Supplementary Materials for Exploring the Interplay Between Fibrillization and Amorphous Aggregation Channels on the Energy Landscapes of Tau Repeat Isoforms

Xun Chen[1,3,6], Mingchen Chen[1,2,6], Nicholas P Schafer[1,4], and Peter G. Wolynes*[1,3,4,5]

[1]Center for Theoretical Biological Physics, Rice University, Houston, TX
[2]Department of Bioengineering, Rice University, Houston, TX
[3]Department of Chemistry, Rice University, Houston, TX
[4]Department of Physics and Astronomy, Rice University, Houston, TX
[5]Department of Biosciences, Rice University, Houston, TX
[6]These two authors contributed equally to this work
*Email: pwolynes@rice.edu, Phone: (713)348-4101

correspondence to: pwolynes@rice.edu

Contents

S1 Supplementary Methods

S1.1 Simulation details using AWSEM.

AWSEM is a predictive coarse-grained protein folding force field. The parameters of the force field were optimized using the energy landscape theory starting from structural information on folded proteins[1]. Readers are referred to Davtyan et al[2] for the detailed construction of AWSEM. The details of protein aggregation simulation using AWSEM may be found in Chen et al[3] and Zheng et al[4]. There the reader will find the descriptions of how AWSEM is used to compute aggregation landscapes of peptides and how we use physical cluster theory from statistical mechanics to infer the grand canonical free energy profiles for a range of monomer concentrations. In this study, all the umbrella sampling simulations for multiple protein chains were performed at 300K for 8 million steps. 8 million steps corresponds to roughly 50 μs in laboratory time for the AWSEM force field. At the nominal simulation concentration (around 80 μM), this time is long enough to ensure reasonable convergence of sampling on this system when using umbrella sampling. The initial configurations used for the umbrella sampling were six monomers randomly distributed over a cubic box of size 500 Angstroms.

S1.2 Definition of hydrogen bonds.

The definition of hydrogen bonds can be found in Davtyan et al[2]. In general, we define the propensity of forming hydrogen bonds between residue pairs as below:

$$\theta_{i,j} = \exp[-\frac{(r_{ij}^{ON} - < r^{ON} >)^2}{2\sigma_{NO}^2} - \frac{(r_{ij}^{OH} - < r^{OH} >)^2}{2\sigma_{HO}^2}]$$

$\theta_{i,j}$ signifies the propensity for the formation of a hydrogen bond between residue i and residue j. In this study, we use the cutoff 0.8 to determine the formation of one hydrogen bond. r_{ij}^{ON} is the distance from the carbonyl oxygen of residue i to the nitrogen of residue j, and r_{ij}^{OH} is the distance from carbonyl oxygen of residue i to the backbone amide hydrogen of residue j. r^{ON} and r^{OH} are corresponding equilibrium bond lengths. σ_{NO}^2 and σ_{NO}^2 are the covariances of the two distances.

S1.3 Free energy perturbation to quantify the effects of phosphorylation.

The thermodynamic effects of phosphorylation on energy landscapes have been studied using the AMH model[5, 6], which was a predecessor of the present AWSEM force field. The phosphorylated serines and threonines were simulated as being "super-charged" glutamic acid (Glu) residues. We adopted the same strategy here, where both the phosphorylated Ser (serine) and phosphorylated Thr(threonine) were modeled as supercharged Glus. Since the pKas of phospho-serine and phospho-threonine have been determined to be around 7 in experiments,

we simulated the phosphorylated residues as having an electrostatic charge of -1.5e. To approximate the ensemble of phosphorylation forms that are found in vivo, we randomly phosphorylated three sites on each monomer by choosing three residues from the available serines and threonines. A higher probability of being selected (as compared to the probabilities that were used for the other serines and threonines) was used for residue Serine 262, which is known from experiments to be phosphorylated with a higher probability than the other residues. This procedure was repeated 5 times to create 5 different mixtures of phosphorylated monomers, and the free energy perturbation calculations described below were carried out individually for each of the 5 mixtures. The mixtures of sequences that were finally used are shown in the SI.

In order to model efficiently the effects of phosphorylation on the aggregation of tau repeats, the ensemble average value of the phosphorylation induced perturbation to the AWSEM hamiltonian for the mixtures is added to the free energy profile for the unphosphorylated versions of the proteins that had already been calculated using the weighted histogram analysis method (WHAM). The resulting free energy profile is correct to first order in the perturbation for the mixtures. By using the first order approximation, we avoid considering how the structural features of the mixture are changed by adding charges, which would entail a large number of additional simulations for many varying compositions.

S2 Supplementary Results

S2.1 Landscapes for the growth of fibrillar/prefibrillar species in both unphosphorylated and phosphorylated R3/4.

