Supporting material

The distinct structural preferences of Tau protein repeat domains

Xuhua Li[a], Xuewei Dong[a] Guanghong Wei[a*], Martin Margittai[b], Ruth Nussinov[c,d*] and Buyong Ma[d]

[a] Department of Physics, State Key Laboratory of Surface physics, Key Laboratory for Computational Physical Science (Ministry of Education), Collaborative Innovation Center of Advanced Microstructures (Nanjing), Fudan University, Shanghai 200433, People's Republic of China.
[b] Department of Chemistry and Biochemistry, University of Denver, Denver, CO 80208.
[c] Sackler Inst. of Molecular Medicine Department of Human Genetics and Molecular Medicine Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.
[d] Basic Science Program, Leidos Biomedical Research, Inc. Cancer and Inflammation Program, National Cancer Institute, Frederick, MD 21702, USA.

[*]E-mail: nussinor@mail.nih.gov; ghwei@fudan.edu.cn

[†] **Electronic Supplementary Information (ESI) available: 3 Tables and 8 figures.**

Materials and Methods

Construction of simulation systems

Based on the cryo-electron microscopy derived (cryoEM-derived) C-shaped R3-R4 structure, we constructed all possible bi-repeat C-shaped motifs including R1-R2, R1-R3 and R1-R3. This C-shaped R3-R4 motif consist of 73 amino acids (residues 306~378). R1-R2 motif comprises repeat R1 and R2 (residues 244~305) and extra ten amino acids following S_{305}. R1-R3 comprises repeat R1 and R3 and extra ten amino acids following Q_{336}. R2-R3 comprises repeat R2 and R3 and extra ten residues also following S_{305}. R1-R2, R1-R3 and R2-R3 are constructed using the cryoEM-derived C-shaped R3-R4 structure as a template (PDB ID entry 503L for PHF and 5O3T for SF).[1, 2] In other words, we construct R1-R2 by replacing every residue of R3-R4 using residues of R1-R2 in the corresponding position. The same process is done for R1-R3 and R2-R3. There are 31, 31, 31 and 32 residues in R1, R2, R3 and R4 repeat, respectively. One more residue in R4 is D_{348}. After sequence alignment, in the corresponding position of D_{348} in R4, there is a gap in R1, R2 and R3. In the initial structure of each bi-repeat motif, there are eight β-sheet regions

termed β1~β8. Using cryoEM-derived R3-R4 as an example, the residues of the β1-8 regions are V_{306}~K_{311}, V_{313}~C_{322}, N_{337}~K_{341}, Q_{336}~S_{341}, K_{343}~K_{347}, R_{349}~I_{354}, S_{356}~V_{363} and N_{368}~F_{378}, respectively. The residues of β-sheet regions of other three bi-repeat motifs are in the same position as those of R3-R4. Each bi-repeat structure includes three type of structures: octameric protofilaments of PHF termed PHF_{34}, PHF_{12}, PHF_{13} and PHF_{23}; filaments made of two tetrameric protofilaments of PHF termed $2PHF_{34}$, $2PHF_{12}$, $2PHF_{13}$ and $2PHF_{23}$, and that of SF termed $2SF_{34}$, $2SF_{12}$, $2SF_{13}$ and $2SF_{23}$. As done in our previous work, [2] we choose octamers to model protofilaments. In order to explore the effect of heparin on structural stability of tau filament, we manually docked heparin hexamers in the grooves of PHF_{34}, PHF_{12}, PHF_{13} and PHF_{23}. The size of groove of these protofilaments matches well with the length and diameter of a heparin. We name the tau-heparin complex as PHF_{34}-HP, PHF_{12}-HP, PHF_{13}-HP and PHF_{23}-HP. We construct four types of K18 in terms of the different shape of R1-R2 motifs. There are seven different K18 (residues 243~372) octameric protofilaments with a C-shaped, linear shape, compact U-shaped or extended U-shaped R1-R2, and a C-shaped R3-R4. The linear shape, compact U-shaped and extended U-shaped R1-R2 of these K18 octamer protofilaments in our work are corresponding to K18-2, K18-8, and K18-9 in our previous works.[2] The structure of R3-R4 within these K18 conformers is from the cryoEM-derived PHF_{34}. For simplicity, we name them K18-1, -2, -3, -4, -5, -6, -7. We totally construct twenty-four systems including four combinations of PHF, 2PHF, 2SF, PHF-HP, seven types of K18 and SF_{34} (Table S1).

