

Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of PHF6 Peptide of Tau Protein

Viet Hoang Man,¹ Xibing He,¹ Jie Gao,² and Junmei Wang^{1*}

¹ *Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening
Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261, USA.*

² *Department of Neuroscience, The Ohio State University Wexner Medical Center, Columbus,
OH 43210, USA.*

E-mail: junmei.wang@pitt.edu

Abstract

Molecular dynamics (MD) simulations play a vital role in revealing the mechanism of amyloid aggregation that is crucial to the therapeutic agent development for Alzheimer's Disease. However, the accuracy of MD simulation results strongly depends on the force field employed. In our previous benchmark for 17 all-atom force fields on modeling of amyloid aggregation using A β ₁₆₋₂₂ dimer, we showed that AMBER14SB and CHARMM36m are suitable force fields for amyloid aggregation simulation, while GROMOS54a7 and OPLSAA are not good for the task. In this work, we continue assessing the applicability of atomistic force fields on amyloid aggregation using VQIVYK (PHF6) peptide which is essential for Tau-protein aggregation. Although, both A β ₁₆₋₂₂ and PHF6 peptides formed fibrils *in vitro*, the PHF6 fibrils are parallel β -sheets while A β ₁₆₋₂₂ fibrils are antiparallel β -sheets. We performed all-atom large-scale MD simulation in explicit water on the PHF6 dimer and octa-peptides systems using five mainstream force fields, including AMBER99SB-disp, AMBER14SB, CHARMM36m, GROMOS54a7 and OPLSAA. The accumulated simulation time is 0.2 milliseconds. Our result showed that the β -sheet structures of PHF6 peptides sampled by AMBER99SB-disp, AMBER14SB, GROMOS54a7 and OPLSAA are in favor of the antiparallel β -sheets, while the dominant type of β -sheet structures is parallel β -sheet by using CHARMM36m. Among the five force fields, CHARMM36m provides the strongest CH- π interaction that was observed in an NMR study. The comparison between our result and experimental observation indicates that CHARMM36m achieved the best performance on modeling the aggregation of PHF6 peptides. In summary, CHARMM36m is currently the most suitable force field for studying the aggregation of both amyloid- β and tau through MD simulations.

Keywords: Force Fields; Molecular Dynamics Simulations; Oligomerization; PHF6; Tau Protein; Amyloid Peptides;

INTRODUCTION

Amyloid- β ($A\beta$) peptide and Tau protein play many important physiological functions. $A\beta$ peptide, proteolytic byproducts of $A\beta$ protein precursor in an amyloidogenic pathway, serves as neuroprotectors and synaptic activity modulators.^{1,2} Tau, an intraneuronal protein, plays a crucial role in cytoskeleton stabilization by binding to microtubules (MT).³ However, $A\beta$ peptide and Tau protein can also aggregate into neurotoxic formations including insoluble fibrils, neurofibrillary tangles (NFT) of Tau protein and $A\beta$ senile plaques, and soluble oligomers.^{4,5} The accumulation of the intracellular NFT and extracellular deposits of $A\beta$ senile plaques are two pathological hallmarks in the brain of Alzheimer's Disease (AD) patients.⁶ $A\beta$ oligomers ($A\beta$ Os) instigate multiple facets of AD neuropathology, and Tau oligomers (TOs) cause neuronal damage, leading to neurodegeneration and traumatic brain injury.^{4,5} Understanding the amyloid aggregation is crucial to the therapeutic agent development for AD which is not curable yet. Therefore, numerous studies, using both experimental and computational approaches, have been conducted to investigate this interesting process. On the other hand, $A\beta$ peptides and Tau protein are intrinsically disordered proteins (IDPs) which are structurally flexible and have high aggregation propensity. These properties make a great challenge for experimental approaches to decode the conformation dynamics of IDPs at atomistic resolution.⁷ Fortunately, molecular dynamics (MD) simulations can complement experimental work to capture the atomistic picture of IDPs aggregation dynamics. In a MD simulation, the force field (FF) for describing atomistic interactions plays a key role to adequately sample conformation ensemble and appropriately model the dynamic procedure. In last decades, many biomolecular FFs have been developed to model folded and unfolded proteins,⁸⁻¹⁴ and specific FFs for modeling IDPs have emerged.¹²⁻¹⁴ Evaluation of those FFs on studying IDPs have been conducted.^{12,15-17} However, the performance benchmark of those FFs in describing a dynamic procedure like aggregation was limited.

Given its importance in the pathogenesis of AD, A β peptide is probably the most studied IDP in FF assessment and a review on this topic was presented in Ref. 17. Most studies focused on how well a FF reproduces experimental structural properties of A β peptides, including secondary structure, chemical shifts, CD spectrum, residual dipolar couplings, NMR spectrum, and collision cross sections. Nevertheless, few studies evaluated the kinetics of A β aggregation. In a pioneering study,¹⁸ Strodel and co-workers investigated the impact of GROMOS54A7 (G54a7),¹⁰ OPLS-AA (OPLS),⁸ AMBER03WS,¹² CHARMM22* (C22*),¹⁹ AMBER99SB*ILDN and AMBER99SB*ILDN with LRDI¹⁸ FFs on the aggregation of A β _{16–22} peptide and its mutants (F19L and F19V/F20V). They found that G54a7 and OPLS most strongly over stabilized protein-protein interactions, and as a result, A β peptides aggregated faster in G54a7 and OPLS simulations than in the other FF simulations. Thus, they recommended that these two force fields not being used to study protein oligomer formation.¹⁸ In our recent work, we compared the effects of 17 FFs from the AMBER, CHARMM, GROMOS, and OPLS FF families on the aggregation kinetics of A β _{16–22} peptides.¹⁷ We also observed that AMBER96,²⁰ OPLS, and the GROMOS family FFs over stabilized protein-protein interactions in the MD simulations, while CHARMM FFs including C22*, CHARMM36 (C36),²¹ and CHARMM36m (C36m),¹³ and AMBER FFs including AMBER99SB-ILDN²² and AMBER14SB (A14SB)¹¹ are the best candidates for studying amyloid aggregation. Recently, Strodel and co-workers evaluated additional FFs including AMBER99SB-DISP (A99DISP),¹⁴ C36m and C36mW,¹³ in studying the aggregation kinetics of the short A β peptides.²³ Note that C36mW is C36m in combination with a modified TIP3P water to increase protein-water interactions, therefore C36mW can be considered as pure C36m used with another water model. A99DISP,¹⁴ which was recently developed by D. E. Shaw Research to study both folded and unfolded proteins. Their results showed that A99DISP inhibited peptide aggregation due to the overestimation of the interactions between the peptides and water, while C36m and C36mW provided promising results and were recommended for peptide aggregation simulations.

