Mapping the configurational landscape and aggregation phase behavior of the tau protein fragment PHF6

Evan Pretti[a] **and M. Scott Shell**[a,1]

This manuscript was compiled on October 5, 2023

The PHF6 (Val-Gln-Ile-Val-Tyr-Lys) motif, found in all isoforms of the microtubule-associated protein tau, forms an integral part of ordered cores of amyloid fibrils formed in tauopathies and is thought to play a fundamental role in tau aggregation. Because PHF6 as an isolated hexapeptide assembles into ordered fibrils on its own, it is investigated as a minimal model for insight into the initial stages of aggregation of larger tau fragments. Even for this small peptide, however, the large length and time scales associated with fibrillization pose challenges for simulation studies of its dynamic assembly, equilibrium configurational landscape, and phase behavior. Here, we develop an accurate, bottom-up coarse-grained model of PHF6 for large-scale simulations of its aggregation, which we use to uncover molecular interactions and thermodynamic driving forces governing its assembly. The model, not trained on any explicit information about fibrillar structure, predicts coexistence of formed fibrils with monomers in solution, and we calculate a putative equilibrium phase diagram in concentration-temperature space. We also characterize the configurational and free energetic landscape of PHF6 oligomers. Importantly, we demonstrate with a model of heparin that this widely-studied cofactor enhances the aggregation propensity of PHF6 by ordering monomers during nucleation, and remaining associated with growing fibrils, consistent with experimentally characterized heparin-tau interactions. Overall, this effort provides detailed molecular insight into PHF6 aggregation thermodynamics and pathways, and furthermore, demonstrates the potential of modern multiscale modeling techniques to produce predictive models of amyloidogenic peptides simultaneously capturing sequence-specific effects and emergent aggregate structures.

tau protein | amyloid aggregation | multiscale modeling

Significance Statement

Amyloid aggregation of the microtubule-associated protein tau is central to the pathology of Alzheimer's disease and other tauopathies, but the molecular details underlying tau fibrillization are poorly understood. The PHF6 hexapeptide forms a crucial part of the cross-β spines found in all pathological tau structures. Here, we show that an entirely predictive, bottom-up coarse-grained model of PHF6 illuminates the hierarchy of structures and driving forces underlying its aggregation. Large-scale simulations quantify the phase behavior, fibrillization thermodynamics, and oligomer conformational landscape of PHF6, essential factors underlying fibril nucleation and growth, and further offer mechanisms of cofactor-induced aggregation. These insights suggest the potential for modern multiscale methods to predictively inform experimental efforts to study tau fragments with detailed molecular pictures of aggregation.

Insight into amyloid aggregation of the microtubule-associated protein tau is critical for establishing a molecular understanding of the progression of Alzheimer's disease and other tauopathies (1, 2), and more practically, for the development of faithful benchtop aggregation models for therapeutic screening. Although neurodegenerative diseases in this class all share the common feature of tau fibril formation, remarkable advances in the field suggest that each tauopathy is distinguished by a unique fold or set of folds of tau within its fibrils' cores (3, 4). Landmark cryo-electron microscopy (cryo-EM) studies (5–9) have revealed that, despite a great diversity in fibrillar folds across diseases, all presently known pathological tau fibril structures incorporate the ^{306}VQIVYK^{311} (PHF6, Val-Gln-Ile-Val-Tyr-Lys) motif into their cores. In these structures, PHF6, found in the R3 C-terminal repeat region present in all tau isoforms (10), participates in cross-β hydrogen bonding, characteristic of amyloid aggregation, that structures successive layers of fibrils (11–13). Interactions of the hydrophobic sidechains of PHF6 between nearby monomers within fibrils contribute a significant stabilizing effect (14, 15).

PHF6, and the related ^{275}VQIINK^{280} (PHF6*, Val-Gln-Ile-Ile-Asn-Lys) motif in the R2 C-terminal repeat region found in 4R tau isoforms (16), are known to play key roles in pathological tau aggregation (11, 17). PHF6 in particular, however, is known to be essential as a motif for fibril formation (18), and PHF6 as the isolated VQIVYK hexapeptide self-assembles on its own into ordered β-sheet-containing fibrils (12, 19). Compared to full-length tau (352-441 residues depending on isoform) (16), or known tau fibril cores, typically consisting of $\sim$ 100 residues, the 6-residue PHF6 fragment has a much smaller conformational landscape. Although the processes of formation of larger ordered cores of tau fibrils necessarily entail more complex conformational changes than those involving PHF6 alone, experimental evidence continues to reveal the importance of this fragment in tau aggregation. In particular, a small region of tau including PHF6 and 9 surrounding residues was recently

Author affiliations: [a]Department of Chemical Engineering, Engineering II Building, University of California, Santa Barbara, Santa Barbara, CA 93106-5080, USA

E.P. and M.S.S. designed research; E.P. performed research and analyzed data; and E.P. and M.S.S. wrote the paper.

The authors declare no competing interest.

[1] To whom correspondence may be addressed. Email: shell@ucsb.edu

proposed to form paired cross-β structures in the initial stages of Alzheimer's amyloid aggregation, based on findings from time-resolved cryo-EM (20). Thus PHF6 fibrillization, which serves as a useful model in its own right, is of major interest for understanding the critical early stages of tau aggregation. Yet, despite its small size, illuminating atomistic simulations of PHF6 fibrillization remain intractable, due to the number of chains needed to overcome the aggregation free energy barrier and the long time scales of fibril nucleation and elongation (21, 22). To gain insight into fibril phase behavior, the configurational landscape of oligomers, and the effects that seeding or aggregation cofactors might have on fibril growth, a multiscale modeling approach that addresses the complex balance between local (*e.g.*, secondary structure) and global (collective assembly) interactions is essential.

Here, we present *de novo* large-scale simulations of PHF6 aggregation, using a bottom-up coarse-grained (CG) modeling approach. The model developed here predicts assembly of PHF6 from monomers in solution into paired hydrophobically- and cross-β-stabilized fibrils, even though no information about fibrillar states is included during model training—such that the preference for fibrillization emerges from the effective interactions determined by underlying physical, atomic-scale interactions. We use the model to map out the detailed hierarchical configurational landscape of oligomers, revealing unexpected cylindrin-like structures intermediate in the aggregation pathway distinct from both single and paired β-sheet oligomeric fibril precursors. We then use large-scale simulations to construct the equilibrium phase diagram of coexisting PHF6 fibrils and solution monomers, which we show is consistent with a statistical-mechanical model of fibrillization. We finally show that heparin as a model polyelectrolyte enhances the propensity of PHF6 to fibrillize, playing an active role in the arrangement of monomers during nucleation and remaining associated with the growing fibrils *via* electrostatic interactions.

While CG protein and peptide simulation models have been used to study fibrillization, the present work demonstrates a multiscale approach free of experimental input, training on known fibril structures, and hand-tuning, and thus provides a systematic way to investigate the conformationally dynamic aggregation pathways. Phenomenological top-down CG models (23–26) have revealed fundamental qualitative insights into aggregation thermodynamics and kinetics, but often ignore sequence-dependent behaviors. Alternatively, sequence-specific CG models have studied amyloid aggregation (27–32), and two CG models (33, 34) examined PHF6 aggregation in particular (22, 35). But many existing models are parameterized using conformational and contact information from experimental native structures of globular proteins, or alternately, by *ad hoc* methods to match CG system properties to higher resolution simulations. In contrast, modern systematic bottom-up coarse-graining approaches (36–40), posit "universal" variational principles to optimize CG models. Notably, the relative entropy approach has been used to model polyalanine (41, 42) and to build a Gō-like model able to reproduce folded protein structures given native contact information, with high structural accuracy (43). However, predictive bottom-up models for biologically relevant aggregating systems that take advantage of advances in systematic coarse-graining, avoiding *a priori* input of native contacts, secondary structures, or *ad hoc* parameterization strategies, have yet to be realized. Such models for aggregating peptides have the potential to inform the effects of mutations or post-translational modifications, or the presence of particular cofactors that have not been previously experimentally characterized.

The relative-entropy-based multiscale strategy presented here suggests a predictive route to quantitatively understanding peptide aggregation, using only information from small-scale atomistic (or all-atom, AA) simulations of interacting PHF6 chains, without requiring any experimental structure input. This approach enables large length- and time-scale molecular dynamics (MD) simulations of PHF6 aggregation and fibril coexistence with solution, as well as extensive configurational sampling to explore the conformational landscape of oligomers. These insights create a detailed molecular picture of PHF6 aggregation, including conformational states, mechanisms, and thermodynamic driving forces, and provide fundamental insight into the dynamic pathways of assembly into fibrils, from monomers to oligomers to large-scale aggregates.

Results and Discussion

PHF6 Fibrillization from Solution. We use relative entropy coarse-graining to generate a CG model of PHF6 from implicit-solvent (IS) AA reference systems of 3 interacting capped PHF6 chains. We employ a CG mapping of 4 sites per residue (Fig. 1a), in order to capture secondary and tertiary assembled structures as shown earlier (41, 43). We use the resulting model to simulate solutions of 512 PHF6 at a concentration $\rho = 10$ mM. Fig. 1b–e shows the systems at four temperatures after 10 μs of CG MD; an animation at 350 K is available as Movie S1 of the Supporting Information (SI). Except at 370 K where the solution phase alone persists for the duration of the simulation (Fig. 1e), we observe nucleation and growth of fibrillar aggregates of PHF6, each consisting of two helically intertwined protofilament-like cross-β-sheet structures (Fig. 1f).