All-atom explicit solvent MD simulations protocols

All-atom explicit-solvent MD simulations were performed using the NAMD[3] program with the CHARMM force field[4, 5] (CHARMM27 for PHF, 2PHF, 2SF, and K18; CHARMM 36 for PHF-HP) and three points (TIP3P) water models. All systems were solvated in the rectangle box with a minimum distance 15 Å from any boundary of the water box to any tau protein atom. Na+ and Cl- ions were added to neutralize the systems and mimic ~150 mM ionic strength corresponding to physiological ion concentration. The total number of atoms in all systems varied from 61220 to 130447 (detailed information shown in Sup-Table 1.). The non-protein groups were energy minimized by 5000 steps of the steepest decent minimization. The systems were then gradually heated from 0 to 300 K and equilibrated for 4 ns. In order to compare with previous work[2], K18 systems were gradually heated up to 310 K and equilibrated for 4 ns. In the energy minimization stage of K18, the positions of inter-chain hydrogen bonds of repeats R1-R2 are restrained. MD simulations of bi-repeat (PHF, SF, 2PHF, 2SF, and PHF-HP) systems were conducted for 100 ns in NPT ensemble at a pressure of 1 atm (1 atm =101.3kPa), and a temperature of 300 K. The MD production runs of K18 systems were conducted for 40 ns in the NPT ensemble with the same conditions as bi-repeat systems. All simulations were performed using periodic boundary conditions. Long range electrostatic interactions were calculated by the particle mesh Ewald method. The van der Waals interactions were treated by the smoothly truncated method via a potential shift at 14 Å. The equations of motion were

integrated using the Verlet integrator with a time step of 2 fs, and the trajectory was save every 20 ps.

Analysis methods

We analyze the trajectories using several parameters, including root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), hydrogen bond (H-bond) number, secondary structure content, contact probability map, salt-bridge, angle between β2-β3 and the angle between β6-β7. The generalized Born method with molecular volume (GBMV)[6] implemented in CHARMM[10] was used to calculate relative free energy of protofilaments. All analyses were performed using the tools within GROMACS-4.5.3[7, 8] and in-house codes. The Cα-RMSDs of protofilaments/filaments were calculated with respect to the energy minimized initial structure. The Cα-RMSF was calculated. A H-bond is considered to be formed if the distance between N and O atoms is less than 3.5 Å and the angle of N-H···O is greater 150°. The secondary structures of protofilaments/filaments were identified using the dictionary secondary structure of protein (DSSP) program[11]. As the C-terminal and N-terminal residues are always in random coil conformation, the average percentage of each secondary structure was calculated by only ignoring the two terminal residues of each chain. The inter-chain and intra-chain residue pairwise contact probabilities were analyzed. Here, a contact is defined when the carbon atoms of two non-sequential residues come within 0.54 nm or any other heavy atoms (ignore hydrogen atoms) of two non-sequential residues lie within 0.46 nm. A salt bridge is formed if the minimum distance between the N atom of the side chain NH_3^+ of K_{353}/K_{290} and the side chain COO^- group of D_{358}/D_{295} (i.e. the K_{353}-D_{358} and K_{290}-D_{295} distance) is less than 0.4 nm. In order to characterize C-shaped structure, two angles and one distance were defined. The two angles consist of the angle between β2-β3 and the angle between β6-β7 of tau protofilaments. They are both defined by two vectors involving four residues. Taking PHF_{34} as an example, β2-β3 angle is the angle between the vector from V_{318} to C_{322} and the vector from H_{332} to I_{328}. β6 and β7 is the angle between the vector from Q_{351} to K_{353} and the vector from H_{362} to D_{355}. The residues that form the two angles in other bi-repeat systems are in the same positions as that in PHF_{34}. The distance is between Q_{351} and I_{371} of octameric PHF_{34}, Q_{288} and I_{308} of PHF_{12}, and T_{319} and V_{339} of PHF_{12} and PHF_{23}. The values of angles and distance are averaged over all eight chains of protofilaments. For all bi-repeat systems, the last 50 ns trajectory was used to calculate the average RMSF, secondary structure content, β-sheet probability of each residue, salt-bridge. The relative free energy of K18 was calculated by GBMV algorithm implemented in the CHARMM program with standard parameters. Single-point energies were calculated after 500 steps of energy minimization to relax the local geometries caused by the thermal fluctuations which occurred in the MD simulations. For each K18 system, a total of 250 conformations extracted from the last 5 ns were used to obtain the average energy. All representations of the studied systems are drawn using Visual Molecular Dynamics (VMD) software.[9]

Discussions

The contribution of β8 for the shape formation

We estimated the contributions of β8 to shape formation of the four bi-repeat combinations by calculating the pairwise residue contact probabilities between β8 and β1~β2 for each PHF system and calculate the free energy contribution of β8 using $-RT\ln P$ (P is the total native contact fraction of β8 and β1~β2) (see Fig S10 below). We assume that each residue contact contributes an unitless energy ε, and that the maximum native contact in the four systems are almost constant since they have the same number residues and share high sequence similarity. Thus P = Sum(ε*$K_{contact}$)/(N_{frame}*Mc*ε), where $K_{contact}$ is the number of contact in a conformer, N_{frame} is the total conformers generated, and Mc is the maximum contact between between β8 and β1~β2.