Although all IDPs can aggregate, their aggregated formations may be different from a protein to another protein. For example, the amyloidogenic fragment NFGAILS of human islet amyloid polypeptide aggregated into anti-parallel β -sheet structure,²⁴ while fragment GNNQQNY from yeast prion Sup35 preferred forming parallel β -sheet fibril.²⁵ Hence, a FF need to be benchmarked on various peptides/proteins. Considering the significant impact of Tau protein aggregation on many neurodegenerative diseases,^{4,26} particularly AD, the aggregation of Tau is of great interest. There are more and more computational studies using different FFs to model Tau protein, recently.²⁷⁻³⁷ In this work, we planned to perform all-atom FF benchmark study on Tau aggregation. Tau is present as six isoforms ranging from 352 to 441 amino acids. The Tau MT binding repeats R1-R4 spanning residues 244-368 are related to many disease-linked modifications.³⁸ This region recapitulates much of the aggregation property of Tau-441 in animal models.³⁹ Experimental evidence has shown that the key hexapeptide, VQIVYK called PHF6 (residues 306-311 in R3), is essential for Tau aggregation and amyloid formation.⁴⁰ Atomic-resolution crystallographic study showed the PHF6 peptide formed fibrils with parallel ordering of the β -strands.⁴¹ Because of the limitation of available computational power, a short peptide, which somehow can represent for a long peptide, is usually chosen for amyloid aggregation study for the sake of computational efficiency. For example, for the case of A β peptides, the short A β ₁₆₋₂₂ peptide containing the central hydrophobic core (residues 17-21) is not sufficient to fully understand full-length A β (A β ₄₀ and A β ₄₂) peptides, but it is the essential segment for the full length A β protein fibrillization. It helps exploring fundamental aspects of the thermodynamics and kinetics of amyloid aggregation.^{42,43} As such, A β ₁₆₋₂₂ is extensively applied to replace the full length A β (A β ₄₀ and A β ₄₂) in studying amyloid aggregation. Similar to A β peptides, we used the PHF6 peptide to benchmark all-atom FFs for Tau aggregation. The FFs we considered in this work include A99DIPS, A14SB, C36m, G54a7 and OPLS. Those FFs are representatives of the major biomolecular FF families and are considered as the best candidates for studying IDPs.

MATERIALS AND METHODS

System design

We designed two simulation systems, 2PHF6 and 8PHF6, which contain two and eight PHF6 peptides, respectively, to study the aggregation mechanisms of PHF6 peptide. To prepare the two simulation systems, we first generated a databank of PHF6 monomeric structures collected from a 100-nanosecond (ns) NPT equilibrium simulation in explicit solvent at 310 K and 1 atm. We applied AMBER Tools⁴⁴ to construct the PHF6 peptide with the sequence of ACE-VQIVYK-NME, and Amber FF A14SB¹¹ to describe the peptide. 2200 TIP3P water molecules⁴⁵ were added to the simulation box and the simulation system was neutralized by adding Cl^- ions. 5000 monomeric structures from the second half of the NPT simulation were collected and stored in the monomeric structure bank. Next, we generated multiple 2PHF6 and 8PHF6 for MD simulations. To construct 2PHF6, two random monomers from the monomeric structure databank were randomly placed but with two conditions satisfied: (i) the distance of the mass centers of two peptides is in range of 9 Å to 12 Å; (ii) the minimum distance of two peptides is larger than 1.2 Å. To construct 8PHF6, eight random monomers from the monomeric databank were placed at 8 vertices of a 16 Å cube. In total, we generated one hundred different 2PHF6 and twenty different 8PHF6 systems to be used in intending MD simulations. Finally, each 2PHF6 or 8PHF6 system was placed at the center of an octahedron box solvated by an explicit water model, and Cl^- ions were added to neutralize the net charge of simulation systems. The minimum distance between the peptides and the edges of water box was at least 10 Å. The peptides were described by one of five FFs including A99DISP,¹⁴ A14SB,¹¹ C36m,¹³ G54a7¹⁰ and OPLS.⁸ The explicit water was represented by SPC for G54a7 and TIP3P for other FFs. The box size, volume, number of water molecules and peptide concentration were 48 Å, 85000 Å³, 2737 and 36 mM for the 2PHF6 simulation system, and 66 Å, 221000 Å³, 6990 and 60 mM for the 8PHF6 simulation system. Altogether, we produced 100 MD trajectories for 2PHF6 and 20 MD trajectories for 8PHF6, to investigate the oligomerization of PHF6 peptides

for each FF. The accumulated MD simulation time is 0.2 milliseconds (ms) for five FFs and the simulation details are presented below.