The CG model captures the fibrillization process, despite no training on fibril states, and is essential to enabling equilibration of these large-scale aggregation simulations to obtain thermodynamic properties. As measured by autocorrelation times for backbone dihedral transitions (Fig. S1a: see the *SI Appendix*), the CG model shows a 77 ± 13 × reduction in time *vs.* an explicit-solvent (ES) model and a 93 ± 16 × reduction *vs.* an IS model. The chain end-to-end distance, with a longer relaxation timescale, shows that the CG model accelerates by 46.0 ± 1.6 × the ES and 3.59 ± 0.13 × the IS model (Fig. S1b). Beyond these timescale accelerations, the CG model shows additional practical computational speedups, especially for larger systems (Fig. S1c). For 32 PHF6, the performance is 23 ns/day for ES AA, 399 ns/day for IS AA, and 2340 ns/day for CG simulations ($\sim 6\times$ faster than IS and $\sim 100\times$ faster than ES). CG performance is greatly enhanced for larger systems necessary to equilibrate fibrils: IS AA MD of 8 PHF6 gives 885 ns/day while nearly the same simulation speed (893 ns/day) is possible for CG simulations of 1024 PHF6. The combination of dynamic acceleration within, and reduced GPU time to advance by, a given interval of simulated time, points to the great utility of coarse-graining

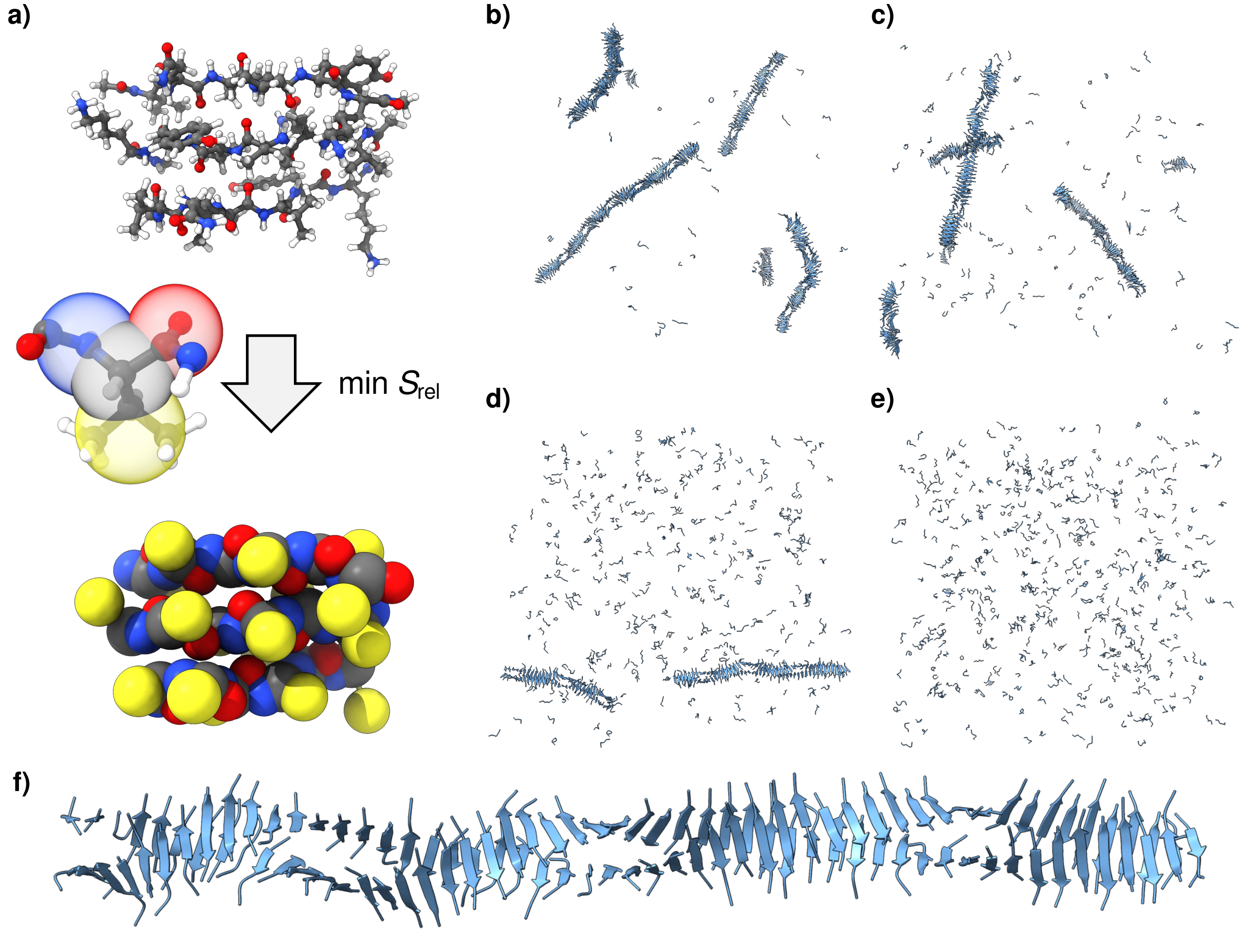

Fig. 1. Illustration of the CG model and representative snapshots of simulated aggregates. (a) Representative configuration of the reference system containing 3 PHF6, with an illustration of the mapping operation to the corresponding CG configuration. (b–e) Systems of 512 PHF6 with $\rho = 10$ mM and $T = 340, 350, 360, 370$ K, respectively, after 10 μs of MD simulation. (f) A representative fibril containing 106 PHF6 from a simulation at 350 K.

in exploring conformational landscapes and carrying out large-scale aggregation simulations.

The structural properties of the simulated aggregates provide a detailed picture of PHF6 fibrils. Fig. S2a calculates the sphericities of the aggregates showing that oligomers of size $\sim$ 10 chains are more compact, while larger clusters have more extended geometries characteristic of fibrils. The spacing between consecutive layers of pairs of chains in fibrils is 4.52 ± 0.10 Å, and the pitch of the helical twist of the paired layers is 127 ± 9 Å; these are relatively independent of temperature (Fig. S2b). This corresponds to 28 ± 2 layers in a full 360° rotation, or 12.8° ± 0.9° of rotation per layer. Compared to typical tau fibril geometries, *e.g.*, 4.7 Å spacing and ≈ 1° rotation in Alzheimer's disease paired helical filaments (PHFs) (5), the simulated PHF6 fibrils show slightly closer spacing and much tighter helical twisting. However, the angles observed here are in very good agreement with the theoretical investigation of PHF6 fibril geometries by Yang *et al.* (15) giving $\sim 11°$–$14°$. The small cross-sectional areas of these fibrils compared to tau fibrils are also likely more permissive of a larger twist angle, consistent with experimental observations of decreasing rotation in β-lactoglobulin amyloid fibrils of increasing numbers of protofilaments (44). Finally, the reduced layer spacing could be explained by a rotation of the chains in a fibril to accommodate twisting about its axis—such a rotation is evident in Fig. 1f.

The temperature dependence of the equilibrium PHF6 clusters reveals underlying thermodynamic and kinetic effects governing aggregation (Fig. 2a). At 370 K, most chains remain dissociated ($M = 1$) or otherwise form small clusters ($2 \leq M \leq 16$), but not larger aggregates ($M > 16$). At lower temperatures, fewer chains are found dissociated or in small oligomers, with most in large aggregates, as is also evident from Fig. 1b–d. The time dependence of these fractions (Fig. S3a–c) reveals that fibrillization spanning 320–350 K begins nearly immediately, and that the distributions of cluster sizes reach equilibrium within a few μs. At 360 K, the fibril growth rate is much lower, and longer simulations are likely necessary to equilibrate the system. It is unclear from these results alone whether or not the absence of aggregation at 370 K is due to thermodynamic instability of the fibrils, or that the timescale of fibril nucleation is simply significantly longer than the simulation; we address this issue further in *Aggregation Phase Behavior*.

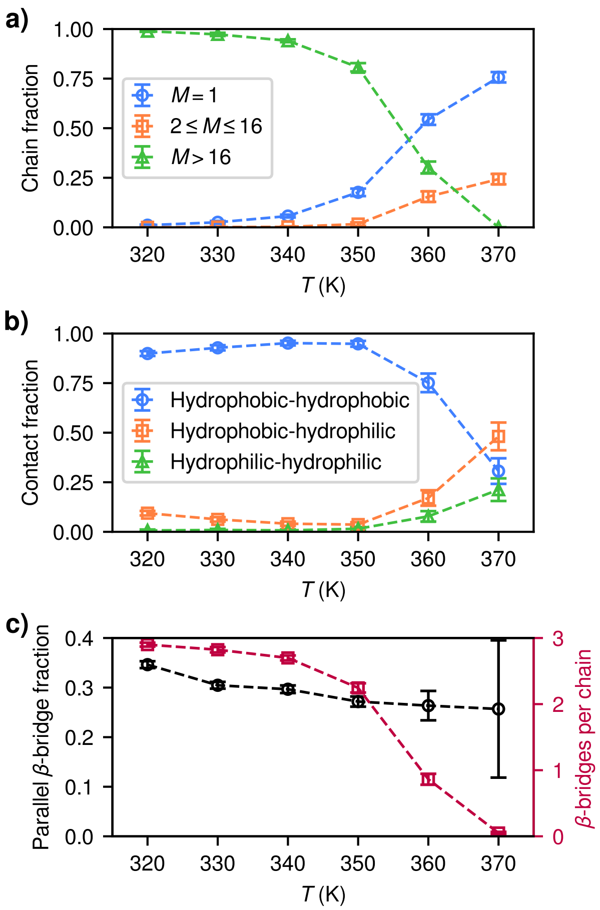

Fig. 2. Temperature dependence of properties of aggregating systems. (a) Fractions of isolated chains ($M = 1$) and chains in clusters of sizes $2 \leq M \leq 16$ and $M > 16$, from the last 5 μs of 10 μs MD simulations of 512 PHF6 with $\rho = 10$ mM. (b) Likewise, fractions of non-hydrogen-bonded contacting chains for which the closest pair of sidechains consisted of residues from the hydrophobic and hydrophilic sides of PHF6. (c) Fractions of β-bridge-participating residues with parallel (*vs.* antiparallel) orientations, and total numbers of β-bridges per chain. Error bars show standard deviations of the measured properties.

