amyloid folds. Grey regions correspond to structures with no overlapping residues resulting in no calculated RMSD scores.

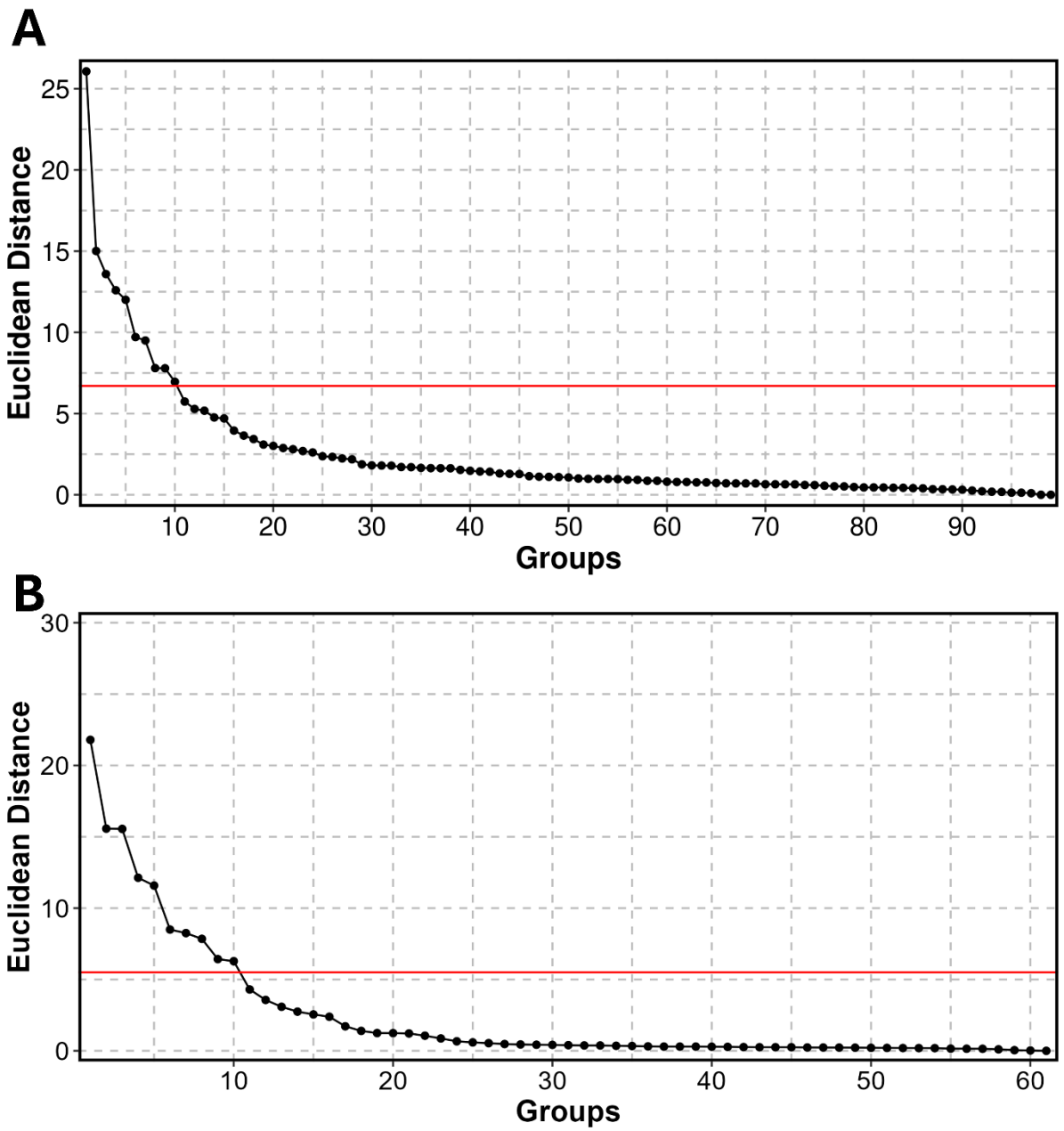


Figure S10: Scree plots showing the Euclidean distance cut height and the resulting number of cluster groups for α-synuclein (A) and tau (B) , related to STAR Methods. The manually assigned cut heights are shown by the red line.