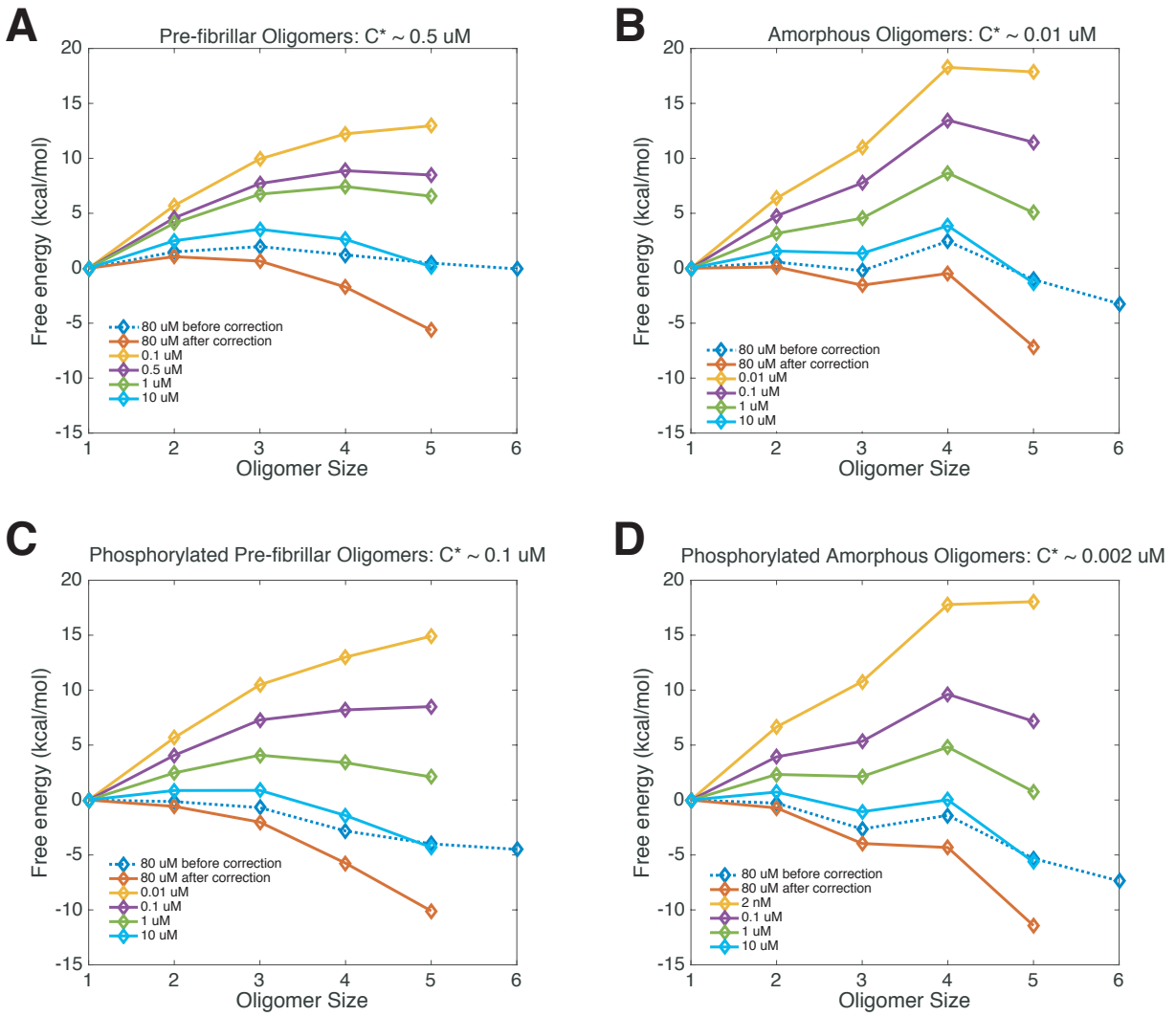

Figure S1: Grand canonical free energy profiles for the growth of different oligomeric species. The grand canonical free energy, $F - n\mu$, as a function of oligomer size and corrected for the monomer concentration changes in the fixed number simulation allows one to infer the critical value of the concentration of free monomers. (A) The 1D aggregation profiles for the growth of pre-fibrillar species in unphosphorylated R3/4. (B) The 1D aggregation profiles for the growth of off-pathway species in unphosphorylated R3/4. (C) The 1D aggregation profiles for the growth of pre-fibrillar species in phosphorylated R3/4. (D) The 1D aggregation profiles for the growth of off-pathway species in phosphorylated R3/4. The uncorrected free energy profiles are shown in blue dashed lines. The solid lines represent the aggregation profiles after accounting for the finite size effect and extrapolating to concentrations other than the one at which the simulation was performed. The last data point ($n = 6$) is the most uncertain because of the sparse sampling of when the initial monomer pool is maximally depleted.