Distinct interaction patterns between β8 and β1~β2 are observed in the contact probability maps of the four systems (Fig S10), indicating different interactions between β8 and β1~β2 in stabilizing the structural preferences of the four systems. In PHF_{34} system, residue pairs with high contact probability are F_{378}/I_{308}, L_{376}/Y_{310}, H_{374}/Y_{310} and K_{370}/D_{314}, indicating that hydrophobic interactions make dominant contributions to the C-shaped conformation. In PHF_{12} system, relatively strong interactions are seen for K_{311}/D_{252}, Q_{307}/K_{254} and Q_{307}/V_{256} residue pairs, including salt-bridge interactions between K_{311} and D_{252}, hydrophobic interactions between the nonpolar aliphatic groups of polar residues Q_{307} and K_{254}, and those between the nonpolar aliphatic groups of Q_{307} and V_{256}. In PHF_{13} system, salt-bridge interactions of residue pairs K_{340}/D_{252} and E_{338}/K_{254} and hydrophobic interactions between L_{344} and M_{250} have a strong contribution to the U-shaped structure. In PHF_{23} system, salt-bridge interactions of residue pairs E_{342}/K_{281} and K_{340}/D_{283} contribute most to the V-shaped structure, although hydrophobic interactions between F_{346} and I_{278} also make non-ignorable contributions. These different interaction patterns can be clearly seen from the snapshots in Fig S10 (B). The values of contact energy reflect the contribution of β8 to the shape formation of PHF_{34} and PHF_{12} systems is larger than that of PHF_{13} and PHF_{23} systems. Although the β8 has a great contribution to shape formation of PHF_{12}, its structure is still not stabilized and tends to form a linear structure.

The contribution of dimerization for the shape formation

In order to examine the contributions of tau dimerization to the C-shape formation of R3-R4, we performed two additional MD simulations on an isolated tetramer protofilament and compared the results with those of filaments consisting of two tetramer protofilaments. Figure S6 (A) shows the time evolution of the average Ca-RMSD of isolated tetramers in the two MD runs and the average Ca-RMSD of two tetramers in the 2PHF34 and 2SF34 filaments. It can be seen from Fig S6 that the Ca-RMSD of the isolated tetramer increases with simulation time and reaches 0.8 nm, whereas the Ca-RMSD of the tetramer in the filaments becomes much smaller (around 0.5 nm). The reduced RMSD of the tetramer in the filaments indicates the important

role of dimerization on the C-shape formation of R3-R4. We also calculated the contact energy between the two tetramers in 2PHF filaments and 2SF filaments. The data show that the dimerization of both $2PHF_{34}$ and $2SF_{34}$ is energetically favorable. For other systems, although dimerization of some systems (i.e. PHF_{23}, SF_{13} and SF_{23}) displays energetically favorable, their structure still cannot maintain the initial structures and exhibit different structural preferences.

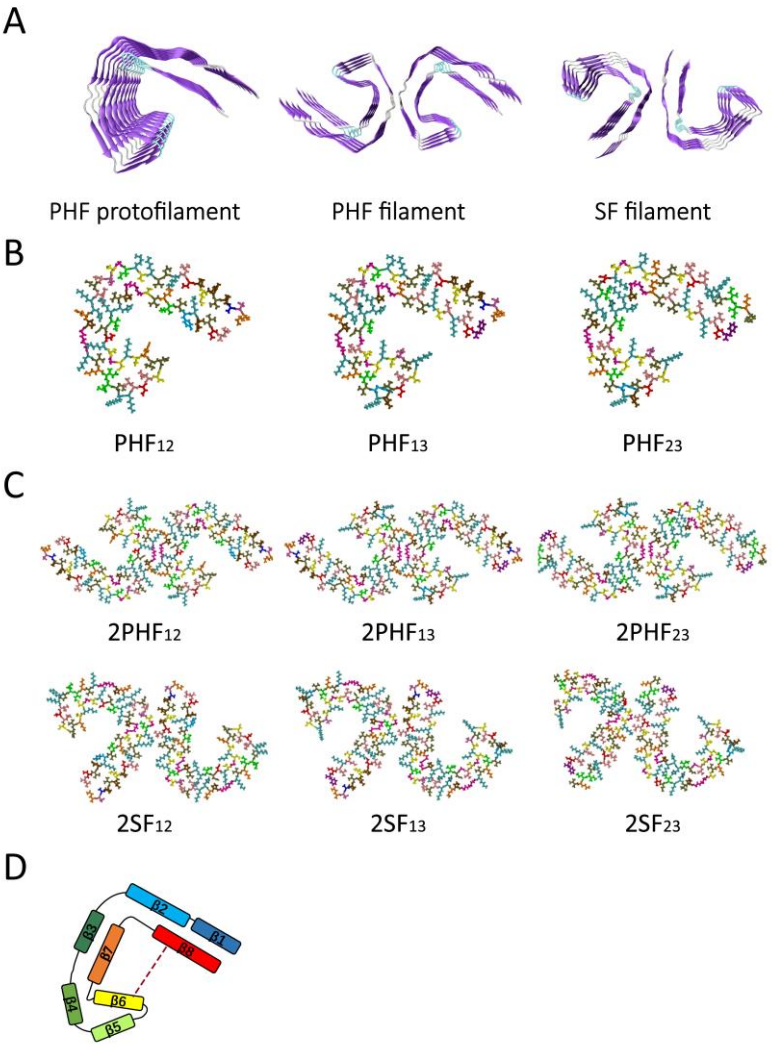

Figure S1. Initial structures of MD simulation systems. (A) Initial structure of the octameric protofilaments and filaments. Initial structures of R1-R2, R1-R3 and R2-R3 PHF protofilaments (B), PHF filaments and SF filaments (C). For clarity, only one chain in each protofilament is shown. (D) A cartoon representation of a single chain in protofilaments with eight β-sheet regions labelled and a representation of the distance mentioned in the paper with red line.