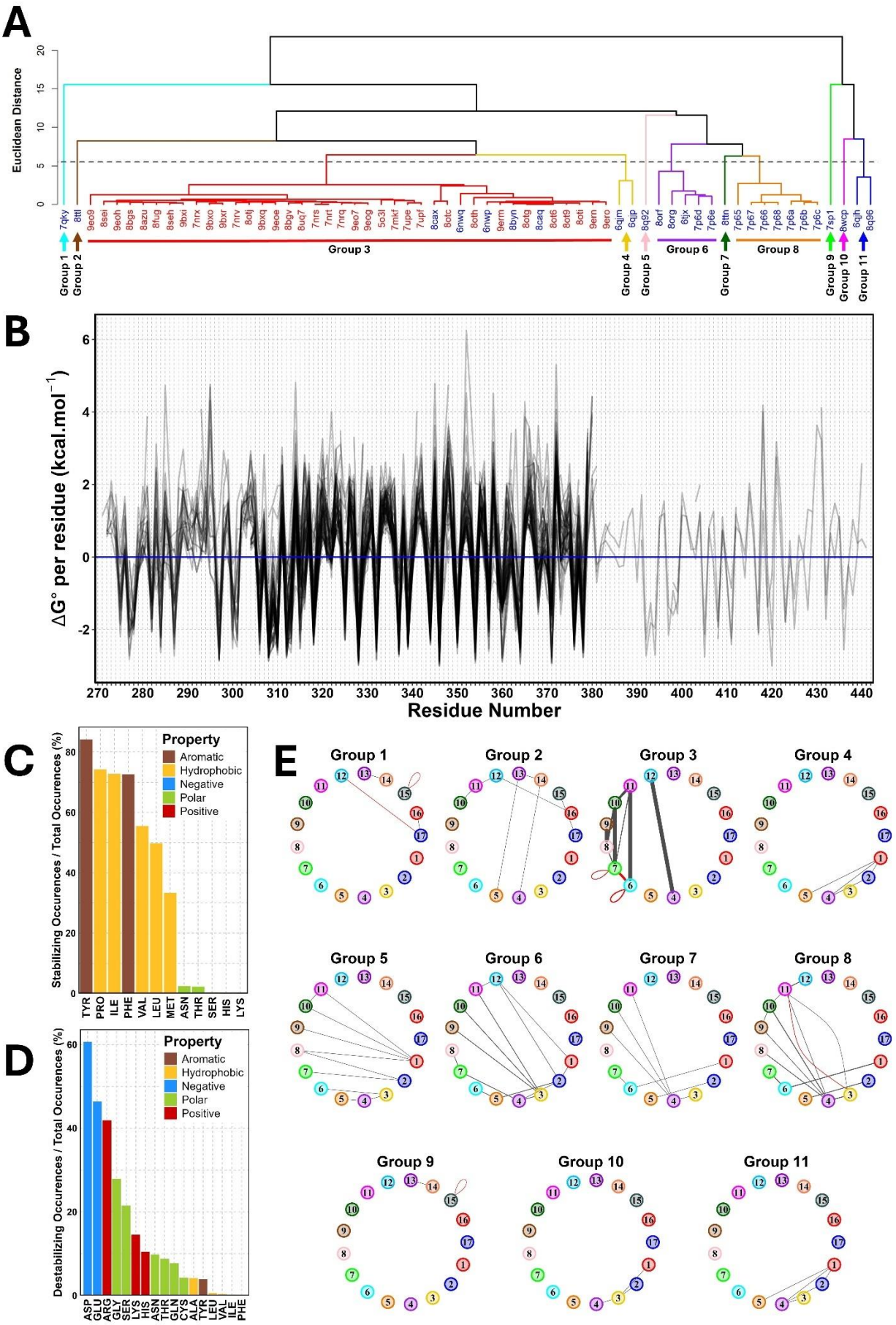

A
Euclidean Distance
0
5
10
15
20
7qky
8ttl
9eo9
8sei
9eoh
8bgs
8azu
8fug
8seh
9bxi
7nrx
9bxo
9bxr
7nrv
8olj
9bxq
9eoe
8bgv
8uq7
7nrs
7nrt
7nrq
9eo7
9eog
5o3l
7mkf
7upe
7upf
8cax
8otc
6nwq
8oth
6nwp
9erm
8byn
8otg
8caq
8ot6
8ot9
8oti
9ern
9ero
6qjm
6qjp
8q92
8orf
8org
6tjx
7p6d
7p6e
8ttn
7p65
7p67
7p66
7p68
7p6a
7p6b
7p6c
7sp1
8wcp
6qjh
8q96
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Group 9
Group 10
Group 11
B
ΔG° per residue (kcal.mol⁻¹)
6
4
2
0
-2
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
Residue Number
C
Stabilizing Occurences / Total Occurences (%)
0
20
40
60
80
Property
Aromatic
Hydrophobic
Negative
Polar
Positive
TYR
PRO
ILE
PHE
VAL
LEU
MET
ASN
THR
SER
HIS
LYS
D
Destabilizing Occurences / Total Occurences (%)
0
20
40
60
Property
Aromatic
Hydrophobic
Negative
Polar
Positive
ASP
GLU
ARG
GLY
SER
LYS
HIS
ASN
THR
CYS
ALA
TYR
LEU
VAL
ILE
PHE
E
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Group 9
Group 10
Group 11

Figure S11: Hierarchical clustering of amyloid fold similarity between solved PDB structures of tau amyloid clustered from their RMSD values, related to Figures 2 and 3. **A)** 11 structural classes result, shown in different colours. The PDB code for each structure is given below. PDB codes are coloured as follows: 4R structures in blue and 3R+4R structures in red. **B)** Distributions of ΔG° per residue for published tau amyloid structures after Q-score validation. The blue line denotes ΔG° per residue of 0, with positive values indicating a destabilizing contribution and negative values indicating overall stabilizing residues. **C,D)** Bar charts showing the normalized number of times each amino acid was found to be either **C)** stabilizing (mean - 1 SD) or **D)** destabilizing (mean + 1 SD). **E)** Network diagrams showing the number of times stabilizing regions are found within <10.8Å. Nodes represent the stabilizing regions and are coloured as in **Figure S13B**. The intra-protofilament (within a single layer) and inter-protofilament contacts are shown by black and red lines, respectively. The width of the edges indicates the number of times a contact between the stabilising regions occurs across all members of the group. The tau structures are separated into distinct polymorphs based on their cluster groups identified in **(A)**. The percentage of structures involved in each contact within a group is presented in **Table S4**.