The model predicts appropriate structural ordering of the PHF6 chains within their fibrils, as illustrated in Fig. 2b showing the fractions of closest, non-hydrogen-bonded contacts (in opposing protofilaments) as a function of temperature. We classify contacts by the participating residues: V^{306}, I^{308}, and Y^{310} make up the “hydrophobic side” of PHF6, and Q^{307}, V^{309}, and K^{311} comprise the “hydrophilic side,” a division based on the steric zipper crystal structure of PHF6 (12, 45). When fibrils form, the majority of closest contacts are between sidechains on the hydrophobic sides of PHF6 (Fig. 2b), confirming that the model favors assembly into fibrils enclosing these sidechains in their interiors. On the other hand, at 370 K where no fibrils form, the distribution of contacts approaches what would be expected for random, transient, and unbiased association of monomers in solution (roughly 50% for hydrophobic-hydrophilic contacts, and 25% each for same-side contacts). A slight enhancement in contacts involving the hydrophobic side suggests that hydrophobic interactions are still slightly favored at high temperatures, even though significant enrichment in hydrophobic contacts only occurs in fibrils. Fig. S3d–f illustrates that these contact fractions evolve from initial solution to final fibril values on similar timescales as fibril assembly itself. It is possible that the slight decrease in hydrophobic-hydrophobic contact fractions at the lowest temperatures is due to kinetic arrest in the quickly growing fibrils: at 320 K, most chains are depleted from the solution by 1 μs and the fibril structure remains effectively “frozen in” from this point on. At, *e.g.*, 350 K, however, this occurs more slowly, likely allowing for more annealing during fibril growth.

Under the conditions here, the overall fibrillization process occurs predominantly through primary nucleation from solution followed by fibril growth from attachment of monomers to fibril ends. To probe the growth process, we characterize chains that irreversibly attach to fibrils (Fig. S4) before and after the moment that contact is made and not subsequently broken. The probability of a chain-fibril contact rises steadily $\sim$ 300 ns prior to this, suggesting reversible attachment and detachment; interestingly, an increase in contacts involving residues on the hydrophilic sides of PHF6 chains occurs during this period. However, formation of β-sheet secondary structure and intermolecular hydrogen bonding, along with the hydrophobic contacts characteristic of the formed fibrils, takes another $\sim$ 300 ns to be established. That reversible association is seen leading up to integration of a chain into a fibril, and that attaching chains do not immediately adopt their final conformations, suggests either a “dock-and-lock” mechanism (46) for attachment, or that it is necessary for additional monomers to attach on top of an associated chain before it is stabilized within a fibril.

The model also suggests that a delicate balance of interactions governs the orientations (parallel and antiparallel) of β-bridges within fibrils. Fig. 2c indicates a bias towards antiparallel alignment, but with a significant amount of orientational disorder, slightly increasing at lower temperatures as the parallel fraction increases towards 50%. In contrast, experiments suggest that PHF6 forms fibrils with parallel β-sheets (12, 19). Interestingly, however, the CG model does reproduce the parallel-antiparallel balance present in the reference AA system, which forms $\sim$ 20–30% parallel β-bridges (Fig. S5). Several factors are known to influence PHF6 fibril stability, such as electrostatic, π-π, and CH-π interactions (47, 48), and these can vary between atomistic force fields (49), which predict different chain alignments within β-sheets (50). Our model suggests a free energetic bias towards antiparallel structures of $\sim$ 1.0–1.5 k_BT, so only a small change in force field interaction strengths is necessary to shift it. N-terminal capping, known to control PHF6 fibrillization (51) as well as affect morphologies of fibrils formed from fragments of islet amyloid polypeptide (52, 53), may also influence experimental findings for PHF6 structures. Sawaya *et al.* (12) provide a structure of uncapped, zwitterionic PHF6 with parallel β-sheet alignment, although notably this structure is for *microcrystalline* rather than *fibrillar* PHF6, and it involves positively-charged lysine sidechains forming favorable electrostatic contacts with negatively-charged C-termini of chains in adjacent β-sheets in the crystal. However, although Goux *et al.* (19) do not attempt to obtain atomistic structures of fibrillar PHF6 with neutral terminal caps, they report FTIR spectra consistent with the absence of antiparallel β-sheets.

In the end, experimental evidence points towards predominantly parallel orientation of the chains in PHF6 fibrils, although the amount of orientational disorder in real fibrils is unknown. That fibrils grow robustly in our simulations despite significant mixing of parallel and antiparallel β-bridges throughout suggests that this kind of order may not be of great importance for PHF6 fibrillization specifically. This contrasts with aggregation of longer tau, which by necessity must form parallel cross-β structures in order to stack into regular fibrils of complex folds; if the PHF6 region is involved in initial aggregation stages as a recent study (20) suggests, this raises the interesting possibility of a further aggregation barrier involving strand alignment ordering. The lack of strong specificity, echoed by the variety of behaviors observed in force field predictions, has been observed in AA simulation studies of PHF6 aggregation (54, 55). Moreover, a PHF6 fibrillization study (22) using the CG model of Bereau and Deserno (33) found parallel orientation of chains within β-sheets and antiparallel orientation across sheets, consistent with experiments. On the other hand, Cheon *et al.* (35) showed that the PRIME20 CG model (34) favors antiparallel PHF6 β-sheet formation. Notably, however, adjustment of the PRIME20 hydrogen bonding interaction distance constants by just 5% causes PHF6 fibrils with nearly perfect parallel chain orientation to grow. This again highlights the delicate balance of interactions affecting cross-β orientation in PHF6 aggregation, compared to the much more robust preference for hydrophobic sidechain contacts inside fibrillar cores that our model captures accurately.

Configurational Landscape for Oligomerization. Although direct MD simulations of aggregation can reveal general information about fibril structure, they do not distinguish clearly between the roles of thermodynamics and kinetics in governing fibril formation. For instance, simulations in which no fibrils nucleate cannot prove that fibrils are thermodynamically unstable at the simulated conditions as opposed to kinetically inaccessible due to a large nucleation barrier. To more rigorously probe the initial stages of PHF6 aggregation, we use advanced sampling replica exchange MD (REMD) simulations of $N = 2$, 4, 8, and 16 total chains to map out the equilibrium configurational landscape of small oligomers. Fig. 3a–c shows representative structures from these simulations for oligomer sizes $M = 4$, 8, and 16 at 300 K.

At low temperature (300 K), the systems spend most of their time in single aggregates that reveal a hierarchy of stable states. For the $N = 2$ system, these are simply dimers of two β-strands. At $N = 4$, the predominant oligomeric structure consists of a single β-sheet (Fig. 3a). The largest system ($N = 16$) forms paired sheets (Fig. 3c) similar to small sections of fibrils, while the $N = 8$ system forms cylindrical structures (Fig. 3b) distinct from either single or paired β-sheets. At 350 K, these aggregates break into multiple oligomers of smaller sizes. We use simulation reweighting techniques incorporating data from all temperatures to reveal the probability of oligomeric structures of all sizes and types at 300 K. Fig. S6 shows representative images of the most probable oligomeric structures for all sizes $M = 2, 3, \ldots, 16$ along with the fractions of all possible structures of these sizes that they comprise. Importantly, these results show that single β-sheets predominate for $2 \le M \le 7$, the unique cylindrical structure appears at $M = 8$, and paired β-sheets are present for $9 \le M \le 16$.

The most probable oligomers of all sizes at 350 K (see Fig. S7), where the systems sample a wide variety of configurations, reveal more diverse behavior. At this temperature, oligomers of size $4 \le M \le 7$ show β-sheets of size $M - 1$ with single chains associated with their surfaces. For $8 \le M \le 10$, the same behavior is seen with 2 or 3 associated chains, until paired sheets appear for $11 \le M \le 16$. Overall, the results suggest that backbone hydrogen bonding and β-sheet formation is crucial for the initial creation of small oligomers from individual chains. As the oligomers grow, hydrophobic interactions between sidechains draw in additional chains to the forming β-sheet surfaces, until eventually paired cross-β structures characteristic of fibrils are able to form. This is supported by the sidechain contact types observed in oligomers of each size at 350 K (Fig. 3d). For small oligomers ($M \lesssim 6$), the distribution of contact types is similar to that in solutions without aggregates (see, *e.g.*, 370 K in Fig. 2b), but evolves towards that of extended fibrils as M increases. A slightly greater preference for antiparallel alignment is seen in smaller oligomers (Fig. S8), suggesting that sidechain interactions may bias larger oligomers towards increased parallel alignment where they are of greater importance compared to hydrogen bonding alone.

The cylindrical β-barrel-like oligomers in the $N = 8$ simulations are particularly striking in several ways beyond their unique hydrogen bonding topology. Almost all of these oligomers show antiparallel alignment for all of their β-bridges; all of the adjacent pairs of β-strands are out of register; moreover, the hydrophobic V^{306}, I^{308}, and Y^{310} residues point to the outside of, rather than towards the interior of, the oligomer (Fig. S9). Notably, the octamers seen here are structurally similar to trimeric and hexameric "cylindrins" first observed in oligomers of the K11V fragment of αB-crystallin (56), and to tetramers of an A$\beta_{30\text{–}36}$-containing construct (57). Specifically, Liu *et al.* (57) demonstrate the formation of cylindrin-like tetramers using model constructs designed to aggregate out-of-register, and posit a mechanism for toxicity involving membrane disruption from insertion of the cylindrins aided by their exposed hydrophobic residues, structurally similar to the PHF6 oligomers predicted here. Whether or not the PHF6 cylindrin octamer has any bearing on the potential toxicity of longer PHF6-containing tau fragments is unclear; however, the unique octamer structure nonetheless provides an interesting hypothesis about potentially functional oligomeric structures of PHF6. In any case, an in-register fibril cannot grow from a cylindrin containing out-of-register hydrogen bonds without it first unrolling, then breaking and re-forming these bonds, suggesting that such oligomers could contribute to a nucleation barrier or off-pathway state in PHF6 fibrillization.