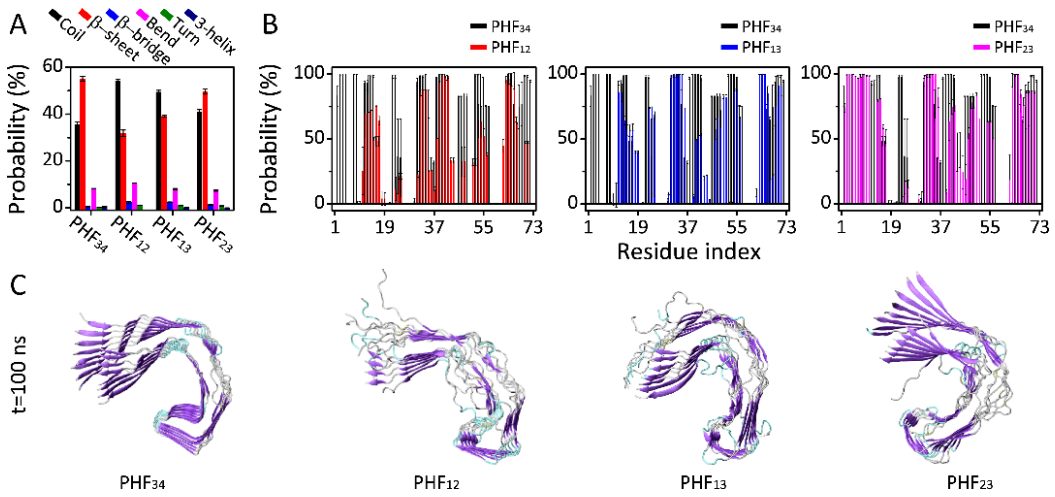

Figure S2. Secondary structure analyses of each PHF protofilament. (A) The probability of each type of secondary structure. (B) The comparison of residue-based β-sheet probability between PHF_{34} and other three PHF protofilaments (PHF12, PHF13 and PHF23). (C) The snapshots of four PHF protofilaments at t=100 ns.

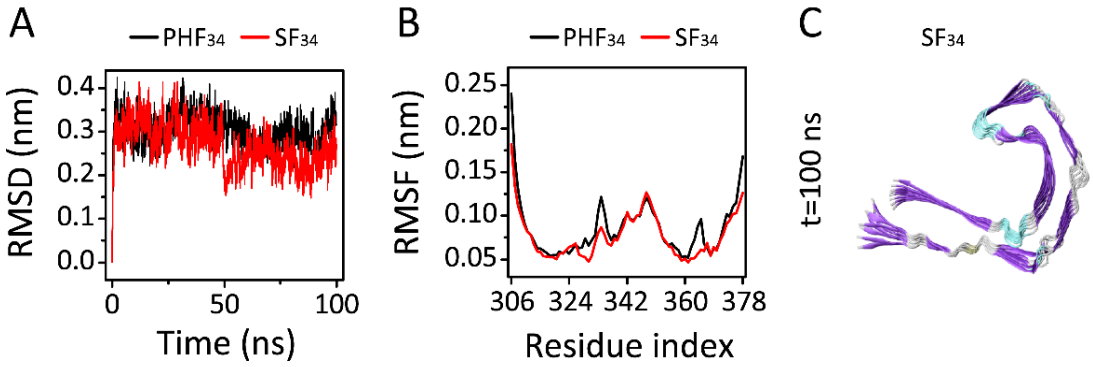

Figure S3. Structural stability of the SF_{34} protofilament. (A) Time evolution of the RMSD for SF_{34}. (B) The RMSF of SF_{34}. (C) The snapshots of SF_{34} protofilaments at t=100 ns. For comparison, the RMSD and RMSF data for PHF_{34} are also given.