Simulation Details

The GROMACS 2018 package⁴⁶ was employed for all simulations. The solvated systems were minimized using the steepest descent method and were equilibrated for 1 ns at constant pressure (P) of 1 atm and temperature (T) of 310 K. The pressure and temperature of the simulations were controlled using the Berendsen coupling method⁴⁷ with a relaxation time of 0.1 picoseconds (ps) and the Bussi-Donadio Parrinello velocity scaling method⁴⁸ with a relaxation time of 1 ps, respectively. The NPT simulations were subsequently extended 200 ns (for 2PHF6 systems) or 1000 ns (for 8PHF6 systems) for sampling snapshots for post-analysis, resulting in 40 microseconds (μ s) of the total simulation time for each FF. The equations of motion were integrated using a leap-frog algorithm⁴⁹ with a time step of 2 femtoseconds (fs). The LINCS algorithm⁵⁰ was used to constrain the lengths of all covalent bonds with a relative geometrical tolerance of 10^{-4} . The van der Waals forces were calculated with a cutoff of 10 Å, and the particle mesh Ewald method⁵¹ was employed to treat the long-range electrostatic interactions. The non-bonded interaction pair list, with a cutoff of 10 Å, was updated every 5 fs. Periodic boundary conditions were applied to all of the simulations.

Data analysis

The association between two short peptides is characterized by intermolecular side chain-side chain contacts (N_{sc}) and intermolecular backbone hydrogen bond (H-bond). A side chain-side chain contact is formed if the distance between the centers of mass of two residue side chains is ≤ 6.5 Å. A H-bond is formed if the acceptor-donor distance is ≤ 3.5 Å and the acceptor-donor-H angle is less than 30° . The secondary structure contents classified into β , helix, turn and coil were calculated by using the STRIDE algorithm.^{52,53} Here, the helix content includes 3-10 helix, Pi helix and α -helix.

The radii of gyration (R_g) and solvent accessible surface areas (SASA) were calculated by using GROMACS tools.

Besides being classified into either parallel or anti-parallel β -sheet based on the directions of the two peptide strands, a β -sheet is further characterized by the length (kb) of extended strands. For a short peptide as PHF6, we classified the dimeric structures into seven states as shown in Figure 1, which are DOS for a disordered structure, PS1 for a parallel β -sheet with $1 \leq kb \leq 2$, PS2 for a parallel β -sheet with $kb = 3$, PS3 for a parallel β -sheet with $kb \geq 4$, AS1 for an antiparallel β -sheet with $1 \leq kb \leq 2$, AS2 for an antiparallel β -sheet with $kb = 3$, and AS3 for an antiparallel β -sheet with $kb \geq 4$. Using this classification scheme, we tracked the populations of the seven dimeric states as well as the transitions between those states in both 2PHF6 and 8PHF6 systems. The population of a state, P_{st} , was calculated by Eq. 1 for a 2PHF6 system, and Eq. 2 for an 8PHF6 system.

$$P_{st} = \frac{N_{st}}{N} \quad (1)$$

$$P_{st} = \frac{1}{N} \sum_{s=1}^N \left(\frac{\sum_{i,j(i \neq j)}^8 D_{i,j,st}}{\sum_{i,j(i \neq j)}^8 D_{i,j}} \right) \quad (2)$$

Where N is total number of the snapshots collected from the NPT simulations, N_{st} is number of dimeric structures in st state, D_{ij} is the number of contacted peptide dimers and $D_{i,j,st}$ is the number of contacted peptide dimers in st state in a snapshot. A peptide pair is considered as a contacted peptide dimer when they have at least one side chain-side chain contact. The parallel- β content ($p\beta$) and antiparallel- β content ($a\beta$) were calculated by Eqs. 3 and 4, where β is β content.

$$p\beta = \beta \frac{PS1 + PS2 + PS3}{PS1 + PS2 + PS3 + AS1 + AS2 + AS3} \quad (3)$$

$$a\beta = \beta \frac{AS1 + AS2 + AS3}{PS1 + PS2 + PS3 + AS1 + AS2 + AS3} \quad (4)$$