To directly assess nucleation barriers, we calculate oligomer free energies βF_M for a $\rho = 10$ mM solution (Fig. 3e). At higher temperatures where aggregation is unfavorable, barriers increase with M, presumably due to the entropic penalty for association of free chains into oligomers, while at lower temperatures, they level out and begin to decrease for sufficiently large M. This is evident for $T \lesssim 360$ K, in agreement with the large-scale assembly simulations. As the temperature further decreases, the critical oligomer sizes and

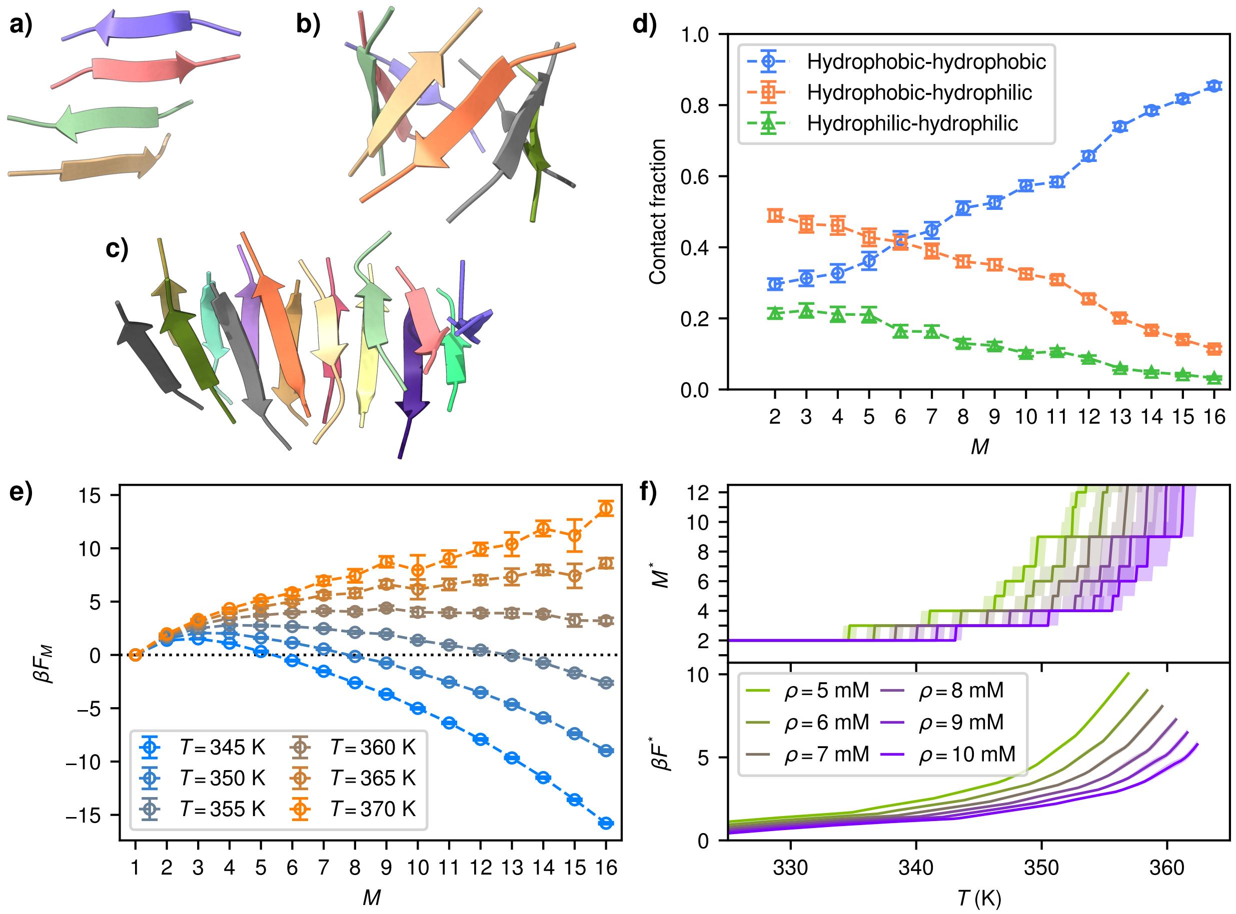

Fig. 3. Oligomer structures and thermodynamic properties. (a-c) A representative tetramer, octamer, and 16-mer, respectively, from replica exchange simulations of PHF6 at $T = 300\ \mathrm{K}$ and $\rho = 10\ \mathrm{mM}$. (d) Fractions of non-hydrogen-bonded contacting chains for which the closest pair of sidechains consists of residues from the hydrophobic and hydrophilic sides of PHF6, for a system with $N = 16$, $\rho = 10\ \mathrm{mM}$, and $T = 350\ \mathrm{K}$. (e) Free energies of oligomers of size M in a $\rho = 10\ \mathrm{mM}$ solution. (f) Critical oligomer sizes estimated from free energies (upper panel), and corresponding free energy barrier heights (lower panel: curves are truncated when the most probable value of M^* reaches the largest system size $N = 16$). Error bars and bands show 95% confidence intervals of mean values, or in the case of M^*, median values.

free energy barrier heights, corresponding to the locations of the maxima in these free energy curves, also decrease. The dependence of these sizes and heights on temperature is shown in Fig. 3f for different solution concentrations. As expected, with increased supersaturation (increasing concentration at constant temperature, or decreasing temperature at a given concentration), fewer monomers must oligomerize to form a critical aggregate, and the free energy penalty is decreased.

Aggregation Phase Behavior. Beyond small PHF6 oligomers representative of the start of fibril growth, we investigate how larger fibrillar aggregates interact with chains in solution. Here, we first initialize MD simulations with seeds (see Fig. 4a) based on a known, parallel β-sheet crystal structure of PHF6 (12) (PDB: 2ON9) containing the hydrophobic interface characteristic of the paired PHF6 fibrils observed to grow in our simulations. Similar seeding approaches have been used with other CG models to understand fibril elongation and protofilament assembly (58, 59). Here, the PHF6 seeds are placed in solutions of additional monomers, and we monitor the dynamic equilibrium of chains exchanging between the fibrillar and solution phases. To begin, we choose a fixed temperature of 350 K and perform five independent replicate runs each at *overall* concentrations of 5, 7, and 9 mM, *i.e.*, the solution concentrations that would be present were the seeds to dissolve completely.

The simulations reveal a significant dynamic exchange of monomers with the fibrils. Fig. 4b shows an example of the state of a fibril in one such simulation after 1 μs (see also Movie S2). Many of the seed chains initially present (orange) remain, while additional chains (blue) join both ends, mixing with the originals. Overall, seeds grow in the 7 and 9 mM simulations, while those in the 5 mM simulations remain roughly around the same size (Fig. 4c, solid lines), illustrating the stability of the seed structures, and the ability of the model to reach an *equilibrium* coexistence between solution and fibril phases. The number of chains starting in the seeds that never detach into the solution phase (Fig. 4c, dashed lines) decreases with time, showing that chains quickly detach from the ends of fibrils and move about in solution, even if they might later rejoin the fibrils. Similarly, the fraction of parallel β-bridges (Fig. S10a) decreases over time from the initially purely parallel state of the seeds towards the expected value for fibrils grown directly from solution with the CG model. This occurs progressively through series of attachment and detachment events rather than complete

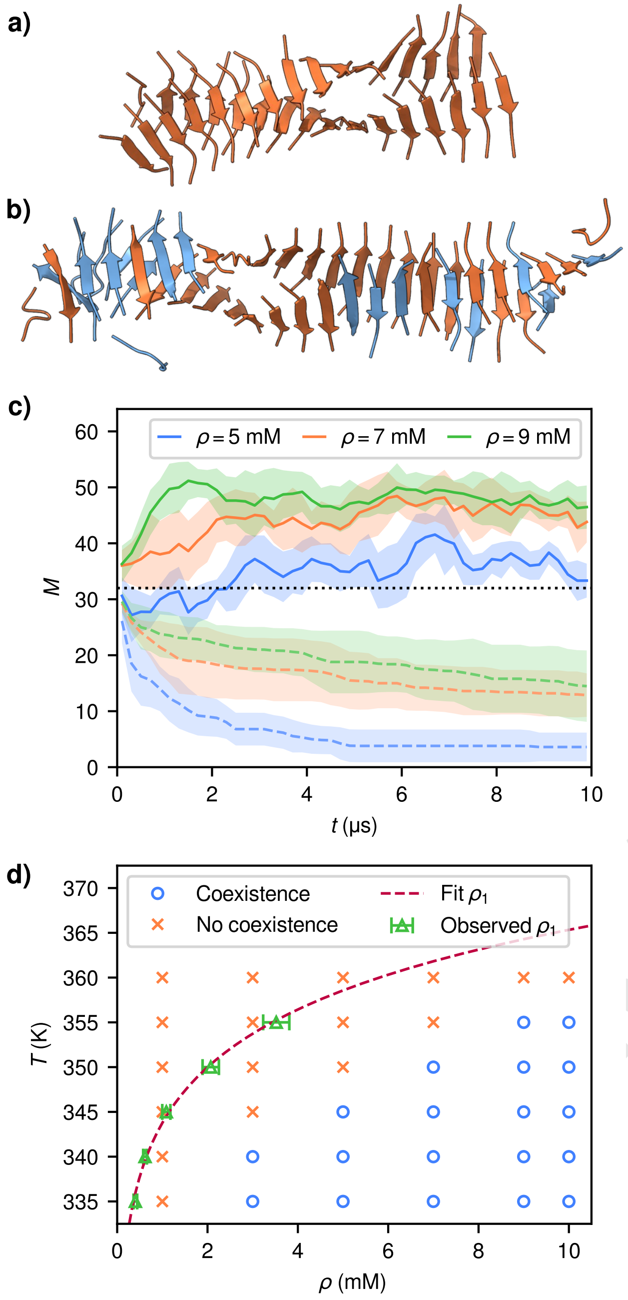

Fig. 4. Fibril seeding and fibril-solution coexistence. (a) Initial state and (b) state after 1 μs, of a fibrillar seed of size 32 (orange) in a solution of 32 PHF6 (blue) in a volume corresponding to a total concentration $\rho = 9$ mM, and $T = 350$ K. (c) Fibril size (solid) and chains never having detached from the fibril (dashed) from 5 replicates each, $T = 350$ K. Means and standard deviations over 200 ns intervals are shown. (d) Theoretical coexistence curve (red – –) of fibrillar PHF6 with solution, fit to observed solution concentrations (green △) in equilibrium with fibrils. Blue ○ and orange × symbols indicate the presence or absence, respectively, of persistent aggregates in 25 μs simulations of 64 PHF6. Error bars show 95% confidence intervals.

loss and recovery of fibril structure, shown by retention of hydrophobic-hydrophobic side contacts (Fig. S10b–d) throughout the equilibration process. However, it is clear that at higher concentrations, fibrils equilibrate at a notably slower rate, reflecting the stability of larger fibrils whose interior monomers are far less kinetically accessible than those at the ends.