S2.2 Landscapes for the growth of fibrillar/prefibrillar species in both unphosphorylated and phosphorylated R1/3/4.

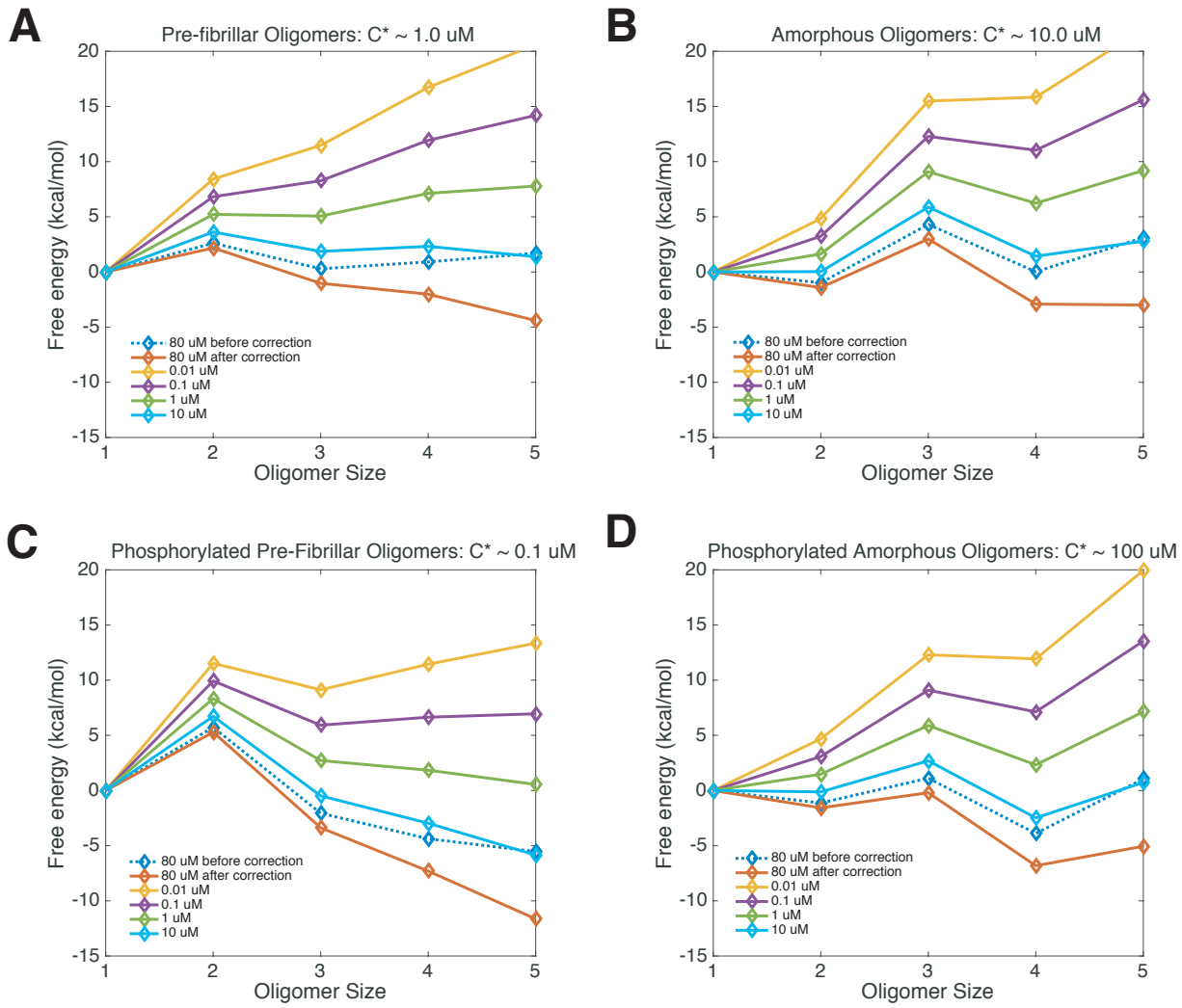

Figure S2: Grand canonical free energy profiles for the growth of different oligomeric species. The grand canonical free energy, $F - n\mu$, as a function of oligomer size and corrected for the monomer concentration changes in the fixed number simulation allows one to infer the critical value of the concentration of free monomers. (A) The 1D aggregation profiles for the growth of pre-fibrillar species in unphosphorylated R1/3/4. (B) The 1D aggregation profiles for the growth of off-pathway species in unphosphorylated R1/3/4. (C) The 1D aggregation profiles for the growth of pre-fibrillar species in phosphorylated R1/3/4. (D) The 1D aggregation profiles for the growth of off-pathway species in phosphorylated R1/3/4. The uncorrected free energy profiles are shown in blue dashed lines. The solid lines represent the aggregation profiles after accounting for the finite size effect and extrapolating to concentrations other than the one at which the simulation was performed. The last data point ($n = 6$) is the most uncertain because of the sparse sampling of when the initial monomer pool is maximally depleted.