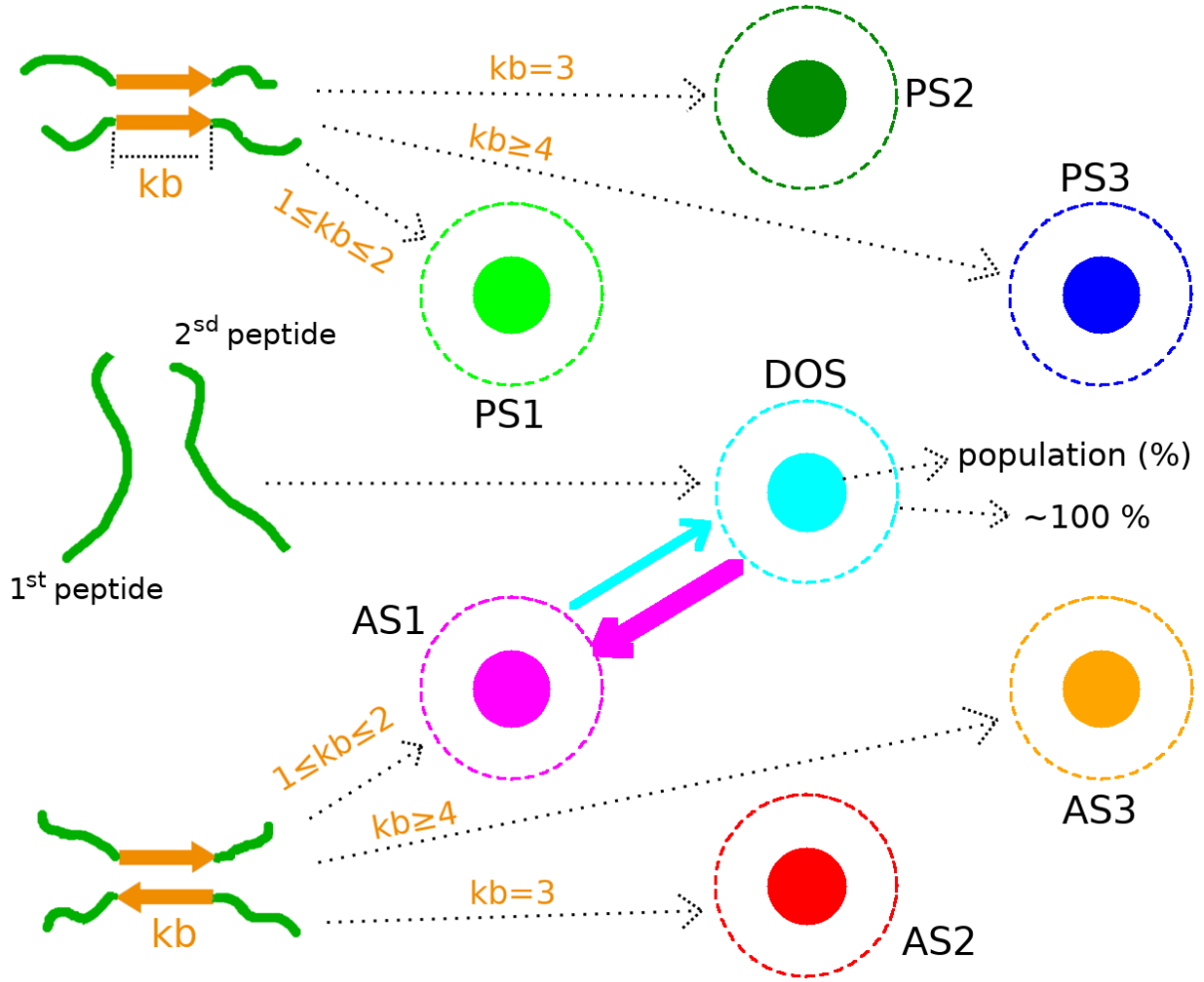


Figure 1: Transition between seven states of PHF6 dimers. The seven states of a dimer, including the disordered structure (DOS, in cyan color), parallel β -sheet with $1 \leq kb \leq 2$ (PS1, in light green color), parallel β -sheet with $kb = 3$ (PS2, in dark green color), parallel β -sheet with $kb \geq 4$ (PS3, in blue color), antiparallel β -sheet with $1 \leq kb \leq 2$ (AS1, in magenta color), antiparallel β -sheet with $kb = 3$ (AS2, in red color) and antiparallel β -sheet with $kb \geq 4$ (AS3, in orange color). The kb is number of residues with β structure in an extended strand of a β -sheet dimer. For each state, the area of solid circle demonstrates the state population, and the area of dash circle means 100% of the population. The transition from state i to state j is represented by an arrow which has color to be the same with the color of state i . The thickness of the arrow is directly proportional to the transition intensity.

Free Energy Landscape (FEL). The free-energy surface along the N-dimensional reaction coordinated $V = (V_1, \dots, V_N)$ is given by $\Delta G = -k_B T [\ln P(V) - \ln P_{max}]$, where $P(V)$ is the probability distribution represented by a histogram of MD data. P_{max} is the maximum of distribution, which is

subtracted to ensure that the lowest free energy minimum has ΔG of 0. The k_B and T are Boltzmann constant and simulation temperature, respectively. In this study, we used SASA per peptide ($SASA_{pp}$) and protein-protein (P-P) interaction energy as reaction coordinates for the two-dimensional FEL.

RESULTS

To ascertain that our data are well-equilibrated, we calculated the distribution of SASA per peptide ($SASA_{pp}$), R_g and number of intermolecular side chain-side chain contact (N_{sc}), using two time-windows $[t_{eq}, t_1]$ and $[t_1, t_{full}]$, and over all of simulation trajectories for each system. For 2PHF6, the equilibration time t_{eq} , dividing time t_1 and full time t_{full} are 50, 125 and 200 ns, respectively. For 8PHF6, the three corresponding time parameters are 200, 600 and 1000 ns. As shown in Figure S1 in supporting information (SI), excellent agreements were observed between results obtained using the two time-windows, confirming the convergence of 2PHF6 and 8PHF8 simulation systems. This allows us to discuss the similarity and difference between the FFs and to compare our results with those obtained by other studies. In the rest of text, except mentioned explicitly, all observables are ensemble-averaged data calculated for the time-window of $[t_{eq}, t_{full}]$ using all 100 trajectories for an 2PHF6 system or 20 trajectories for an 8PHF6 system.

The overall structures

To get the first glance of the FF effect on the aggregation of PHF6 peptide, we analysed the key parameters that characterize aggregation including R_g , SASA, number of intermolecular side chain-side chain contacts (N_{sc}) and number of intermolecular hydrogen bonds (N_{hb}) for 2PHF6 and 8PHF6 systems. The time dependence of those reaction coordinate parameters is shown in Figure S2 and their mean values in the time-windows $[t_{eq}, t_{full}]$ of the simulations are presented in Table 1. It is clear that the overall structures of hexapeptide PHF6 are different from one FF to another. This result is in line with the findings in previous studies for IDPs.^{15,17,23} The R_g and SASA values vary with the

same trend, with the largest ones predicted by A99DISP, the smallest ones by OPLS and G54a7, and the values in-between by C36m and A14SB. This indicates that the PHF6 structures predicted by OPLS and G54a7 are more compact than the other three FFs. The compactness of the peptides in 2PHF6 and 8PHF6 systems employed the same FF is further compared using the SASA per peptide. The $SASA_{pp}$ distribution shown in Figure 2 reveals that PHF6 peptides form more compact structures in 8PHF6 than in 2PHF6. This conclusion is further supported by comparing R_g per peptide between the two systems (Figure S3 in SI). The peptide-peptide (P-P) interactions predicted by A99DISP are weak, with N_{sc} of 1.4 and 11.6 in 2PHF6 and 8PHF6, respectively, and with N_{hb} of 0.4 and 5.7 for the two corresponding PHF6 peptide systems. In contrast, the P-P interactions predicted by G54a7 and OPLS are much stronger as their N_{sc} and N_{hb} values are significantly larger (Table 1). As expected, SASA is negatively correlated to N_{sc} for both 2PHF6 and 8PHF6.