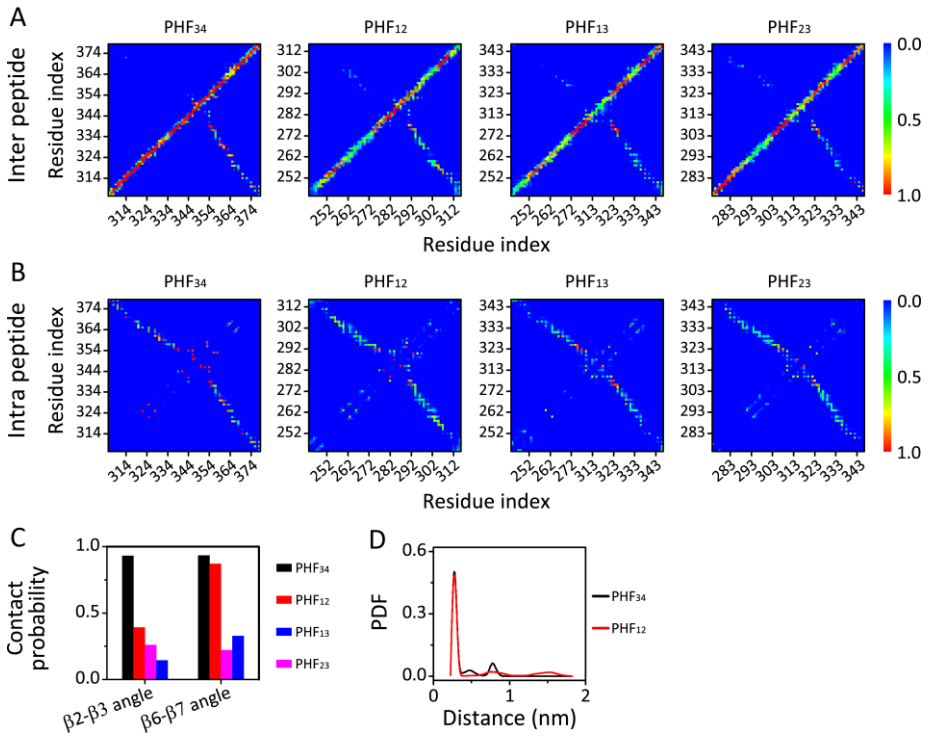

Figure S4. Analyses of inter-chain and intra-chain interactions for four PHF protofilaments. (A) Inter-chain residue pairwise contact probability maps. (B) Intra-chain residue pairwise contact probability maps. (C) The contact probability between identical residues in each two neighboring chains. The average value of intra-chain contact probabilities among C_{322}, L_{325}, I_{328}, P_{364} and G_{367}in β2-β3 angle region and K_{353}, S_{356} and D_{358} in β6-β7 angle region of $PHF_{34}I_{260}$, T_{263},L_{266}, $P_{301}$and G_{304} in β2-β3 angle and K_{290}, S_{293} and D_{295} in β6-β7 angle of PHF_{12}, and C_{291}, K_{294}, I_{297}, P_{332}， and G_{335} in β2-β3 angle and K_{341}, S_{344} andG_{346}in β6-β7 angle of PHF_{13} and PHF_{23}.(E) The analyses of K_{353}-D_{358} salt-bridge in PHF_{34} and K_{290}-D_{295} salt-bridges in PHF_{12}.

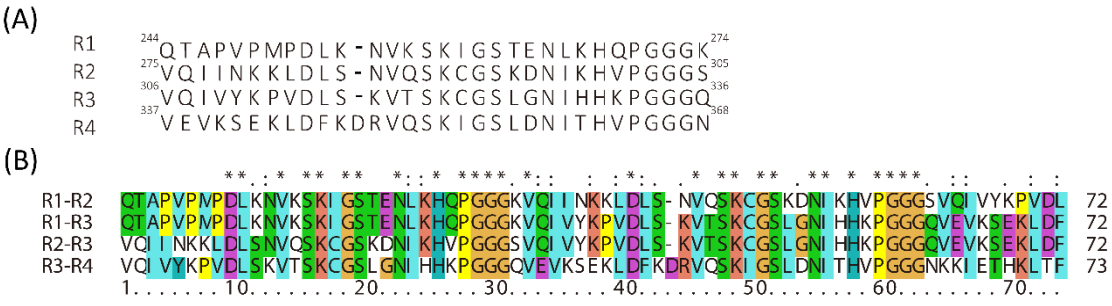

Figure S5. Sequence alignment of (A) four repeats (R1, R2, R3 and R4) and (B) four bi-repeat combinations (R1-R2, R1-R3, R2-R3 and R3-R4). Identical residues are marked by star symbol (*). The high similar residues are marked by two dots symbol (:) and the similar residues are marked by a dot symbol (.).