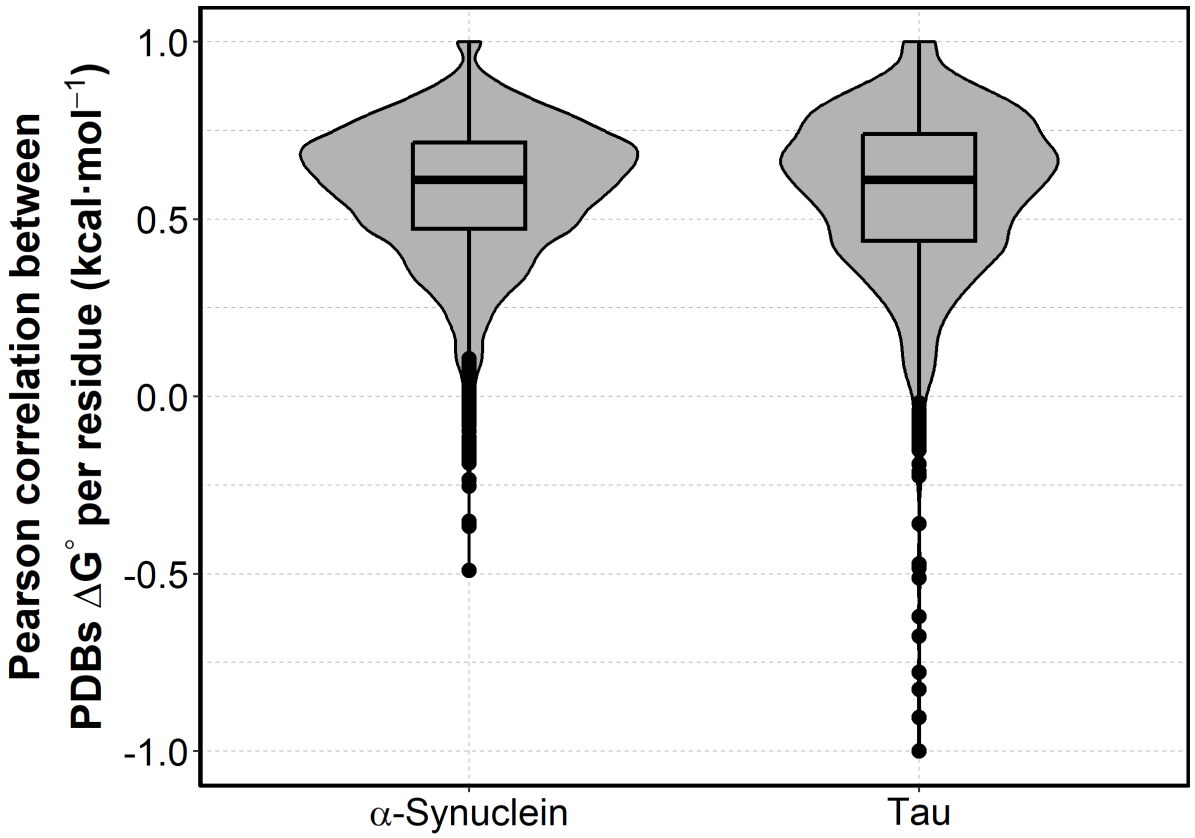


Figure S12: Violin plots showing the Pearson correlation between α-synuclein and Tau ΔG° per residue profiles, related to Figure 3. Pearson correlations were calculated between ΔG° per residue profiles traces in a pairwise fashion for every unique amyloid fold. Boxplots display the median and interquartile range for each group.

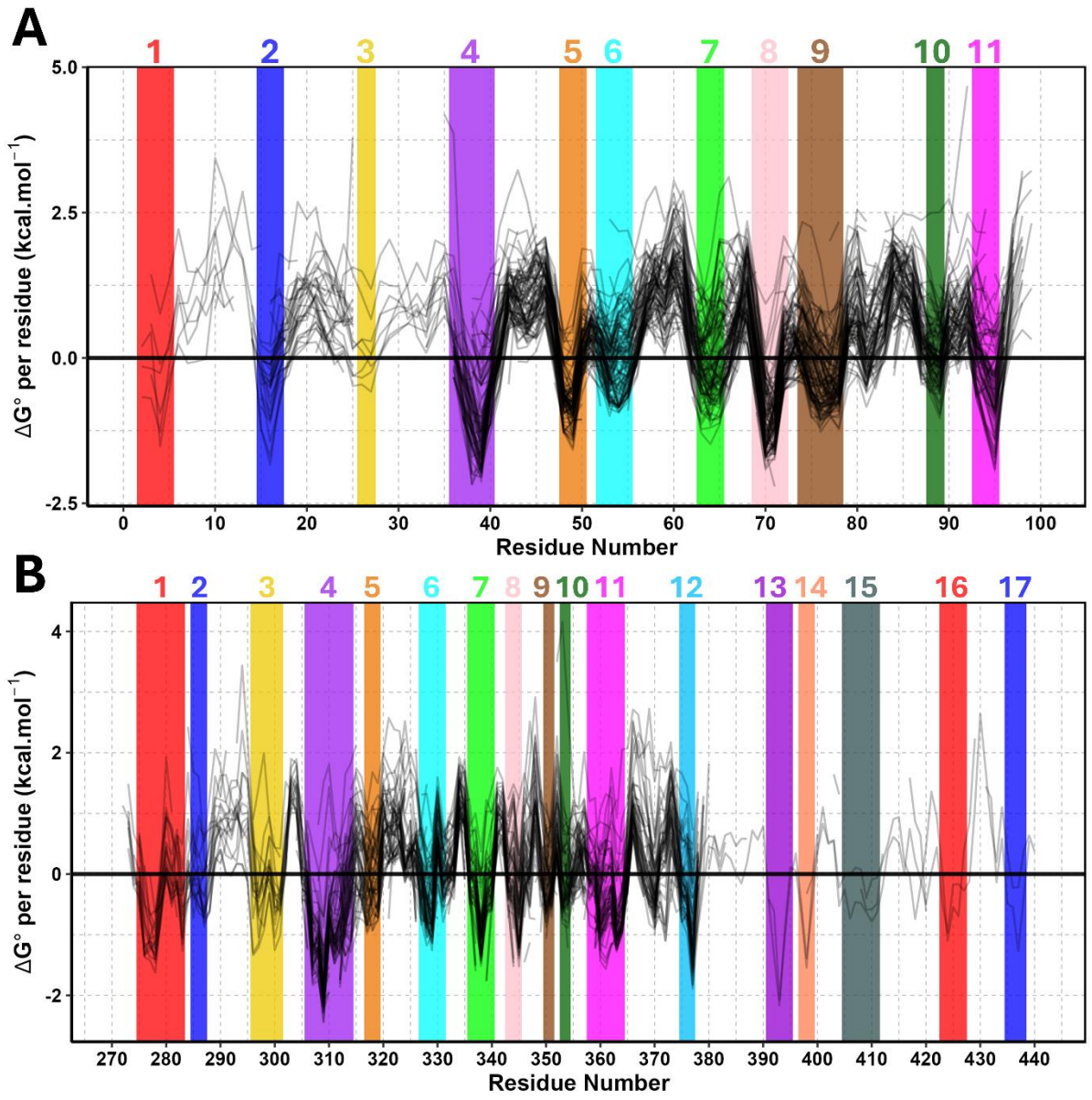


Figure S13: Defining stabilizing regions in amyloid structures related to Figure 3 and STAR methods. The thermodynamic profiles of **A)** α-synuclein and **B)** tau obtained using FoldX were smoothed using a sliding window of size 3. Stabilizing regions are indicated by the vertical coloured bars.