Table 1: The ensemble averages of the eight structural parameters characterizing the reaction coordinates including radii of gyration (R_g , in Å), solvent accessible surface area (SASA, in Å²), number of intermolecular side chain-side chain contacts (N_{sc}), number of intermolecular hydrogen bonds (N_{hb}) and the secondary structures (β , helix, turn and coil) of the peptides found in the 2PHF6 and 8PHF6 systems.

System	FFs	R_g	SASA	N_{sc}	N_{hb}	β	helix	turn	coil
2PHF6	A99DISP	11.2	1982.2	1.4	0.4	1.7	0.3	9.5	88.5
	A14SB	8.5	1730	3.7	1.2	8.9	0.7	17.9	72.5
	C36m	9.5	1797	3.2	1.3	13.2	0.2	7.2	79.3
	G54a7	7.2	1553	6.5	3.2	43.8	0.1	9.5	46.6
	OPLS	7.4	1576	5.7	2.0	25.1	0.2	14.2	60.5
8PHF6	A99DISP	21.0	6729	11.6	5.7	10.9	0.5	10.9	77.7
	A14SB	12.0	4881	23.0	10.3	15.4	0.6	19.4	64.6
	C36m	12.8	4889	24.0	16.2	32.6	0.0	5.5	61.8
	G54a7	10.9	4326	21.6	17.7	34.6	0.2	16.7	48.5
	OPLS	11.2	4465	24.6	10.7	13.6	0.3	23.5	62.6

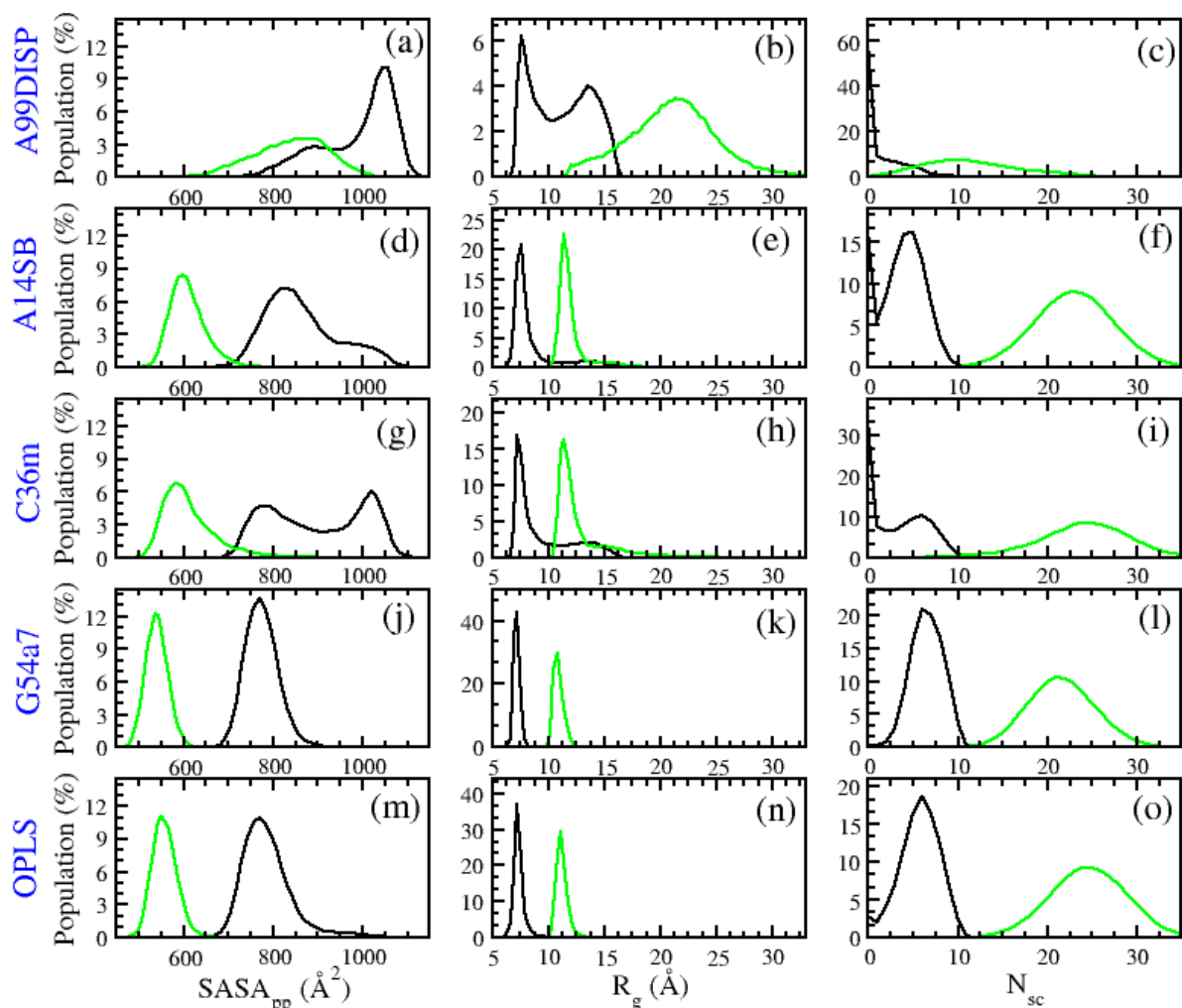


Figure 2: The distributions of surface accessible solvent area per peptide ($SASA_{pp}$), radii of gyration (R_g) and number of intermolecular side chain-side chain contact (N_{sc}) of 2PHF6 (black lines) and 8PHF6 (green lines).