Based on these findings, the model supports a picture of PHF6 aggregation involving a dynamic equilibrium between fibrils and solution. To map its phase diagram, we find the concentrations of PHF6 solutions at given temperatures that are in equilibrium with fibrils, *i.e.*, the concentrations at which fibrils will neither grow nor shrink on average over time. The blue and orange symbols in Fig. 4d show the temperatures and concentrations at which fibrils do or do not, respectively, both grow over the first 5 μs of 25 μs MD simulations of 64 PHF6, starting from solutions, and persist for the duration of the simulations. Once a fibril grows from a solution of initial concentration ρ, the chains not incorporated into the fibril remain in solution, at a concentration $\rho_1 < \rho$.

Fig. 4d shows the simulation-measured coexistence concentrations ρ_1 for each T, alongside a fit to a thermodynamic phase boundary scaling law for aggregation discussed in *Methods*. The latter predicts $\rho_1 \sim e^{\beta \Delta \epsilon_F}$, where $-\Delta \epsilon_F$ is the latent heat of fibrillization. This simple van't Hoff form, which assumes that the heat capacity of a chain in solution is identical to that of one in a fibril, agrees with prior studies of the phase behavior of amyloid systems (60–62). We find here $-\Delta \epsilon_F = 26.7 \pm 2.3$ kcal/mol. Although we are not aware of a previously measured value of this quantity for PHF6, it is of a reasonable magnitude, considering, *e.g.*, the value of ~ 22 kcal/mol reported by Wang *et al.* (62) for paired β-sheet aggregates of the Aβ_{16-22} (KLVFFAE) peptide, although unlike these aggregates, we do not observe PHF6 fibrils with more than two β-sheets in cross-section.

We find evidence of significant metastability in fibril equilibrium. Many of the state points falling just to the right of the coexistence curve in Fig. 4d do not produce fibrils in the 25 μs simulations. This behavior is consistent with the finite size and length of the simulations and the anticipated barrier to fibril nucleation observed in oligomer free energies (Fig. 3f), since a sufficiently weakly supersaturated solution may fail to nucleate fibrils on any timescale if the concentration ρ_1 of chains remaining in the solution after formation of a critical nucleus falls below the coexistence concentration. Alternatively, timescales associated with fibril formation and dissolution may simply be very long, illustrated by the 7 mM, 355 K case (Fig. S11a), contrasted with 5 mM at the same temperature where no aggregation occurs, and the 9 mM case where it occurs immediately. At each of these concentrations, we take configurations containing the largest fibrils grown in simulations at 350 K and simulate 5 independent replicates at 355 K. Fig. S11b–d shows that at 5 mM, 4 out of 5 seeds dissolve by 10 μs, at 7 mM, 1 seed dissolves, and at 9 mM, all remain stable despite shrinking on average. Thus, at conditions close to the saturation line, fibrils in these systems of finite size may be unstable or metastable. Nonetheless, we are still able to extract the temperature-dependent equilibrium concentrations from simulations that do eventually produce fibrils reaching coexistence with a solution phase.

As a final check to demonstrate that these simulations indeed reach equilibrium, we extract rate constants for association and dissociation of monomers from the seeded growth simulations in Fig. 4c as well as simulations of isolated seeds partially dissolving. We solve a master equation describing the time evolution of the seed length probability distribution, similar to approaches previously applied to amyloid systems, often in the context of investigating primary and secondary nucleation mechanisms and fragmentation (63, 64). For this system, we always observe primary nucleation followed by growth *via* attachment, so our main interest is to consider the growth kinetics. Demonstrations of correspondence between seeded simulations and solutions to the master equation are provided in Fig. S12 (see *Methods* for details). From the rate constants, we then estimate at $T = 350$ K that $\rho_1 = 2.17 \pm 0.17$ mM, consistent with $\rho_1 = 2.00 \pm 0.16$ mM from the fitted model shown in Fig. 4d.

Influence of Heparin as a Model Polyelectrolyte. Experimentally, tau fibrillization is often induced by an anionic polyelectrolyte such as heparin or RNA (65). Heparin in particular has been shown not only to act as an initiator for tau aggregation, but also to remain associated with the grown fibrils, reacting stoichiometrically to form stabilizing complexes with tau rather than simply catalyzing fibril nucleation (66, 67). Heparin has furthermore been shown experimentally to enhance PHF6 aggregation and affect fibril morphology in a capping group-dependent manner (51). To quantify the molecular mechanisms behind the influence of polyanions like heparin on aggregation, we develop a CG model for heparin based on known experimentally-determined solution conformational ensembles (see *Methods* for details). In this model, each saccharide unit in a heparin chain is represented as a single negatively charged CG site, with a screened electrostatic attractive interaction with the positively charged lysine sidechain sites of PHF6. We do not attempt to capture here the chemical details of PHF6-heparin interactions, but instead probe the influence of heparin as an idealized, model anionic polyelectrolyte on PHF6 aggregation.

Simulations of 64 PHF6 at $T = 350$ K and a range of concentrations, along with single heparin chains of 60 saccharide units (~ 17 kDa), show that heparin is capable of initiating PHF6 aggregation under conditions where it would not occur spontaneously. Heparin serves as a nucleation site for aggregation (see Fig. 5a), but also remains associated with elongated fibrils (see Fig. 5b and Movie S3). At the lowest PHF6 concentration tested, $\rho = 2.0$ mM, we do not observe fibril formation in any of 5 independent replicate simulations after 10 μs: Fig. 5c shows that most chains are found as isolated monomers in solution in this case. At $\rho = 2.5$ mM, 4 of 5 replicates form fibrils within 10 μs; at $\rho = 3.0$ mM, all replicates formed fibrils within the first 8 μs, and at higher concentrations, fibrillization begins almost immediately (see Fig. S13). In contrast, in identical conditions in the absence of heparin, no fibrillization occurs within 10 μs for $\rho = 3.0$ mM and 4.0 mM, while fibrils only form spontaneously without heparin within this timespan once the PHF6 concentration is increased to $\rho = 5.0$ mM (Fig. S14a–c). Replacing the single heparin 60-mer at $\rho = 3.0$ mM (where Fig. S13b shows fibrillization in all replicates) instead with 15 heparin tetramers, also leads to no aggregation (Fig. S14d); the same is true when the heparin-PHF6 electrostatic attraction is explicitly excluded from the force field (Fig. S14e).

As these electrostatic interactions between heparin and the lysine sidechain sites are the only attractive interactions between the modeled heparin and PHF6 chains, they appear to play a crucial role in enhancing the ability of fibrils to nucleate. The loss of heparin's cofactor effect upon its fragmentation highlights the importance of its *poly*anionic

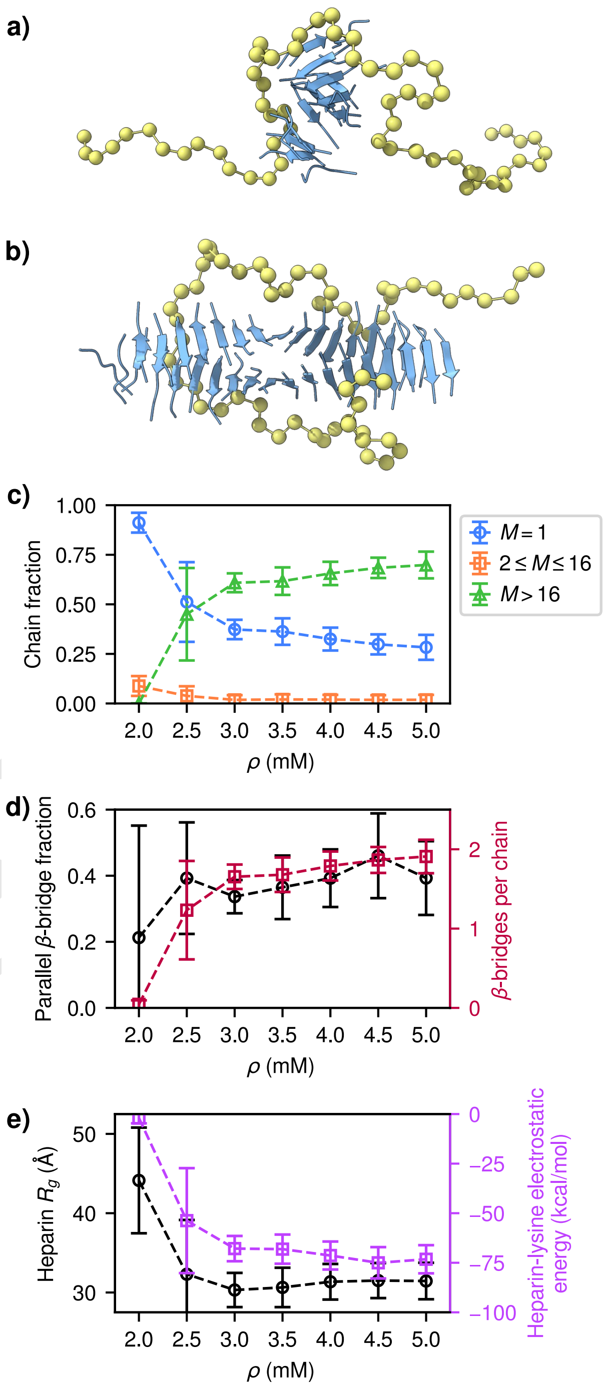

Fig. 5. Heparin-induced aggregation of PHF6. (a) Initial stages of heparin-induced aggregation at $T = 350$ K and $\rho = 2.5$ mM from a system of 64 PHF6 containing a heparin chain of 60 saccharide units. (b) Fully grown fibril under these conditions. (c) Fractions of chains in clusters of different sizes from the last 2 μs of 5 replicates each of 10 μs simulations of 64 PHF6 with a heparin chain of 60 saccharide units at $T = 350$ K. (d) Likewise, fractions of β-bridge-participating residues with parallel (*vs.* antiparallel) orientations, and total numbers of β-bridges per chain. (e) Radius of gyration of heparin chains and total electrostatic interaction energy between heparin and lysine sidechain CG sites. Error bars show standard deviations of the measured properties.

nature: localization of charges in heparin is important for attracting PHF6 chains to form a sufficiently large oligomer. A natural question is whether heparin chains only serve to increase the local concentration of PHF6, or if heparin also orders assembling chains in some way that affects fibril structure. We do not observe any significant differences in the layer spacing or pitch of heparin-induced *vs.* spontaneously aggregating fibrils. However, an increased fraction of parallel cross-β structures might be expected if the negatively charged polyelectrolyte were to cause the positively charged lysine sidechains to align along it. We do observe some such alignment (Fig. 5d): for transient association of monomers in solution in the absence of aggregation, the parallel fraction is similar to what is observed without any heparin present, but concentrations yielding heparin-induced aggregates show an enrichment in parallel β-sheets. The time-dependence of the alignment of chains in heparin-associated fibrils (Fig. S15) shows that some nucleation events produce fibrils initially more enriched in parallel β-sheets, but this effect is not universal. Others form initial aggregates with similar parallel alignment fractions to those not induced by heparin, and either remain predominantly antiparallel or increase in parallel cross-β content over time. It is likely that the particular bias observed in a given fibril is dependent on the conformations of heparin at the time of nucleation and while associated with the extended fibril.