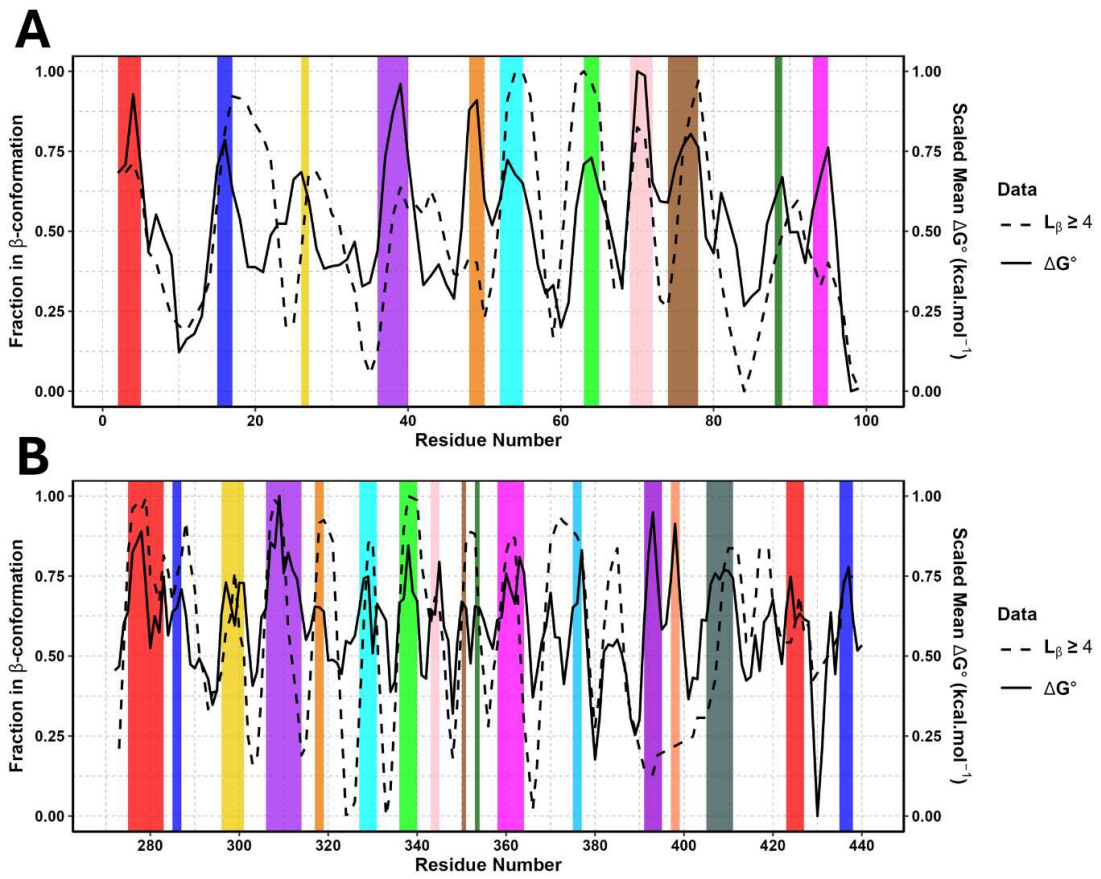


Figure S14: Comparing stability and β-strand propensity for α-synuclein (A) and 3R+4R or 4R tau (B), realted to STAR methods. The dashed line represents the number of times a given residue was found in a β-strand consisting of at least 4 residues as a proportion of the total number of occurrences for each residue across all analysed PDBs. The solid line represents the mean ΔG° per residue across all analysed PDBs which is then scaled so the minimum (most destabilizing) and maximum value (the most stabilizing) equal 0 and 1, respectively. The coloured bars represent the defined stabilizing regions as in **Figure S11**.

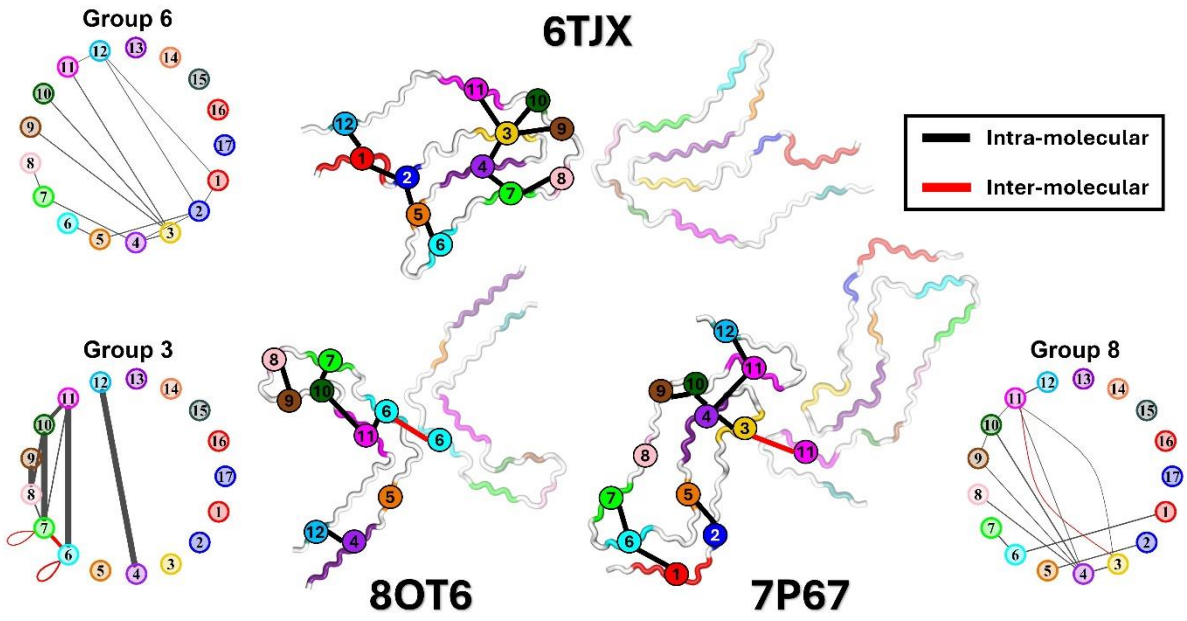


Figure S15: Visualization of the network diagrams for example tau structures, related to Figure 4. Shown are three example structures; 8OT6[8], 6TJX[7] and 7P67[9] taken from the three largest cluster groups (**Figure S11A**), group 3, 6 and 8, respectively. Network diagrams showing the number of times contacting residues are found within <10.8Å. Nodes represent single residues and are coloured based on the stabilizing regions identified in **Figure S13B**. All structures are formed from two protofilaments. For each structure, a single protofilament has been labelled with its stabilizing regions denoted by small numbered circles and contacts within <10.8Å shown as solid lines. For both the network diagrams and the annotated structures, the intra- and inter-protofilament contacts (per layer) are shown by black and red lines, respectively. For simplicity, the second protofilament for each structure is unlabelled except for instances of inter-molecular contacts.