S2.3 Probing the effects of phosphorylations on the aggregation landscapes of R3/4.

A

```
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF

VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF

VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF

VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF

VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
VQIVYKPVDL SKVTSKCGSL GNIHHKPGGG QVEVKSEKLD FKDRVQSKIG SLDNITHVPG GGNKKIETHK LTF
```

B

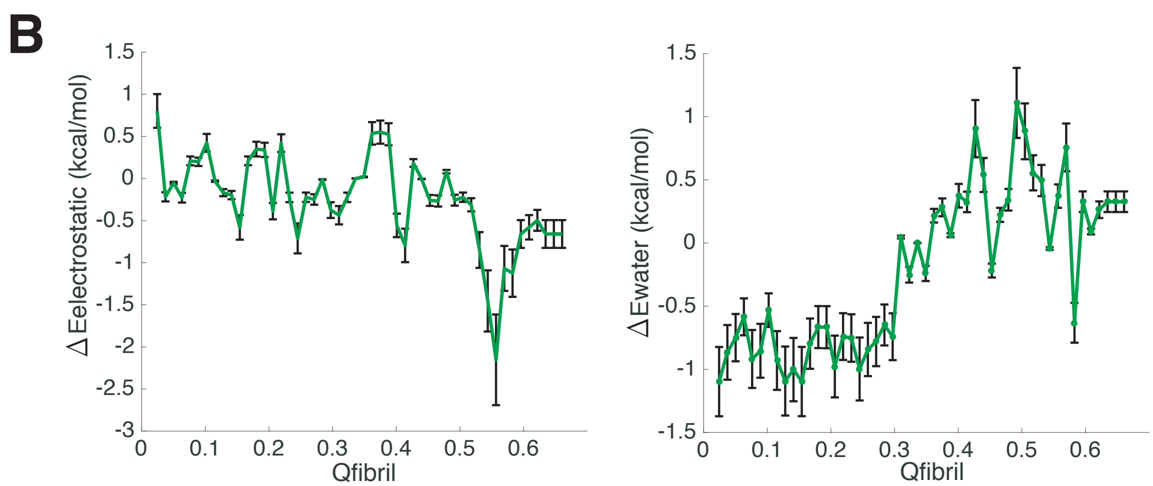

Figure S3: Probing the effects of phosphorylation on the aggregation landscapes of R3/4. (A) Three sites are randomly chosen to be phosphorylated in the thermodynamic perturbation simulations for five repeats. The chosen sites are shown in red. (B) The resulted change in electrostatic energies and water mediated interactions in AWSEM along the reaction coordinate (Qfibril, similarity to the fibrillar form) is shown after five repeated simulations, and the error bar is shown in black. Negative values implicate better favoring the phosphorylated forms.

S2.4 Probing the effects of phosphorylations on the aggregation landscapes of R1/3/4.

A

```
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF

KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF

KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF

KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF

KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
KNVKSKIGST ENLKHQPGGG KVQIVYKPVD LSKVTSKCGS LGNIHHKPGG GQVEVKSEKL DFKDRVQSKI GSLDNITHVP GGGNKKIETH KLTF
```