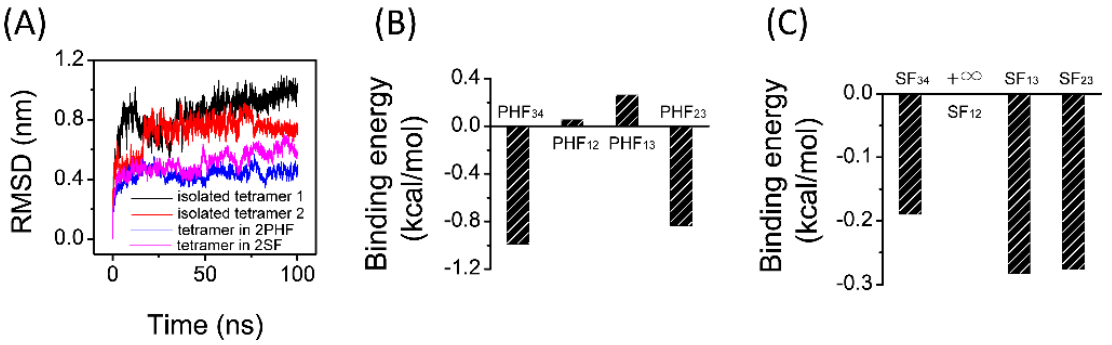

Figure S6. Analyses of the contributions of tau dimerization to the shape formation of the segments. (A) Time evolution of the RMSD for tetramer protofilaments of R3-R4. Binding energies between the tetramer protofilament of each 2PHF filaments (B) and 2SF filaments (C).

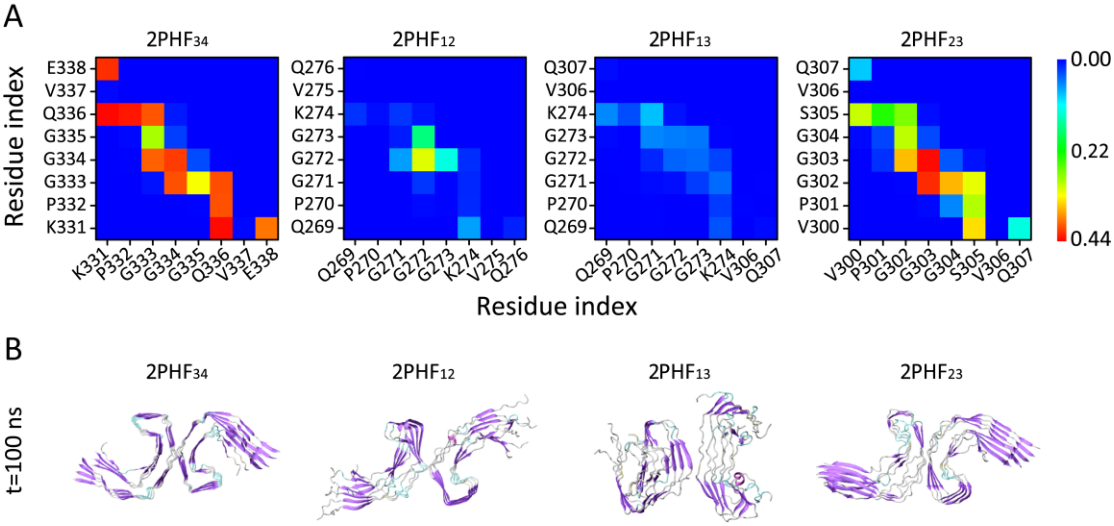

Figure S7. Analyses of interface interactions between the two protofilaments in each PHF filament. (A) The interface contact probability maps. (B) The snapshots of 2PHF filaments at t=100 ns. To make it clear, we extend the interaction regions involving eight residues.

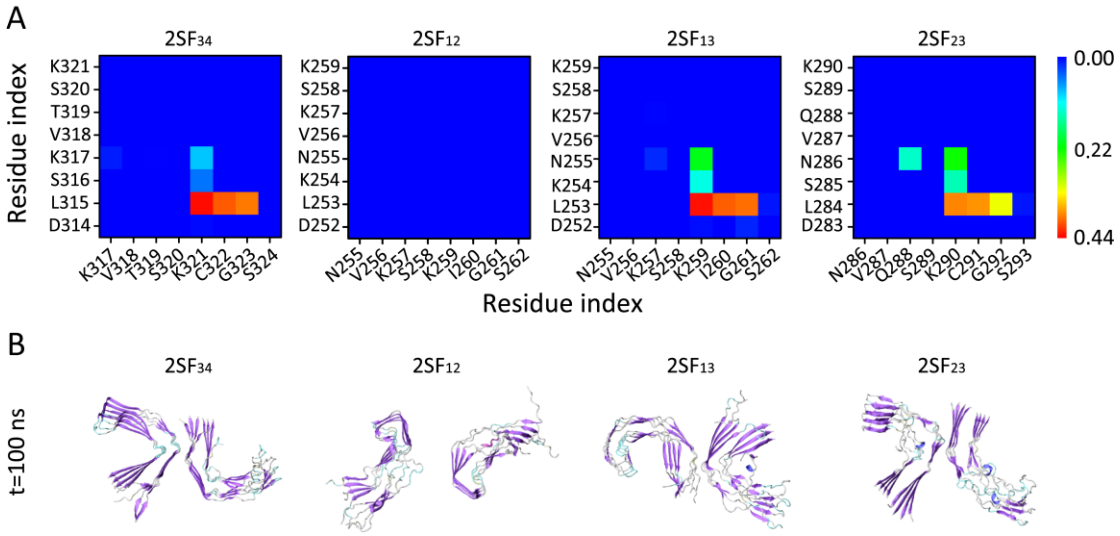

Figure S8. Analyses of interface interactions between the two protofilaments in each SF filament. (A) The interface contact probability maps. (B) The snapshots of 2SF filaments at t=100 ns.