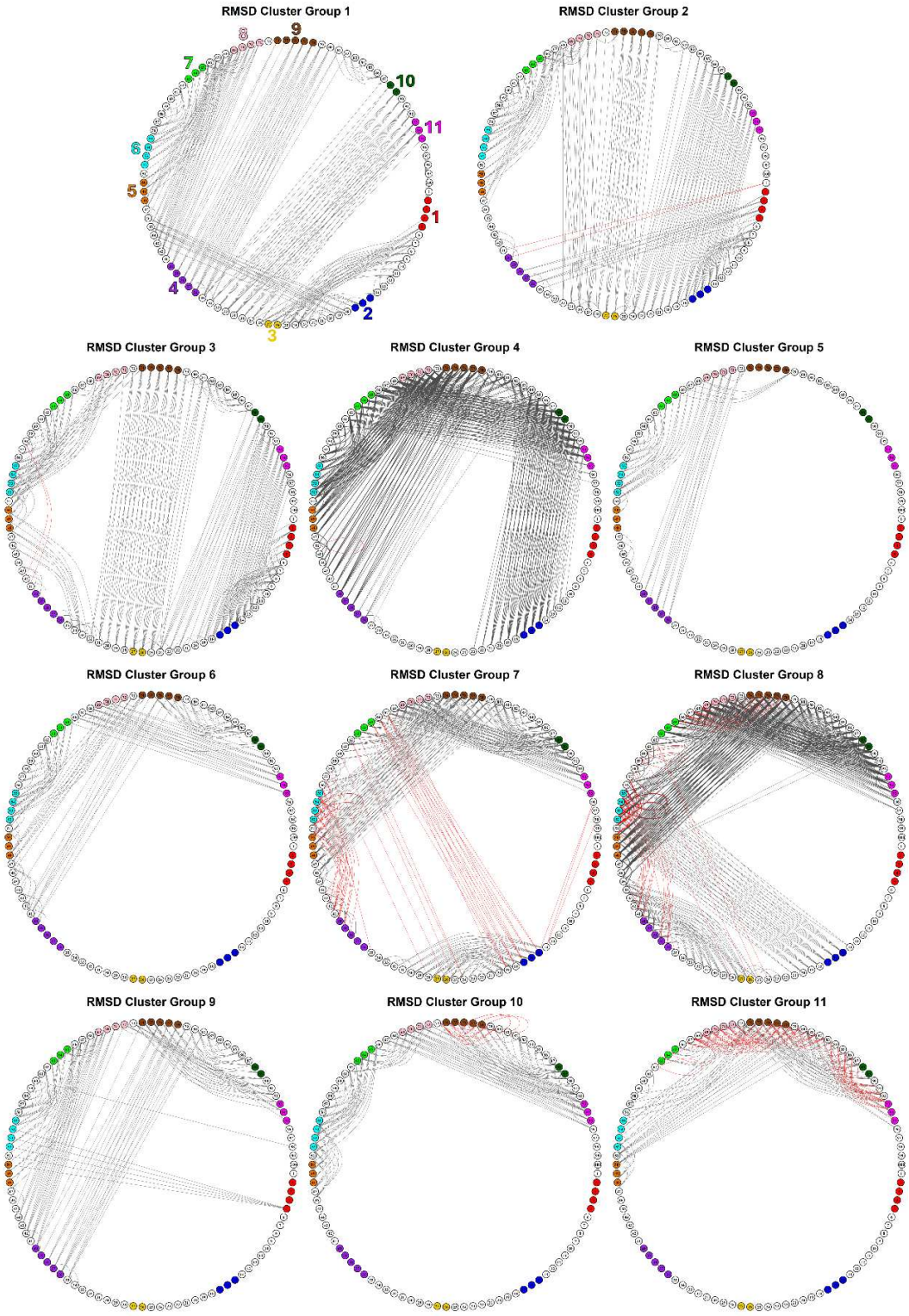


Figure S16: All-residue contact maps of different α-synuclein amyloid polymorphs, related to Figure 4. Network diagrams showing the number of times residues are found within <10.8Å. Nodes represent single residues and are coloured (and numbered for one group) based on the stabilizing regions identified in **Figure S13A**. Grey edges denote intra-protofilament interactions within a layer, with red edges indicating inter-protofilament interactions. The α-synuclein structures are separated into distinct polymorphs based on the cluster groups identified in **Figure 2**.