CH- π , Peptide-Peptide and Peptide-Solvent interactions

In a previous study, Sogawa *et al.* observed for the first time that the CH- π interaction between the γ CH of I308 and the aromatic ring of Y310 in PHF6 using NMR spectroscopy.⁵⁴ They found that the CH- π interaction stabilizes paired helical filament (PHF), and it further supports an extended amphipathic structure for molecular self-association during the process of PHF formation of tau protein.⁵⁴ Therefore, we also considered the impact of FFs on the CH- π interaction. A stringent criterion for the formation of CH- π interaction was introduced by Brandl *et. al.* in a previous study.⁵⁵

The criterion includes the distance from the carbon atom of the γ CH to the center-of-mass of the aromatic ring indicated by the point X (CH- π distance), angle defined by the carbon and hydrogen atoms of the γ CH and the center of the aromatic ring ($\angle\gamma\text{C-H-X}$), and the distance between X and the position at which the hydrogen atom is projected vertically onto the ring-plane ($d_{\text{HP-X}}$) (for more detailed description please see Figure 2 in Ref.55). However, considering there are no hydrogen atoms provided in the experimental structure of PHF6 fibrils and the hydrogen atoms in the CH group are not explicitly modelled in the G54a7 FF, in this work we used the CH- π distance and θ , the angle formed between the normal line of the aromatic ring crossing X and the line crossing carbon atom (i.e., the CG2 in the FFs) of γ CH and X as shown in Figure 3a, to monitor the CH- π interaction. In the PHF6 fibrils, the values of CH- π distance and θ are 4.7 Å and 48°, respectively. Therefore, we used the cutoffs, 4.9 Å for CH- π distance and 50° for θ , to determine the formation of a CH- π interaction. In other words, a CH- π interaction is formed, if the CH- π distance is smaller than or equal 4.9 Å and the θ angle is smaller than or equal to 50 degrees. The distributions of the CH- π distance are respectively shown in Figure 3b for 2PHF6 systems and Figure 3c for 8PHF6 systems. It is obvious that all the five FFs have a population peak around 4.7 Å, the CH- π distance found in PHF6 fibrils. Table 2 lists the population of conformations which have the CH- π interaction. It clearly shows that C36m has the largest population of CH- π interaction. In a summary, although the CH- π interaction was observed in all the FFs, C36m provided the highest peak of CH- π distance around 4.7 Å and the largest population of CH- π interaction. Therefore, the CH- π interaction predicted by C36m is stronger than the other FFs.

Beside N_{sc} , N_{hb} and SASA parameters, which directly or indirectly reflect P-P and peptide-solvent (P-S) interactions, we also estimated P-P and P-S interaction energies, $E_{\text{P-P}}$ and $E_{\text{P-S}}$, and the ratio between them, $E_{\text{P-P}}/E_{\text{P-S}}$ (Table 2). For $E_{\text{P-P}}$, A99DISP had the highest values (-46 kJ/mol in 2PHF6 system and -673 kJ/mol in 8PHF6 system), and G54a7 provided the lowest values (-199 kJ/mol in 2PHF6 system and -1428 kJ/mol in 8PHF6 system). Complementary to $E_{\text{P-P}}$, A99DISP

and G54a7 have the lowest and highest values of E_{P-S} , respectively. For both the 2PHF6 and 8PHF6 systems, E_{P-S} of five FFs follows an increasing order as A99DISP $\rightarrow$ C36m $\rightarrow$ OPLS $\rightarrow$ A14SB $\rightarrow$ G54a7, while E_{P-P} follows a decreasing order as A99DISP $\rightarrow$ C36m $\rightarrow$ A14SB $\rightarrow$ OPLS $\rightarrow$ G54a7. E_{P-P}/E_{P-S} follows the same trend as E_{P-P} , with the smallest value found with A99DISP.