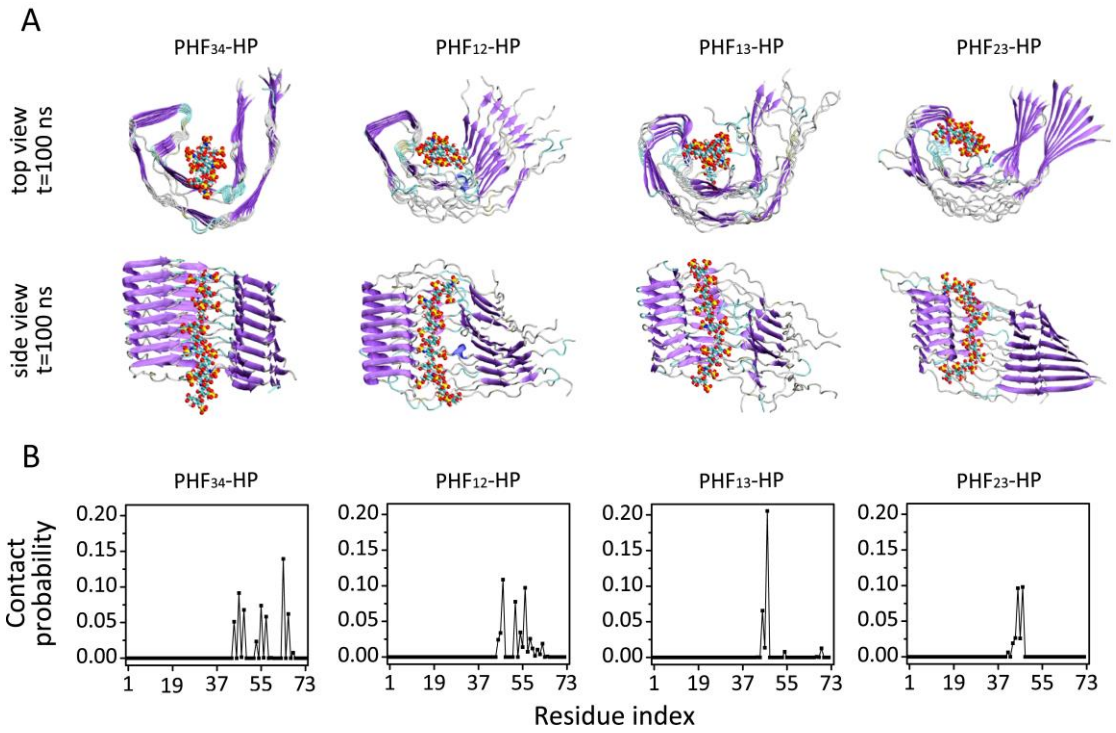

Figure S9. The effects of heparin on PHF protofilaments. (A) The snapshots PHF-HP structures with two different views (top view and side view). (B) The contact probability between heparin and each residue of tau filament.

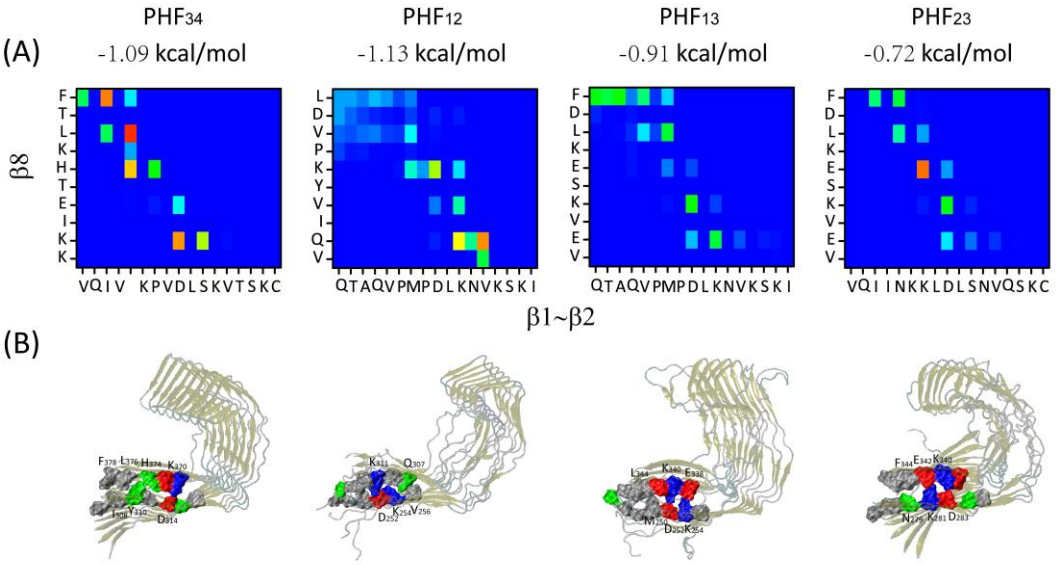

Figure S10. Analyses of interaction patterns between residues in β1~β2 and β8 regions of the four protofilaments. (A) The contact probability maps. (B) The snapshots of PHF protofilaments at t=100 ns. The residues in β1~β2 and β8 regions are shown in surface representation. Basic, acidic, polar and hydrophobic residues are colored in red, blue, green and silver, respectively.