B

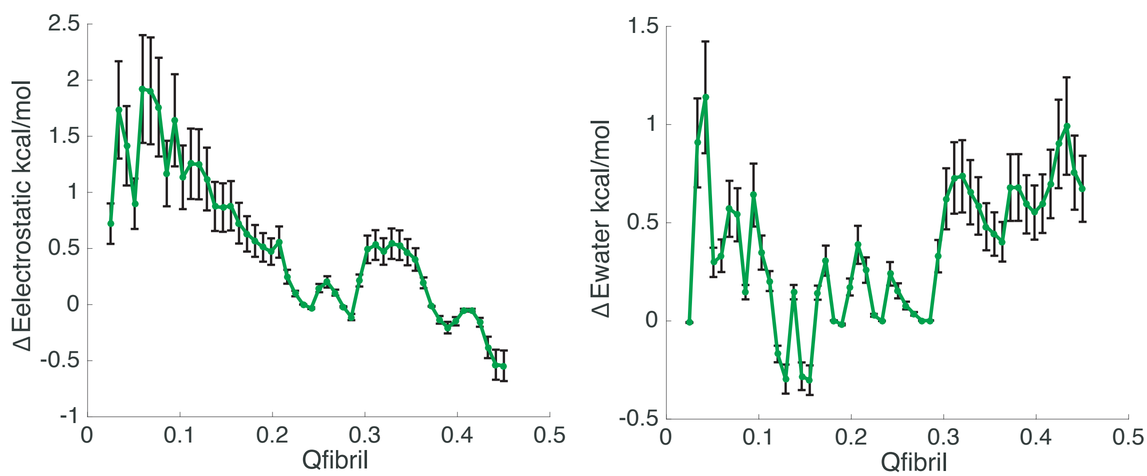

Figure S4: Probing the effects of phosphorylation on the aggregation landscapes of R1/3/4. (A) Three sites are randomly chosen to be phosphorylated in the thermodynamic perturbation simulations for five repeats. The chosen sites are shown in red. (B) The resulted change in electrostatic energies and water mediated interactions in AWSEM along the reaction coordinate (Qfibril, similarity to the fibrillar form) is shown after five repeated simulations, and the error bar is shown in black. Negative values implicate better favoring the phosphorylated forms.

S2.5 Distribution of pairwise distance between C260 in each type of oligomers of R1/3/4.

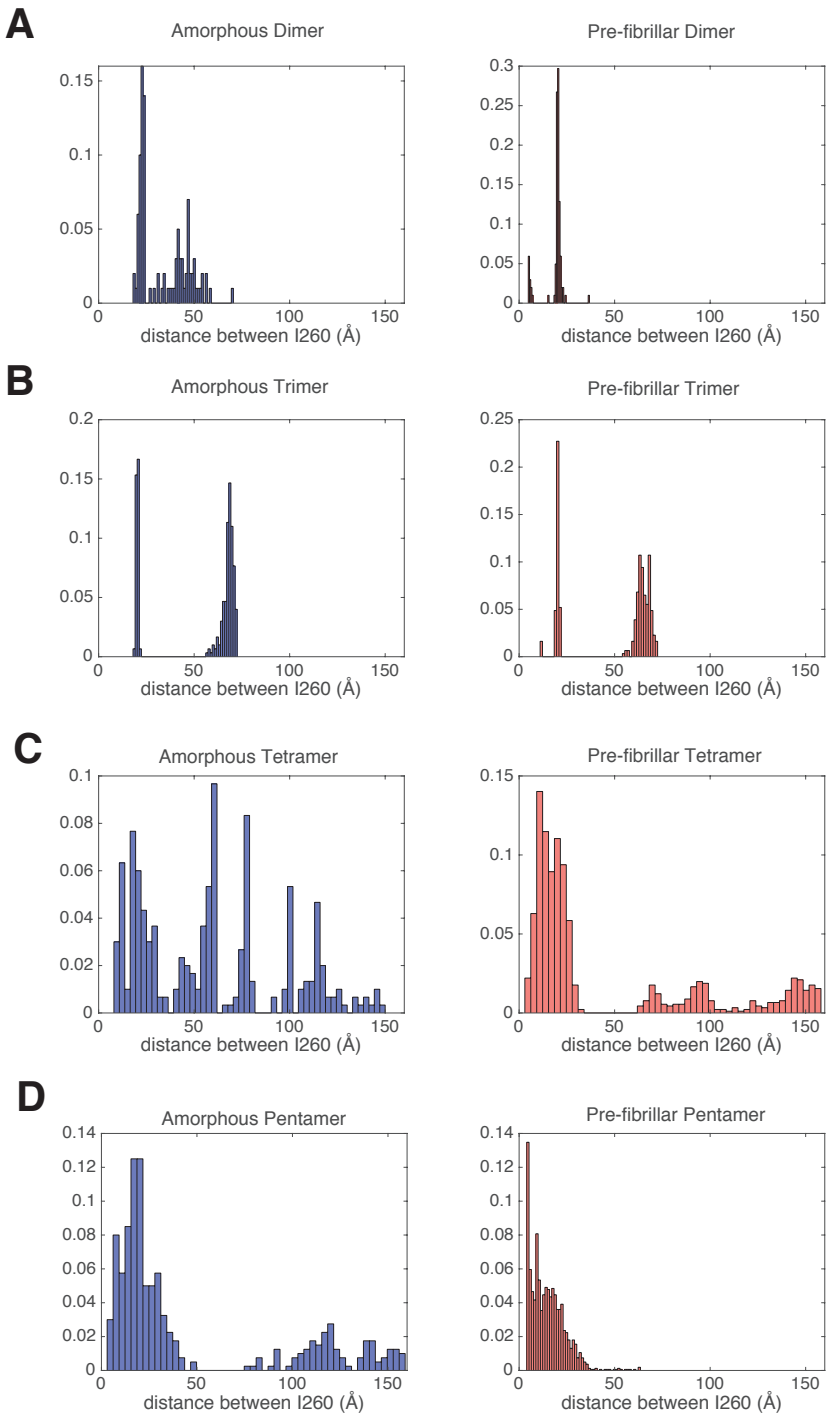

Figure S5: Distribution of pairwise distances between C260 in each monomer chain from oligomeric ensembles of R1/3/4.