In general, the interplay between the structures of the fibrils and heparin may be complex. In fact, the fibrils themselves affect the conformations explored by heparin, illustrated by a decrease in the radius of gyration R_g of fibril-associated heparin compared to heparin in systems where fibrils are absent (Fig. 5e, left axis). This collapse, likely due to the favorable electrostatic interactions between heparin and the lysine sidechains (Fig. 5e, right axis), is also captured by a decrease in the average persistence length of heparin: at $\rho = 3.0$ mM, it decreases from 25.0 ± 1.2 Å before fibril formation (comparable with 30.5 ± 1.4 Å for the reference conformations and ~ 21 Å (68, 69) and ~ 45 Å (70) from other experiments) to 6.7 ± 0.4 Å after fibril formation. Notably, heparin's interaction with PHF6 fibrils is consistent with the observed behavior of heparin in other amyloid systems, such as full length tau as discussed above, as well as with an atomistic simulation study by Khurshid *et al.* (71) of heparin-assisted aggregation of the Aβ_{16-21} (KLVFFA) hexapeptide. The latter study observed significant flexing of aggregate-associated heparin chains compared to their normal, more extended solution conformations, and furthermore showed that artificially constraining heparin to remain extended decreases its ability to induce aggregation. We note, given that we do not use reference systems including both heparin and PHF6 interacting, that the magnitude of the changes in heparin chain properties upon PHF6 aggregation may not be quantitative; however, the results nonetheless provide insight into the influence of an anionic polyelectrolyte on PHF6 aggregation, and suggest future work to probe the specific role of chain stiffness. Specifically, these results highlight the active role of such a polyelectrolyte in the aggregation process, not only as a catalyst but also as a reactant in the structure of formed fibrils.

Conclusions

In this study, we have shown that a purely bottom-up CG model of PHF6 developed using relative entropy minimization is able to accurately capture the interactions necessary to model the fibrillar aggregation of PHF6. Notably, no manual tuning of interaction parameters, or input of experimental or simulated fibril geometry data, is needed; realistic fibrillization behavior arises from interactions determined solely from small-scale reference atomistic simulations not including fibril formation. The resulting simulated fibrils show important hallmarks of amyloid aggregation, namely, structures with long-range fibrillar order maintained through cross-β hydrogen bonding along their lengths, and in this case, helically wound paired protofilaments held together by hydrophobic interactions. Importantly, the model enables simulations of systems with sizes larger and timescales longer than are generally accessible through atomistic simulations alone. Significant acceleration of what would otherwise be intractably slow timescales for PHF6 chains exploring configurational space permits the mapping out of the conformational and free energetic landscape of oligomers, and the determination of equilibrium fibril phase behavior.

The CG simulations here reveal important features of PHF6 aggregation that may be informative in understanding the wider issue of fibril formation of longer tau fragments. While the model does not reproduce a strong bias of PHF6 towards parallel β-sheets as suggested by experiments, it demonstrates that, unlike in amyloid aggregation of much longer peptides where parallel alignment is necessary for any kind of fibrillar order, this is not of critical importance for PHF6. We observe significant disorder within fibrils with respect to parallel *vs.* antiparallel alignment, see that purely parallel alignment of chains in an initial fibrillar seed readily becomes more disordered upon exchange with chains in solution, and show that it can even be influenced by association of heparin with the fibrils. All of these processes can occur without loss of the properties actually crucial for fibril formation, namely, the formation of cross-β hydrogen bonds along protofilaments and of hydrophobic contacts in the interface between protofilaments.

Results highlighting the structural properties of oligomers point to the importance of these hydrophobic contacts for sufficiently large aggregates once paired β-sheets begin to form. The transition from a mixture of various kinds of sidechain contacts to the almost exclusive presence of contacts between sidechains on the hydrophobic sides of PHF6 chains as oligomer size increases suggests a cooperative effect governing oligomer structures: sufficiently many hydrogen bonds are needed to stabilize small oligomers that paired sheets do not form, but as the number of chains grows, more such contacts can form within these paired structures. Such a mechanism could be relevant in explaining the high kinetic barrier to primary nucleation of fibrils from larger tau fragments, where the pairing of protofilaments is partially replaced by peptides folding in particular ways so as to optimize the sidechain contacts formed. That heparin-induced PHF6 fibrils are stabilized by electrostatic attraction of heparin to lysine sidechains is also consistent with the known behavior of heparin complexation with larger tau fragments. Further studies with this coarse-graining methodology applied to larger fragments could shed more light on the structures of oligomers in these aggregates, the effects of seeding by multiple distinct fibrillar folds, and the role of heparin and other polyelectrolytes in structuring larger aggregates. In particular, extension to longer constructs may be particularly informative for understanding the interplay between the PHF6 region and other parts of the chain in organizing the initial stages of tau aggregation, as suggested by very recent experimental work (20).

Methods

Relative Entropy Coarse-Graining of PHF6. We used relative entropy minimization to parameterize a CG model of PHF6. The relative entropy of a CG system with respect to an AA reference system is

$$S_{\mathrm{rel}}\left(\boldsymbol{\lambda}\right) = \int \mathcal{P}_{\mathrm{AA}}\left(\mathbf{R}\right) \ln \frac{\mathcal{P}_{\mathrm{AA}}\left(\mathbf{R}\right)}{\mathcal{P}_{\mathrm{CG}}\left(\mathbf{R};\boldsymbol{\lambda}\right)}\, d\mathbf{R}, \qquad [1]$$

where $\mathcal{P}_{\mathrm{CG}}\left(\mathbf{R};\boldsymbol{\lambda}\right)$ is the probability of CG configurations $\mathbf{R}$ given CG model parameters $\boldsymbol{\lambda}$, and $\mathcal{P}_{\mathrm{AA}}\left(\mathbf{R}\right)$ is the corresponding distribution of AA configurations $\mathbf{r}$ mapped to CG coordinates by the mapping $\mathbf{M} : \mathbf{r}^n \mapsto \mathbf{R}^N$. For a given AA reference system, model parameters $\boldsymbol{\lambda}$ are found by minimizing S_{rel}. The theory and implementation of this approach are detailed elsewhere (38–40). The relative entropy is extensible to multiple AA reference ensembles as

$$S_{\mathrm{rel,multi}}\left(\boldsymbol{\lambda}\right) = \sum_i S_{\mathrm{rel},i}\left(\boldsymbol{\lambda}\right), \qquad [2]$$

where $S_{\mathrm{rel},i}$ is the relative entropy for a given AA system and the corresponding CG ensemble. Such an extended-ensemble approach was proposed by Mullinax and Noid (72) in the context of force matching, and the form of Eq. 2 for $S_{\mathrm{rel,multi}}$ was justified by Sanyal *et al.* (43).

Here, we used an AA reference system of 3 interacting PHF6 peptides. Each PHF6 chain is capped with neutral acetyl and *N*-methylamine groups on its N- and C-termini (*i.e.*, Ac–VQIVYK–NHCH$_3$). N- and C-terminal capping has been shown to affect PHF6 fibril formation *in vitro*, with uncapped, zwitterionic PHF6 exhibiting lower aggregation propensity (51), motivating our choice of charge-neutral capping groups. We used the Amber ff14SBonlysc force field (73) with the GB-Neck2 generalized Born implicit solvation model (74) and the pwSASA method (75) for solvent-accessible surface area estimation in all reference AA simulations. The use of an implicit-solvent AA three-chain reference system is motivated by the need to achieve extensive sampling of conformational space using REMD, including associated and dissociated states, which becomes significantly more challenging with explicit-solvent models and particularly for larger chains. Further simulation details are given in SI Sec. A.

We used a formulation (76) of the multistate Bennett acceptance ratio (MBAR) method (77) for analysis and reweighting of REMD trajectories. The DSSP algorithm and its criterion for backbone hydrogen bonding (78, 79) identified hydrogen bonds and secondary structures. The REMD simulation of the reference system was reweighted to two temperatures $T_1 = 300.0$ K and $T_2 = 348.9$ K for use as two reference ensembles in Eq. 2. We chose T_1 as a representative lower temperature to include configurations with significant cross-β hydrogen bonding, and we set the higher temperature T_2 at the inflection point of the β-strand secondary structure fraction to include partially associated states (see Fig. S16a). T_2 also corresponds closely with the maximum of the trace of the Fisher information matrix of the relative entropy with respect to the model parameters (see Fig. S16b), which was shown by Shen *et al.* (80) to yield improved predictions from CG models.

We chose a mapping from centers of mass of atoms to CG site coordinates (Fig. 1a) including only heavy (non-H) atoms. The CG force field consists of bond, angle, and proper dihedral interactions, as well as pairwise interactions between sites, excluding 1-2, 1-3, and 1-4 bond connections. SI Sec. B gives details of the CG mapping and force field terms, and SI Sec. C describes the CG simulations used during model optimization. Figs. S17–21 show plots of the CG potentials, along with AA and CG probability distributions for associated degrees of freedom, and a Python script to tabulate potentials from their parameters is provided as SI.

Coarse-Grained Model Simulations and Analysis. We performed 5–25 μs NVT MD simulations with the optimized CG model at temperatures ranging between 320–370 K and concentrations between 1–10 mM, with 64–512 peptides in the systems; we chose other simulation parameters identically to the CG simulations used for model optimization. Initial configurations distributed peptides at random in the periodic simulation boxes and were equilibrated at 500 K for 1 μs, or used fibrillar seeds generated from the crystal structure of PHF6 due to Sawaya *et al.* (12) (PDB: 2ON9). Seeds containing 32 or 64 PHF6 (16 or 32 fibril layers) were relaxed for 150 ns using CG MD simulations with temperatures increasing from 20 K to 300 K in equal-length steps of 20 K (see Fig. S22). In addition to simulations at single temperatures, we performed CG REMD simulations of 2, 4, 8, and 16 peptides over a range of temperatures between 290–370 K, at concentrations of 10 mM (see SI Sec. D for details). Finally, we performed CG simulations to compare performance with implicit- and explicit-solvent AA systems (see SI Sec. E).