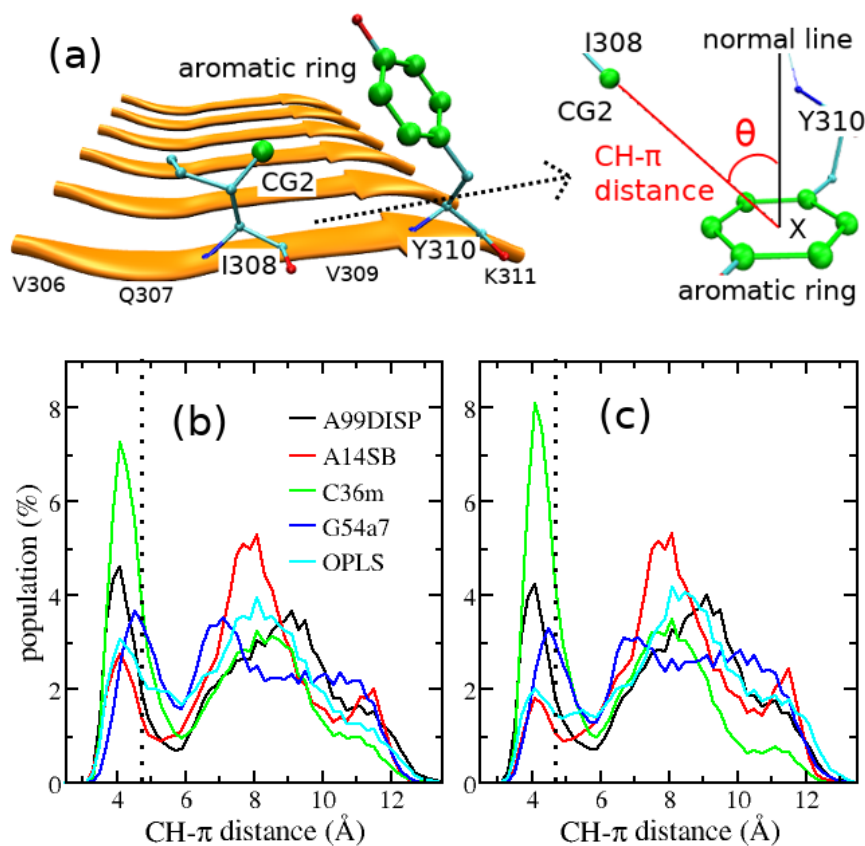


Figure 3: β -sheet structure and CH- π interaction in PHF6 fibrils (a). The distribution of CH- π distance in 2PHF6 (b) and 8PHF6 (c) systems. The CH- π distance in PHF6 fibrils is around 4.7 Å, which is indicated by the dash lines in the Panels b and c. The θ value of in the fibrils is 48°.

Table 2: The population of conformations forming CH- π interaction, P-P interaction energy (E_{P-P}), P-S interaction energy (E_{P-S}), and P-P/P-S interaction ratio in 2PHF6 and 8PHF6 systems.

System	FFs	CH- $\pi \leq 5\text{\AA}$ (%)	E_{P-P} (kJ/mol)	E_{P-S} (kJ/mol)	E_{P-P}/E_{P-S}
2PHF6	A99DISP	29	-46	-1863	0.025
	A14SB	21	-133	-1424	0.093
	C36m	43	-99	-1533	0.065
	G54a7	26	-199	-1228	0.162
	OPLS	29	-194	-1374	0.141
8PHF6	A99DISP	26	-673	-6666	0.101
	A14SB	17	-1293	-4526	0.286
	C36m	46	-1273	-4658	0.273
	G54a7	23	-1428	-3849	0.371
	OPLS	20	-1368	-4567	0.300

Secondary Structures

The evolution of secondary structures of PHF6 peptides in 2PHF6 and 8PHF6 systems during the simulation time is shown in Figure S4, and the values of the secondary structure contents within averaged over the time-window of $[t_{eq}, t_{full}]$ is listed in Table 1. For 2PHF6 systems, the β content predicted by different FFs varied significantly, with the sequence order of A99DISP (1.7%) < A14SB (8.9%) < C36m (13.2%) < OPLS (25.1%) < G54a7 (43.8%). The sequence order of coil content is roughly opposite to that one of the β content. C36m predicted the smallest turn content, while A14SB predicted the most. OPLS also predicted a large percent of the turn content. As expected, the helix structures were rarely detected, and the helix content is less than 1% for all the five FFs. For 8PHF6 systems, the predicted β content of PHF6 peptides is still strongly FF-dependent, but the difference between them is much smaller compared to 2PHF6 systems. The sequence order follows A99DISP (10.9%) < OPLS (13.6%) < A14SB (15.4%) < C36m (32.6%) < G54a7 (34.6%). Interestingly, the β content significantly decreased in 8PHF6 for the OPLS and G54a7 FFs. As to the turn content, the values are roughly the same as those in 2PHF6 for A99DISP, A14SB and C36m, while they are much larger for G54a7 and OPLS. The distribution pattern of coil content in 8PHF6 is similar to that in 2PHF6. Taken together, different FFs predicted significantly

different amount of β content for both 2PHF6 and 8PHF6; when the simulation system is getting larger, more β content was observed with A99DISP, A14SB and C36m, while an opposite trend was found for OPLS and G54a7.

The secondary structure propensities of each amino acid of PHF6 peptide in 2PHF6 and 8PHF6 systems are presented in Figure 4. Among the 6 residues of PHF6, I308 residue shows the highest β propensity in all of the systems. It is understandable due to two following reasons. First, I308 residue is a hydrophobic residue. Second, I308 residue stays in the middle of the short peptide, therefore it has a high probability to participate both in-register and out-of-register β -sheet formations (please see our previous publication¹⁷ for the definition details of the in-register and out-of-register β -sheet formations). V309 residue is similar to I308 residue in terms of residue position and hydrophobicity property. However, the β propensity of V09 is slightly smaller than I308 as V309 is closer to the K311 residue. As a charge residue, K311 is detrimental to parallel β -sheet formation due to unfavorable electrostatic interactions between two positively charged residues. The β propensity of V306 and K311 residues in both 2PHF6 and 8PHF6 systems are zero in A99DISP, A14SB, C36m and OPLS. G54a7 is the only FF predicted non-zero β propensity for V306 in both 2PHF6 and 8PHF6, and K311 in 8PHF6. The helix and turn profiles follow a general trend that the propensities increase from V306 to I308 and decrease from V309 to K311 for all FFs in both 2PHF6 and 8PHF6 systems.