S2.6 Structurally-distinct species of oligomers of R3/4 tau and their transitions.

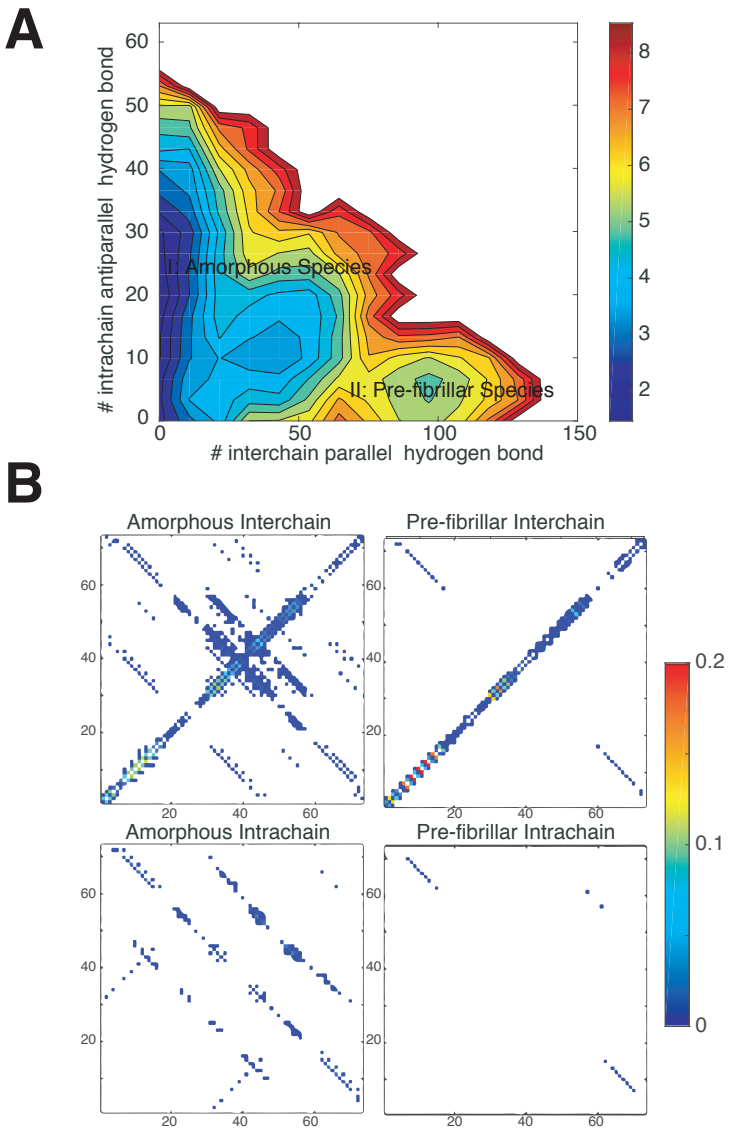

Figure S6: Structurally-distinct species of oligomers of R3/4 tau and their transitions. (A) The grand canonical free energy surface at the concentration of 80 μM at 300 K is plotted using the number of intermolecular parallel hydrogen bonds and the number of intramolecular antiparallel hydrogen bonds as the two dimensions. The two labeled basins represent the amorphous species (labeled class I), while the other is the pre-fibrillar species (labeled class II). (B) Contact probability maps of the amorphous hexamer ensemble (Left) and pre-fibrillar hexamer ensemble (Right). The color indicates the probability that a particular pair of contacts forms in configurations within the structural basin. The top two panels show intramolecular contacts, whereas the bottom two panels show intramolecular contacts. The axes are labeled by residue index. Representative structures for both states can be found in Figure 1.

S2.7 Structurally-distinct species of oligomers of R11/3/4 tau and their transitions.