CG trajectories were analyzed by first mapping CG N, C, and O sites back to N, H, Cα, C, and O atoms. Motivated by other simple protein backmapping schemes (81, 82), we devised a geometric approach to identify coordinates of H, C, and O atoms from those of CG sites (see SI Sec. F). We then performed backbone hydrogen bond and secondary structure analysis with the AA trajectories. Fibrils, oligomers, and other aggregates were identified *via* cluster analysis, with two PHF6 chains considered to be in contact if any of their sidechain sites were within 7 Å. This cutoff was chosen based on sidechain-sidechain radial distribution functions (see Fig. S23) of the trajectories of Fig. 1. Oligomers isolated with this clustering criterion were characterized by the eigenvalue spectra of the adjacency matrices of their backbone hydrogen bond graphs; such spectra provide an efficient approximate method to distinguish graph topologies (83). Helix geometries and aggregate sphericities were found using principal moments and axes from C sites of individual peptides and C site centroids of peptides in clusters (see SI Sec. G). All renderings of CG trajectories were made with the UCSF ChimeraX (84, 85) molecular visualization software.

We used conformational ensembles of heparin chains containing 18, 24, 30, and 36 saccharide units (PDB: 4J8H, 4J8I, 4J8J, 4J8K) derived from solution scattering data (86, 87) as reference systems for relative entropy optimization of the CG polyelectrolyte model. We mapped all heavy atoms of a saccharide unit to a single site. We chose a screened Coulomb potential for heparin-heparin and heparin-lysine sidechain interactions, assigning -2 charges to heparin sites, and $+1$ to lysine sidechains. To avoid overlaps, an r^{-12} repulsive potential was applied between all heparin and PHF6 sites. Interaction functional forms used and their optimized or selected parameters are given in SI Sec. H.

Thermodynamics of Oligomerization and Aggregation and Growth Kinetics. To relate oligomer size distributions to free energies, we suppose that a canonical system of N molecules in a volume V is arranged such that there are N_1 unassociated monomers, N_2 dimers, *etc.*, and in general, N_M oligomers of size M, and that it behaves as an ideal mixture of oligomers. Its partition function can be written as

$$Q = Q_1^N \sum_{\{\mathbf{N}\}\in\mathcal{D}_N} \prod_{M=1}^{N} \frac{e^{-\beta N_M \Delta A_M}}{N_M!}, \qquad [3]$$

where Q_1 is the partition function for a single molecule, $\mathcal{D}_N$ denotes the set of all oligomer size distributions $\{\mathbf{N}\} \in \mathbb{N}^N$ such that $\sum_{M=1}^{N} M N_M = N$, and $\Delta A_M = A_M - M A_1$ where A_M is the free energy of an oligomer of size M in the volume V. The average number of oligomers of size M in the system is

$$\overline{N_M} = \frac{\sum_{\{\mathbf{N}\}\in\mathcal{D}_N} N_M \prod_{M'=1}^{N} e^{-\beta N_{M'}\Delta A_{M'}}/N_{M'}!}{\sum_{\{\mathbf{N}\}\in\mathcal{D}_N} \prod_{M'=1}^{N} e^{-\beta N_{M'}\Delta A_{M'}}/N_{M'}!}; \qquad [4]$$

a system of these equations can be solved for the ΔA_M given values of $\overline{N_M}$ from simulations. Unlike the $\overline{N_M}$, which depend on the finite size N of the system, these free energies can be used to predict the behavior of systems of any size (88). More details are given in SI Sec. I.

For large N, it is impractical to evaluate these expressions due to the magnitude of $|\mathcal{D}_N|$. Working instead from the grand canonical ensemble, it is possible to estimate the free energy for forming an oligomer of size M from a solution of monomers of concentration ρ in the limit $N \to \infty$ as

$$\beta F_M = \beta \Delta A_M - (M-1)\ln(\rho V_0), \qquad [5]$$

where V_0 is the volume to which ΔA_M corresponds. Alternatively, to analyze fibril-solution coexistence, we can assume constant and equivalent heat capacities of PHF6 in both phases, giving

$$\rho_1 = \rho_0 e^{\beta \Delta \epsilon_F}, \qquad [6]$$

where $\rho_1 = N_1/V$, $\Delta\epsilon_F$ is a fibrillization-associated energy, and ρ_0 is a constant. Details are provided in SI Sec. I.

To find rate constants for monomer-wise growth and dissolution of fibrils from the seeded simulations, we first assume that a system with a fixed number of chains N has a single fibril of size M along with $N_1 = N - M$ monomers. We consider the reactions $\mathrm{PHF6}_M + \mathrm{PHF6} \rightleftharpoons \mathrm{PHF6}_{M+1}$, with the rates $r_{M\to M+1} = \kappa_+\rho_1$ and $r_{M+1\to M} = \kappa_-$. For a given simulation, the constants are fit to fibril length data taken at periodic intervals $\tau = 100$ ns using a maximum likelihood approach on the probabilities $\mathcal{P}\left(M', t+\tau | M, t\right) = \mathbf{e}_M^T e^{\mathbf{A}\tau}\mathbf{e}_{M'}$ with $A_{M,M'} = r_{M\to M'} - \delta_{M,M'}\sum_{M''} r_{M\to M''}$. We find $\kappa_+ = 0.0223 \pm 0.0012$ ns^{-1} mM^{-1} and $\kappa_- = 0.0485 \pm 0.0026$ ns^{-1}. Since relative entropy coarse-graining establishes a rigorous AA-to-CG correspondence only with equilibrium configurations, not dynamics (89), unless a modified theoretical approach is used to explicitly account for them (90, 91), such rate constants should only be considered qualitatively. However, at equilibrium, we can meaningfully interpret the ratio $\kappa_-/\kappa_+ = \rho_1$. More details on this and the master equation model are given in SI Sec. J.

Data Availability. All data referenced in the article are presented within or in the *SI Appendix*. Values of the PHF6 CG model parameters are provided within the Python script given as SI.

ACKNOWLEDGMENTS. We gratefully acknowledge funding support from the National Science Foundation through Award No. CHEM-1800344. We acknowledge the W. M. Keck Foundation for a Medical Research Grant. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1650114. Use was made of computational facilities purchased with funds from the National Science Foundation (OAC-1925717) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara. We are grateful to Pritam Ganguly, Songi Han, Kenneth Kosik, Samuel Lobo, Andrew Longhini, Saeed Najafi, Joan-Emma Shea, and Michael Vigers for helpful discussions during the preparation of this work.

1. C Ballatore, VMY Lee, JQ Trojanowski, Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* **8**, 663–672 (2007).
2. M Goedert, DS Eisenberg, RA Crowther, Propagation of tau aggregates and neurodegeneration. *Annu. Rev. Neurosci.* **40**, 189–210 (2017).
3. D Sanders, et al., Distinct tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* **82**, 1271–1288 (2014).
4. J Vaquer-Alicea, MI Diamond, LA Joachimiak, Tau strains shape disease. *Acta Neuropathol.* **142**, 57–71 (2021).
5. AWP Fitzpatrick, et al., Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
6. B Falcon, et al., Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
7. B Falcon, et al., Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
8. W Zhang, et al., Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
9. Y Shi, et al., Structure-based classification of tauopathies. *Nature* **598**, 359–363 (2021).
10. P Ganguly, et al., Tau assembly: The dominant role of PHF6 (VQIVYK) in microtubule binding region repeat R3. *J. Phys. Chem. B* **119**, 4582–4593 (2015).
11. M von Bergen, et al., Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif ($^{306}\mathrm{VQIVYK}^{311}$) forming β structure. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5129–5134 (2000).
12. MR Sawaya, et al., Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
13. V Daebel, et al., β-Sheet core of tau paired helical filaments revealed by solid-state NMR. *J. Am. Chem. Soc.* **134**, 13982–13989 (2012).
14. H Liu, et al., Disclosing the mechanism of spontaneous aggregation and template-induced misfolding of the key hexapeptide (PHF6) of tau protein based on molecular dynamics simulation. *ACS Chem. Neurosci.* **10**, 4810–4823 (2019).
15. J Yang, MV Agnihotri, CJ Huseby, J Kuret, SJ Singer, A theoretical study of polymorphism in VQIVYK fibrils. *Biophys. J.* **120**, 1396–1416 (2021).
16. M Goedert, MC Spillantini, D Rutherford, RA Crowther, Multiple isoforms of human microtubule-associated protein tau: Sequences and localization in neurofibrillar tangles of Alzheimer's disease. *Neuron* **3**, 519–526 (1989).
17. AM Giannetti, et al., Fibers of tau fragments, but not full length tau, exhibit a cross β-structure: Implications for the formation of paired helical filaments. *Protein Sci.* **9**, 2427–2435 (2000).
18. W Li, VMY Lee, Characterization of two VQIXXK motifs for tau fibrillization in vitro. *Biochemistry* **45**, 15692–15701 (2006).
19. WJ Goux, et al., The formation of straight and twisted filaments from short tau peptides. *J. Biol. Chem.* **279**, 26868–26875 (2004).
20. S Lövestam, et al., Disease-specific tau filaments assemble via polymorphic intermediates (bioRxiv: 10.1101/2023.07.24.550295) (2023).
21. C Wu, JE Shea, Coarse-grained models for protein aggregation. *Curr. Opin. Struct. Biol.* **21**, 209–220 (2011).
22. FX Smit, JA Luiken, PG Bolhuis, Primary fibril nucleation of aggregation prone tau fragments PHF6 and PHF6*. *J. Phys. Chem. B* **121**, 3250–3261 (2017).
23. MS Li, DK Klimov, JE Straub, D Thirumalai, Probing the mechanisms of fibril formation using lattice models. *J. Chem. Phys.* **129**, 175101 (2008).
24. S Abeln, M Vendruscolo, CM Dobson, D Frenkel, A simple lattice model that captures protein folding, aggregation and amyloid formation. *PLoS ONE* **9**, e85185 (2014).
25. NS Bieler, TPJ Knowles, D Frenkel, R Vácha, Connecting macroscopic observables and microscopic assembly events in amyloid formation using coarse grained simulations. *PLoS Comput. Biol.* **8**, e1002692 (2012).
26. A Šarić, YC Chebaro, TPJ Knowles, D Frenkel, Crucial role of nonspecific interactions in amyloid nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17869–17874 (2014).
27. A Davtyan, et al., AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **116**, 8494–8503 (2012).
28. M Chen, NP Schafer, W Zheng, PG Wolynes, The associative memory, water mediated, structure and energy model (AWSEM)-amylometer: Predicting amyloid propensity and fibril topology using an optimized folding landscape model. *ACS Chem. Neurosci.* **9**, 1027–1039 (2018).
29. F Sterpone, et al., The OPEP protein model: From single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **43**, 4871–4893 (2014).
30. J Nasica-Labouze, M Meli, P Derreumaux, G Colombo, N Mousseau, A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide GNNQQNY from the yeast prion Sup-35. *PLoS Comput. Biol.* **7**, e1002051 (2011).
31. GG Maisuradze, P Senet, C Czaplewski, A Liwo, HA Scheraga, Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J. Phys. Chem. A* **114**, 4471–4485 (2010).
32. A Rojas, A Liwo, D Browne, HA Scheraga, Mechanism of fiber assembly: Treatment of Aβ peptide aggregation with a coarse-grained united-residue force field. *J. Mol. Biol.* **404**, 537–552 (2010).
33. T Bereau, M Deserno, Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.* **130**, 235106 (2009).
34. M Cheon, I Chang, CK Hall, Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins* **78**, 2950–2960 (2010).
35. M Cheon, I Chang, CK Hall, Influence of temperature on formation of perfect tau fragment fibrils using PRIME20/DMD simulations. *Protein Sci.* **21**, 1514–1527 (2012).
36. S Izvekov, GA Voth, A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **109**, 2469–2473 (2005).
37. S Izvekov, GA Voth, Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **123**, 134105 (2005).
38. MS Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **129**, 144108 (2008).
39. A Chaimovich, MS Shell, Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Phys. Chem. Chem. Phys.* **11**, 1901 (2009).
40. A Chaimovich, MS Shell, Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **134**, 094112 (2011).
41. SP Carmichael, MS Shell, A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *J. Phys. Chem. B* **116**, 8383–8393 (2012).
42. SP Carmichael, MS Shell, Entropic (de)stabilization of surface-bound peptides conjugated with polymers. *J. Chem. Phys.* **143**, 243103 (2015).
43. T Sanyal, J Mittal, MS Shell, A hybrid, bottom-up, structurally accurate, Gō-like coarse-grained protein model. *J. Chem. Phys.* **151**, 044111 (2019).
44. J Adamcik, R Mezzenga, Adjustable twisting periodic pitch of amyloid fibrils. *Soft Matter* **7**, 5437 (2011).
45. JJW Wiltzius, et al., Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16**, 973–978 (2009).