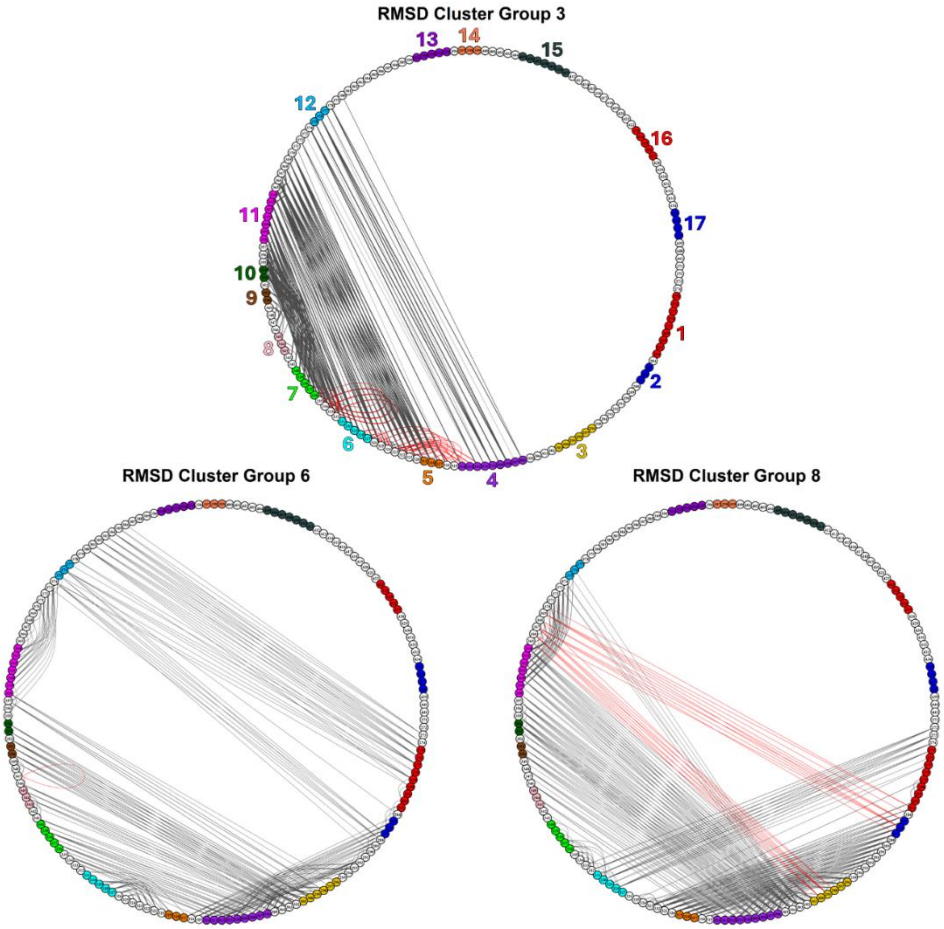


Figure S17: All-residue contact maps of different tau amyloid polymorphs, related to Figure 4. Network diagrams showing the number of times residues are found within <10.8Å for the 3 largest tau cluster groups. Nodes represent single residues and are coloured (and numbered for one group) based on the stabilizing regions identified in **Figure S13B**. Grey edges denote intra-protofilament interactions per layer with red edges indicating inter-protofilament interactions. The tau structures are separated into distinct polymorphs based on the cluster groups identified in **Figure S11A**.