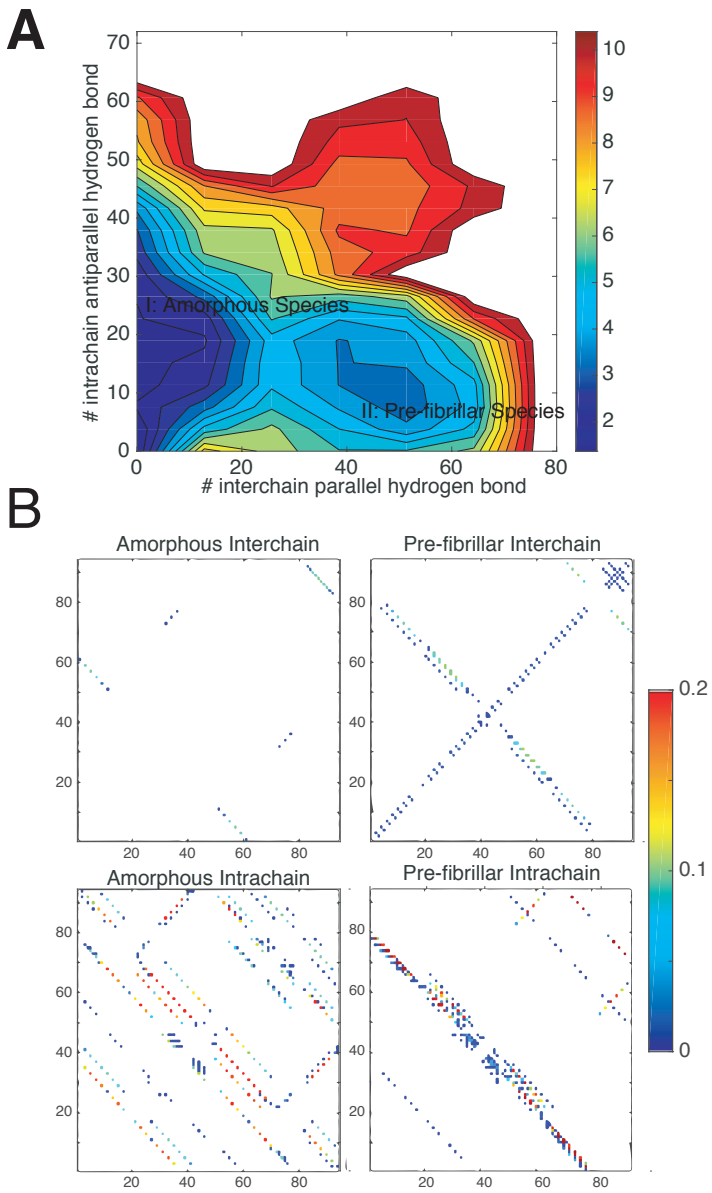

Figure S7: Structurally-distinct species of tau R1/3/4 and their transitions. (A) The grand canonical free energy surface at the concentration of 80 μM at 300 K is plotted using the number of intermolecular parallel hydrogen bonds and the number of intramolecular antiparallel hydrogen bonds as the two dimensions. The two labeled basins represent the amorphous species (labeled class I), while the other is the pre-fibrillar species (labeled class II). (B) Contact probability maps of the amorphous hexamer ensemble (Left) and pre-fibrillar hexamer ensemble (Right). The color indicates the probability that a particular pair of contacts forms in configurations within the structural basin. The top two panels show intramolecular contacts, whereas the bottom two panels show intramolecular contacts. The axes are labeled by residue index. Representative structures for both states can be found in Figure 1.

S2.8 Effects of phosphorylations on the aggregation of R3/4.