46. PH Nguyen, MS Li, G Stock, JE Straub, D Thirumalai, Monomer adds to preformed structured oligomers of Aβ-peptides by a two-stage dock–lock mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 111–116 (2007).
47. JH Zhao, et al., Molecular dynamics simulations to investigate the stability and aggregation behaviour of the amyloid-forming peptide VQIVYK from tau protein. *Mol. Simul.* **36**, 1013–1024 (2010).
48. K Sogawa, et al., CH-π interaction in VQIVYK sequence elucidated by NMR spectroscopy is essential for PHF formation of tau. *Biopolymers* **102**, 288–295 (2014).
49. Y Li, X Peng, Comparison of the force fields on monomeric and fibrillar PHF6 of tau protein. *Biophys. Chem.* **277**, 106631 (2021).
50. VH Man, X He, J Gao, J Wang, Effects of all-atom molecular mechanics force fields on amyloid peptide assembly: the case of PHF6 peptide of tau protein. *J. Chem. Theory Comput.* **17**, 6458–6471 (2021).
51. S Arya, et al., Terminal capping of an amyloidogenic tau fragment modulates its fibrillation propensity. *J. Phys. Chem. B* **124**, 8772–8783 (2020).
52. ST Wang, et al., Sequence-dependent self-assembly and structural diversity of islet amyloid polypeptide-derived β-sheet fibrils. *ACS Nano* **11**, 8579–8589 (2017).
53. X Zottig, et al., Guiding the morphology of amyloid assemblies by electrostatic capping: From polymorphic twisted fibrils to uniform nanorods. *Small* **15**, 1901806 (2019).
54. DW Li, S Mohanty, A Irbäck, S Huo, Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment. *PLoS Comput. Biol.* **4**, e1000238 (2008).
55. C Fagnen, J Giovannini, M Catto, AS Voisin-Chiret, J Sopkova-de Oliveira Santos, On the tracks of the aggregation mechanism of the PHF6 peptide from tau protein: Molecular dynamics, energy, and interaction network investigations. *ACS Chem. Neurosci.* **13**, 2874–2887 (2022).
56. A Laganowsky, et al., Atomic view of a toxic amyloid small oligomer. *Science* **335**, 1228–1231 (2012).
57. C Liu, et al., Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20913–20918 (2012).
58. J Zhang, M Muthukumar, Simulations of nucleation and elongation of amyloid fibrils. *J. Chem. Phys.* **130**, 035102 (2009).
59. J Sørensen, X Periole, KK Skeby, SJ Marrink, B Schiøtt, Protofibrillar assembly toward the formation of amyloid fibrils. *J. Phys. Chem. Lett.* **2**, 2385–2390 (2011).
60. S Auer, D Kashchiev, Phase diagram of α-helical and β-sheet forming peptides. *Phys. Rev. Lett.* **104**, 168105 (2010).
61. LG Rizzi, S Auer, Amyloid fibril solubility. *J. Phys. Chem. B* **119**, 14631–14636 (2015).
62. Y Wang, et al., Thermodynamic phase diagram of amyloid-β (16–22) peptide. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2091–2096 (2019).
63. TPJ Knowles, et al., An analytical solution to the kinetics of breakable filament assembly. *Science* **326**, 1533–1537 (2009).
64. TCT Michaels, TPJ Knowles, Mean-field master equation formalism for biofilament growth. *Am. J. Phys.* **82**, 476–483 (2014).
65. B Nizynski, W Dzwolak, K Nieznanski, Amyloidogenesis of Tau protein. *Protein Sci.* **26**, 2126–2150 (2017).
66. Y Fichou, et al., Cofactors are essential constituents of stable and seeding-active tau fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13234–13239 (2018).
67. Y Fichou, et al., Tau-cofactor complexes as building blocks of tau fibrils. *Front. Neurosci.* **13**, 1339 (2019).
68. SS Stivala, Physico-chemical studies of fractionated bovine heparin. V. Some low-angle X-ray scattering data. *Arch. Biochem. Biophys.* **127**, 795–802 (1968).
69. B Khorramian, S Stivala, Small-angle x-ray scattering of high- and low-affinity heparin. *Arch. Biochem. Biophys.* **247**, 384–392 (1986).
70. G Pavlov, S Finet, K Tatarenko, E Korneeva, C Ebel, Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *Eur. Biophys. J.* **32**, 437–449 (2003).
71. B Khurshid, et al., Heparin-assisted amyloidogenesis uncovered through molecular dynamics simulations. *ACS Omega* **7**, 15132–15144 (2022).
72. JW Mullinax, WG Noid, Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **131**, 104110 (2009).
73. H Nguyen, J Maier, H Huang, V Perrone, C Simmerling, Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* **136**, 13959–13962 (2014).
74. H Nguyen, DR Roe, C Simmerling, Improved generalized Born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* **9**, 2020–2034 (2013).
75. H Huang, C Simmerling, Fast pairwise approximation of solvent accessible surface area for implicit solvent simulations of proteins on CPUs and GPUs. *J. Chem. Theory Comput.* **14**, 5797–5814 (2018).
76. X Ding, JZ Vilseck, CL Brooks, Fast solver for large scale multistate Bennett acceptance ratio equations. *J. Chem. Theory Comput.* **15**, 799–802 (2019).
77. MR Shirts, JD Chodera, Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
78. W Kabsch, C Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
79. WG Touw, et al., A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **43**, D364–D368 (2015).
80. K Shen, et al., Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information. *J. Chem. Phys.* **153**, 154116 (2020).
81. LE Lombardi, MA Martí, L Capece, CG2AA: Backmapping protein coarse-grained structures. *Bioinformatics* **32**, 1235–1237 (2016).
82. MR Machado, S Pantano, SIRAH tools: Mapping, backmapping and visualization of coarse-grained models. *Bioinformatics* **32**, 1568–1570 (2016).
83. P Wills, FG Meyer, Metrics for graph comparison: A practitioner's guide. *PLoS ONE* **15**, e0228728 (2020).
84. TD Goddard, et al., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
85. EF Pettersen, et al., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
86. S Khan, J Gor, B Mulloy, SJ Perkins, Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: New insight into heparin–protein complexes. *J. Mol. Biol.* **395**, 504–521 (2010).
87. S Khan, J Gor, B Mulloy, SJ Perkins, Corrigendum to "Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: New insight into heparin–protein complexes" [J. Mol. Biol. 395 (2010) 504–521]. *J. Mol. Biol.* **425**, 1847 (2013).
88. JT Kindt, Accounting for finite-number effects on cluster size distributions in simulations of equilibrium aggregation. *J. Chem. Theory Comput.* **9**, 147–152 (2013).
89. J Jin, KS Schweizer, GA Voth, Understanding dynamics in coarse-grained models. I. Universal excess entropy scaling relationship. *J. Chem. Phys.* **158**, 034103 (2023).
90. P Español, I Zúñiga, Obtaining fully dynamic coarse-grained models from MD. *Phys. Chem. Chem. Phys.* **13**, 10538 (2011).
91. MA Katsoulakis, P Plecháč, Information-theoretic tools for parametrized coarse-graining of non-equilibrium extended systems. *J. Chem. Phys.* **139**, 074115 (2013).