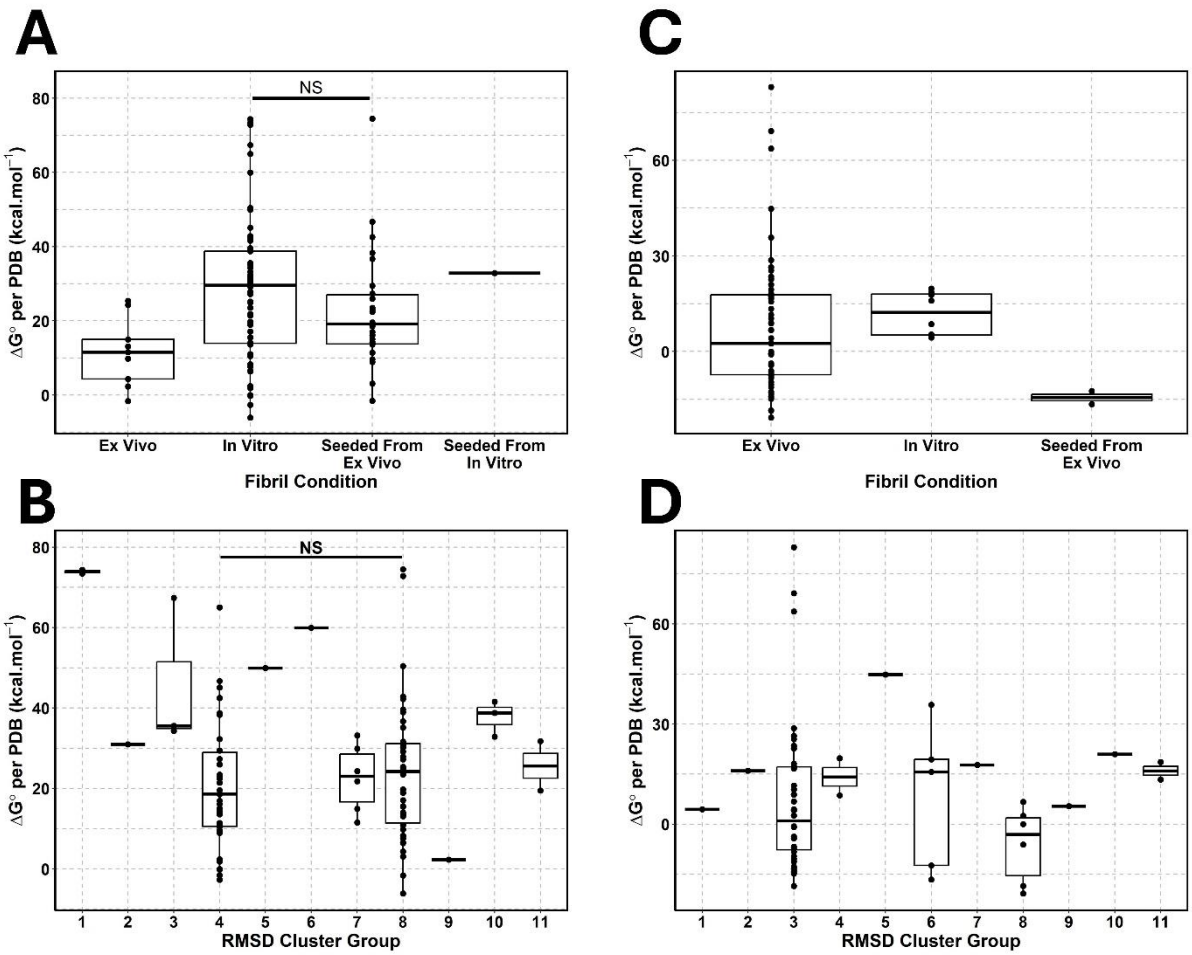


Figure S18: Thermodynamic stability of α-synuclein and tau amyloid fibrils formed under different conditions, related to Figure 5. Data are clustered by **A,C)** growth condition and **B,D)** RMSD cluster group. Box-and-whisker plots show the mean and standard deviation for each group. Individual points represent the ΔG° for a single PDB calculated by summing the mean ΔG° per residue score for each position across all interior chains. **A,B)** data for α-synuclein amyloid fibrils, **C,D)** data for tau fibrils. Statistical analysis was not performed due to small numbers of entries in some classes and errors in calculating stability from structures of different resolution that limits statistical power. However, two pairs of groupings of α-syn contain sufficient numbers for statistical analysis: fibrils generated completely in vitro (64 members) and those seeded from ex vivo fibrils (26 members) and RMSD cluster groups 4 and 8 (38 and 42 members, respectively). As both in vitro and group 8 data are not normally distributed (assessed using a Shapiro-Wilk test) a non-parametric Wilcoxon rank sum test was used and no significant difference was found for either pair ($W = 988$, p-value = 0.166 and $W = 705$, p-value = 0.375 for in vitro and seeded ex vivo and groups 4 and 8, respectively).

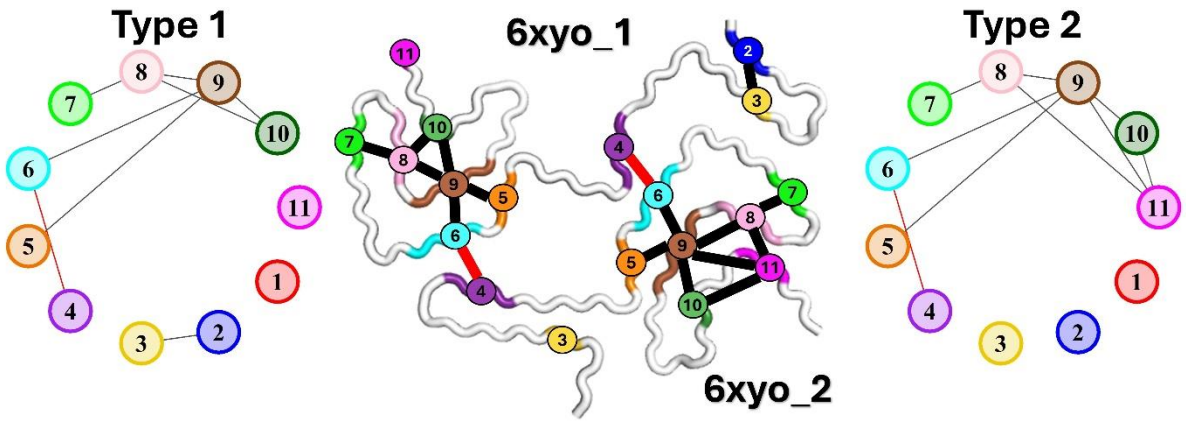
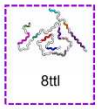


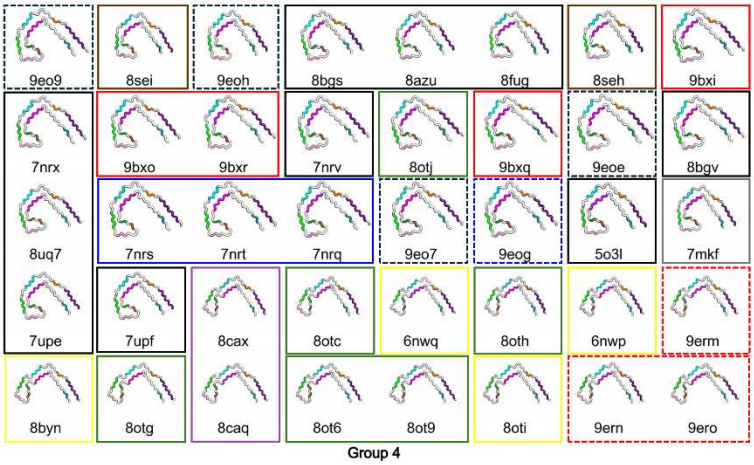
Figure S19: Visualization of the α-synuclein ex vivo MSA structures with a network diagram for each topology, related to Figure 5. A single layer of the 6xyo[15] fibril has been labelled with its stabilizing regions denoted by small, numbered circles and contacts within <10.8Å shown as solid lines. For both the network diagrams and the annotated structures, the intra-protofilament (within a single layer) and inter-protofilament contacts are shown by black and red lines, respectively. Separate network diagrams are shown to display the different stabilizing region contacts within each of the two protofibrils.

Group 1

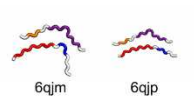
Group 2



Group 3

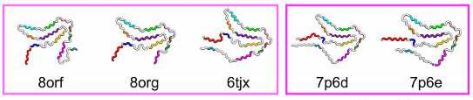


Group 4



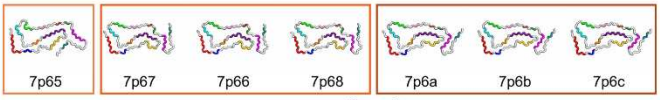
Group 5

Group 6



Group 7

Group 8



Group 9

Group 10

Group 11

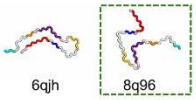


Figure S20: High-resolution cryo-EM structures of tau amyloid grouped into RMSD clusters and coloured by stable regions, related to Figure 5. Solid line boxes denote structures solved from ex vivo samples with Alzheimer's Disease in black, Down syndrome in red, Alzhiemer's disease + Down Syndrome in brown, Cerebral Amyloid Angiopathy in grey, Age-Related Tauopathy in blue, Amyotrophic Lateral Sclerosis/Parkinsonism-dementia complex in green, Chronic Traumatic Encephalopathy in yellow, Subacute Sclerosing Panencephalitis in purple, Corticobasal Degeneration in pink, Argyrophilic Grain Disease in dark pink, Progressive Supranuclear Palsy in light orange, Globular Glial Tauopathy in orange, Limbic-predominant neuronal inclusion body tauopathy in dark orange. Dashed line boxes denote structures solved from point mutant samples with V337M in black, R406W in blue, D395G in red, P301S in green, P301L in brown, S396E + T403E + S404E in orange, S202E + T205E + S208E in purple. The order in which structures are shown in each group is based on their similarity (RMSD), with the order matching that shown in the dendrogram in **Figure S11A** (from left to right). Stable regions are coloured as shown in **Figure S13B**.