Table S1. Simulation details for each system.

Systems	Description	Number of Atoms	Force fields	Simulation time
SF_{34}	R3-R4 protofilament	64347	CHARMM27	100 ns
PHF_{34}	R3-R4 protofilament	66115	CHARMM27	100 ns
PHF_{12}	R1-R2 protofilament	63401	CHARMM27	100 ns
PHF_{13}	R1-R3 protofilament	61682	CHARMM27	100 ns
PHF_{23}	R2-R3 protofilament	61220	CHARMM27	100 ns
$2PHF_{34}$	R3-R4 filament	90195	CHARMM27	100 ns
$2PHF_{12}$	R1-R2 filament	88958	CHARMM27	100 ns
$2PHF_{13}$	R1-R3 filament	86268	CHARMM27	100 ns
$2PHF_{23}$	R2-R3 filament	85551	CHARMM27	100 ns
$2SF_{34}$	R3-R4 filament	100613	CHARMM27	100 ns
$2SF_{12}$	R1-R2 filament	88958	CHARMM27	100 ns
$2SF_{13}$	R1-R3 filament	94109	CHARMM27	100 ns
$2SF_{23}$	R2-R3 filament	89306	CHARMM27	100 ns
PHF_{34}-HP	R3-R4 protofilament with heparin	68463	CHARMM36	100 ns
PHF_{12}-HP	R1-R2 protofilament with heparin	67606	CHARMM36	100 ns
PHF_{13}-HP	R1-R3 protofilament with heparin	65975	CHARMM36	100 ns

PHF_{23}-HP	R2-R3 protofilament with heparin	65872	CHARMM36	100 ns
K18-1	C-shaped R1-R2	93855	CHARMM27	40 ns
K18-2	Linear shape R1-R2	115827	CHARMM27	40 ns
K18-3	Compact U-shaped R1-R2	104839	CHARMM27	40 ns
K18-4	Extended U-shaped R1-R2	131697	CHARMM27	40 ns
K18-5	Linear -shape R1-R2	113789	CHARMM27	40 ns
K18-6	Compact U-shaped R1-R2	111026	CHARMM27	40 ns
K18-7	Extended U-shaped R1-R2	130447	CHARMM27	40 ns

Table S2. Average RMSD values in four combination PHF systems of the last 50 ns.

System	RMSD (nm)		
	Without heparin	With heparin	Reduction Rate
PHF34	0.29	0.23	20.7%
PHF12	0.78	0.76	2.6%
PHF13	0.90	0.71	21.1%
PHF23	0.83	0.80	3.6%

Table S3. The Cα-RMSF difference between PHF protofilaments with and without heparin in β1~β8 regions.

System	RMSF (nm)							
	β1	β2	β3	β4	β5	β6	β7	β8
PHF34	0.003	0.005	-0.002	0.002	0.013	0.012	0.005	0.017
PHF12	0.090	-0.031	-0.015	0.048	0.076	0.048	-0.008	0.128
PHF13	-0.044	-0.003	-0.024	0.004	0.086	0.094	0.022	0.204
PHF23	-0.013	-0.015	-0.023	-0.012	0.003	0.083	-0.035	0.120

A positive number means a rigidity increase and a negative number indicates a reduction in rigidity when heparin exists with protofilaments.

1. A. W. P. Fitzpatrick, B. Falcon, S. He, A. G. Murzin, G. Murshudov, H. J. Garringer, R. A. Crowther, B. Ghetti, M. Goedert and S. H. W. Scheres, *Nature*, 2017, **547**, 185-190.
2. X. Yu, Y. Luo, P. Dinkel, J. Zheng, G. Wei, M. Margittai, R. Nussinov and B. Ma, *J Biol Chem*, 2012, **287**, 14950-14959.
3. L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan and K. Schulten, *Journal of Computational Physics*, 1999, **151**, 283-312.
4. M. Buck, S. Bouguet-Bonnet, R. W. Pastor and A. D. MacKerell Jr, *Biophysical journal*, 2006, **90**, L36-L38.

5. J. Huang and A. D. MacKerell, *Journal of computational chemistry*, 2013, **34**, 2135-2145.
6. M. S. Lee, F. R. Salsbury Jr and C. L. Brooks III, *The Journal of chemical physics*, 2002, **116**, 10606-10614.
7. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *Journal of computational chemistry*, 2005, **26**, 1701-1718.
8. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson and D. Van Der Spoel, *Bioinformatics*, 2013, **29**, 845-854.
9. W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, **14**, 33-38.
10. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of computational chemistry*, 1983, **4**, 187-217.
11. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.