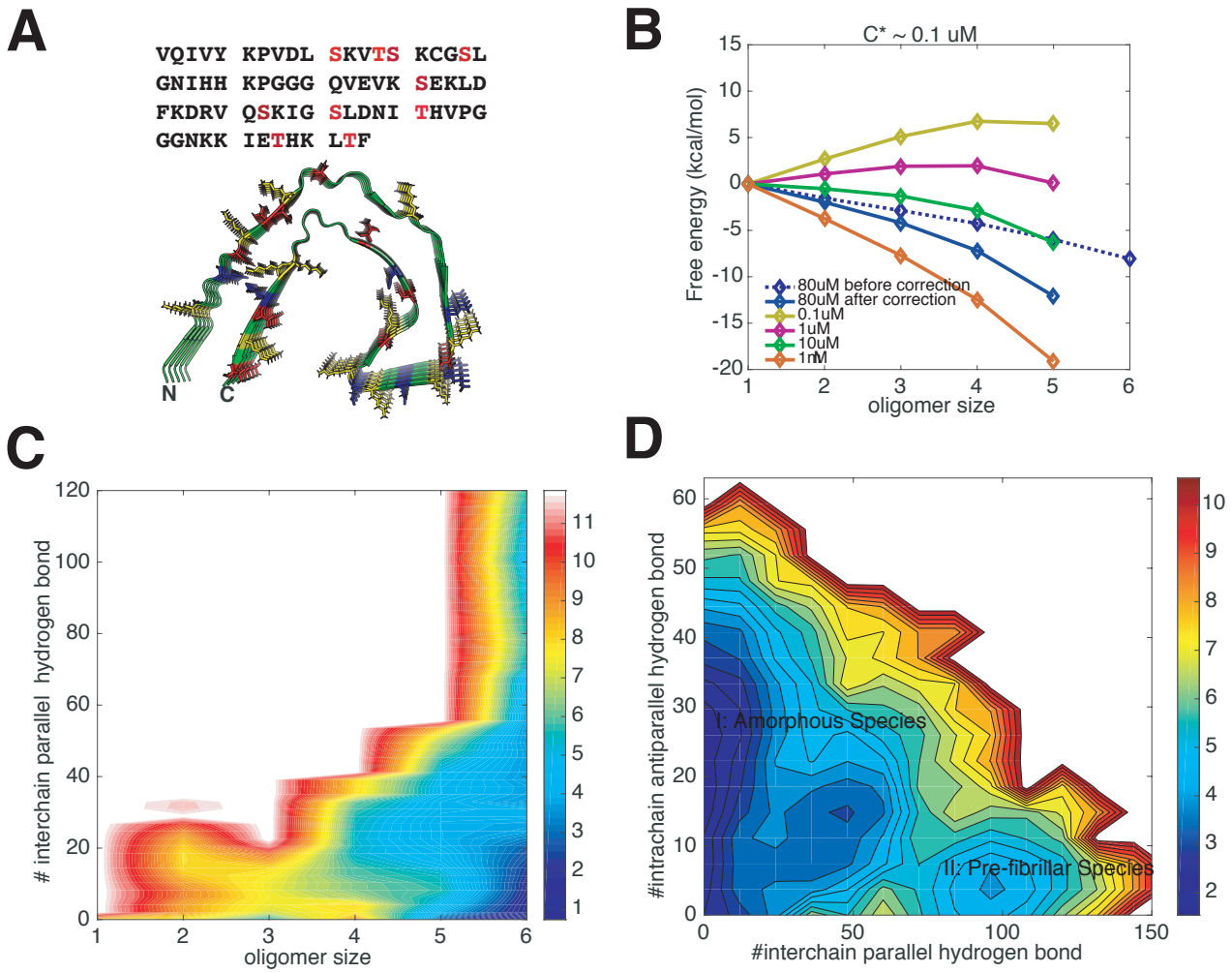

Figure S8: Effects of phosphorylations on the aggregation of R3/4. (A) The amino acid sequence of R3/4 in tau. All of the potential phosphorylated Serines/Threonines are shown in red. The solved fibrillar structure of R3/4 is also illustrated, with positively charged residues colored in yellow, negatively charged residues in blue, and potentially phosphorylated sites in red. Three of these eight residues are randomly phosphorylated before recalculating the free energy surfaces using thermodynamic perturbation theory (see Figure S1). (C) The perturbation-based grand canonical free energy profiles for different oligomer states as corrected for the monomer concentration changes in the fixed number simulation. (B) Perturbation-based grand canonical free energy surface at the concentration of 80 μM at 300 K is plotted using the number of intermolecular parallel hydrogen bonds and the oligomer size as the two dimensions. (D) The recalculated grand canonical free energy surface is also plotted using the number of intermolecular parallel hydrogen bonds and the number of intramolecular antiparallel hydrogen bonds as the two dimensions to be compared with the landscape for the unphosphorylated assembly.

References

[1] Schafer NP, Kim BL, Zheng W, Wolynes PG (2014) Learning to fold proteins using energy landscape theory. *Israel journal of chemistry* 54(8-9):1311–1337.

[2] Davtyan A, et al. (2012) Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The Journal of Physical Chemistry B* 116(29):8494–8503.

[3] Chen M, Wolynes PG (2017) Aggregation landscapes of huntingtin exon 1 protein fragments and the critical repeat length for the onset of huntington's disease. *Proceedings of the National Academy of Sciences* 114(17):4406–4411.

[4] Zheng W, Tsai MY, Wolynes PG (2017) Comparing the aggregation free energy landscapes of amyloid beta (1–42) and amyloid beta (1–40). *Journal of the American Chemical Society* 139(46):16666–16676.

[5] Shen T, Zong C, Hamelberg D, McCammon JA, Wolynes PG (2005) The folding energy landscape and phosphorylation: modeling the conformational switch of the nfat regulatory domain. *The FASEB journal* 19(11):1389–1395.

[6] Lätzer J, Shen T, Wolynes PG (2008) Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles. *Biochemistry* 47(7):2110–2122.