Table S4: Tau contact frequency between stable regions for each RMSD cluster group, related to Figure 3 and STAR methods

RMSD cluster group	Stable Regions Within (<10.8Å)	Occurrences	Unique Structures per Group	Contact Frequency (%)
1	12 -- 17	1	1	100
1	13 -- 14	1	1	100
1	15 -- 15	1	1	100
1	16 -- 17	1	1	100
2	10 -- 11	1	1	100
2	11 -- 12	1	1	100
2	12 -- 16	1	1	100
2	13 -- 14	1	1	100
2	15 -- 17	1	1	100
2	4 -- 14	1	1	100
2	5 -- 13	1	1	100
3	10 -- 11	25	40	63
3	4 -- 12	40	40	100
3	6 -- 11	40	40	100
3	6 -- 6	9	40	23
3	6 -- 7	21	40	53
3	7 -- 10	40	40	100
3	7 -- 11	7	40	18
3	7 -- 7	5	40	13
3	7 -- 8	9	40	23
3	8 -- 10	16	40	40
3	8 -- 9	34	40	85
4	1 -- 2	1	2	50
4	1 -- 4	2	2	100
4	1 -- 5	2	2	100
4	2 -- 4	2	2	100
5	1 -- 10	1	1	100

5	1 -- 11	1	1	100
5	1 -- 8	1	1	100
5	1 -- 9	1	1	100
5	10 -- 11	1	1	100
5	2 -- 7	1	1	100
5	2 -- 8	1	1	100
5	3 -- 4	1	1	100
5	3 -- 6	1	1	100
5	4 -- 5	1	1	100
6	1 -- 12	1	5	20
6	1 -- 2	3	5	60
6	11 -- 12	1	5	20
6	2 -- 12	2	5	40
6	2 -- 4	2	5	40
6	2 -- 5	5	5	100
6	3 -- 10	5	5	100
6	3 -- 11	5	5	100
6	3 -- 4	5	5	100
6	3 -- 9	5	5	100
6	4 -- 7	5	5	100
6	5 -- 6	5	5	100
6	7 -- 8	5	5	100
7	1 -- 6	1	1	100
7	2 -- 5	1	1	100
7	3 -- 4	1	1	100
7	4 -- 10	1	1	100
7	4 -- 11	1	1	100
7	4 -- 9	1	1	100
7	6 -- 7	1	1	100
8	1 -- 6	7	7	100
8	10 -- 11	1	7	14
8	11 -- 12	6	7	86

8	2 -- 5	6	7	86
8	3 -- 11	4	7	57
8	3 -- 4	7	7	100
8	4 -- 10	7	7	100
8	4 -- 11	3	7	43
8	4 -- 8	5	7	71
8	4 -- 9	4	7	57
8	6 -- 7	7	7	100
8	9 -- 10	3	7	43
9	13 -- 14	1	1	100
9	15 -- 15	1	1	100
10	1 -- 3	1	1	100
10	2 -- 3	1	1	100
10	3 -- 4	1	1	100
11	1 -- 2	1	2	50
11	1 -- 4	1	2	50
11	1 -- 5	1	2	50
11	2 -- 4	1	2	50
11	3 -- 4	1	2	50

References

1. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R., and Eisenberg, D. S. Amyloid Atlas (2024). https://people.mbi.ucla.edu/sawaya/amyloidatlas/.

2. Guerois, R., Nielsen, J. E., and Serrano, L. (2002). Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. J Mol Biol *320*, 369–387. doi:10.1016/S0022-2836(02)00442-4.

3. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Research *33*, 382–388. doi:10.1093/nar/gki387.

4. Eisenberg, D., and McLachlan, A. D. (1986). Solvation energy in protein folding and binding. Nature *319*, 199-203. doi:10.1038/319199a0.

5. Eisenberg, D., Wesson, M., and Yamashita, M. (1986). Interpretation of protein folding and binding with atomic solvation parameters. Chemica Scripta *29A*, 217–221.

6. Li, Y., Zhao, C., Luo, F., Liu, Z., Gui, X., Luo, Z., Zhang, X., Li, D., Liu, C., and Li, X. (2018). Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. Cell Research *28*, 897-903. doi:10.1038/s41422-018-0075-x.

7. Zhang, W., Tarutani, A., Newell, K. L., Murzin, A. G., Matsubara, T., Falcon, B., Vidal, R., Garringer, H. J., Shi, Y., Ikeuchi, T., Murayama, S., Ghetti, B., Hasegawa, M., Goedert, M., and Scheres, S. H. W. (2020). Novel tau filament fold in corticobasal degeneration. Nature *580*, 283-287. doi:10.1038/s41586-020-2043-0.

8. Qi, C., Verheijen, B. M., Kokubo, Y., Shi, Y., Tetter, S., Murzin, A. G., Nakahara, A., Morimoto, S., Vermulst, M., Sasaki, R., Aronica, E., Hirokawa, Y., Oyanagi, K., Kakita, A., Ryskeldi-Falcon, B., Yoshida, M., Hasegawa, M., Scheres, S. H. W., and Goedert, M. (2023). Tau filaments from amyotrophic lateral sclerosis/parkinsonism-dementia complex adopt the CTE fold. Proc Natl Acad Sci USA *120*, e2306767120. doi:10.1073/pnas.2306767120.

9. Shi, Y., Zhang, W., Yang, Y., Murzin, A. G., Falcon, B., Kotecha, A., van Beers, M., Tarutani, A., Kametani, F., Garringer, H. J., Vidal, R., Hallinan, G. I., Lashley, T., Saito, Y., Murayama, S., Yoshida, M., Tanaka, H., Kakita, A., Ikeuchi, T., Robinson, A. C., Mann, D. M. A., Kovacs, G. G., Revesz, T., Ghetti, B., Hasegawa, M., Goedert, M., and Scheres, S. H. W. (2021). Structure-based classification of tauopathies. Nature *598*, 359–363. doi:10.1038/s41586-021-03911-7.