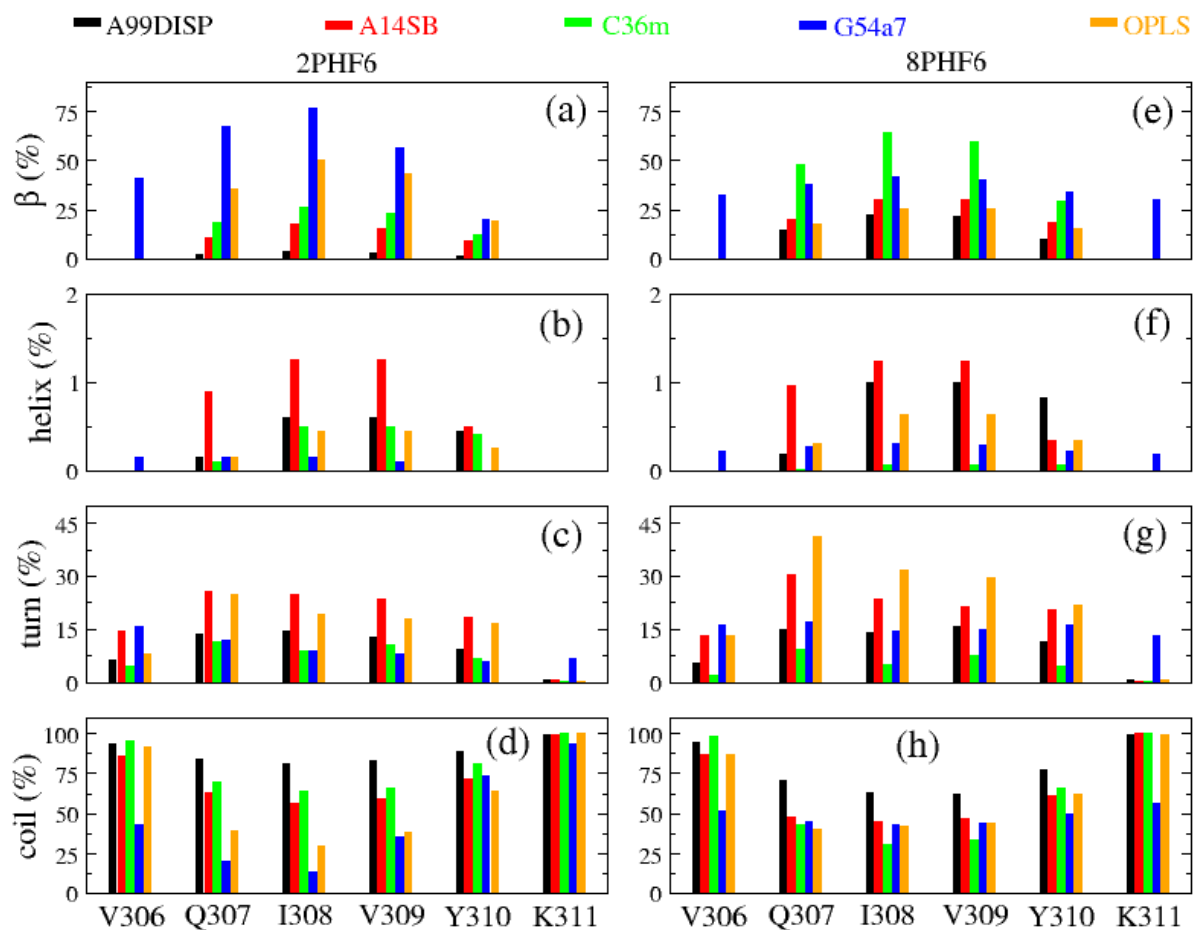


Figure 4: Secondary structural propensities of each amino acid in PHF6 calculated using the MD snapshots sampled for 2PHF6 and 8PHF6 systems.

Dimeric β -sheet Formation

Dimerization is the first step in an amyloid aggregation procedure, and dimeric β -sheet is a building block of amyloid fibrils. To track the dimeric β -sheet formation, we classified a dimeric structure into seven states including disordered state (DOS), weak parallel β -sheet state (PS1), medium parallel β -sheet state (PS2), strong parallel β -sheet state (PS3), weak antiparallel β -sheet state (AS1), medium antiparallel β -sheet state (AS2) and strong antiparallel β -sheet state (AS3) as shown in Figure 1. The populations of each state for 2PHF6 or 8PHF6 systems are explicitly shown in Table 3. For 2PHF6, 95.6% of the equilibrium dimeric structures is disordered in A99DISP, while 86.7% of the structures is in β -sheet formations in G54a7. The antiparallel β -sheet structures (AS=AS1+AS2+AS3) occur more frequently than parallel β -sheet structures (PS=PS1+PS2+PS3) in A14SB, C36m and OPLS FFs. Particularly, in OPLS, AS population is greater than PS population by 35%. Only in G54a7, the population of PS is slightly greater than that of AS. Still, the AS3 population is slightly larger than PS3 with G54a7. In short, PHF6 dimer in 2PHF6 has a strongly tendency to form antiparallel β -sheet structures.

Table 3: The population (%) of the seven states of the dimer found in 2PHF6 and 2PHF8 systems with different force fields. The population of parallel β -sheet structures, PS, is the sum of PS1, PS2 and PS3, and the population of antiparallel β -sheet structures, AS, is the sum of AS1, AS2 and AS3.

System	FFs	DOS	PS	AS	PS1	PS2	PS3	AS1	AS2	AS3
2PHF6	A99DISP	95.6	2.1	2.4	1.1	1.0	0.0	1.8	0.6	0.0
	A14SB	78.3	5.8	15.9	4.1	1.7	0.0	8.5	6.2	1.2
	C36m	70.3	11.6	18.1	4.7	6.9	0.0	7.2	8.6	2.3
	G54a7	13.2	45.9	40.8	9.2	26.1	10.6	9.4	13.7	17.7
	OPLS	43.6	10.1	46.3	4.7	5.4	0.0	19.1	22.8	4.4
8PHF6	A99DISP	74.7	3.8	21.6	2.3	1.5	0.0	11.3	8.0	2.3
	A14SB	63.4	5.8	30.8	3.8	2.0	0.0	17.7	10.5	2.6
	C36m	30.0	46.9	23.1	19.0	27.9	0.0	7.6	10.9	4.6
	G54a7	27.7	32.8	39.5	8.3	20.3	4.2	12.1	16.0	11.4
	OPLS	68.0	9.2	22.9	6.9	2.3	0.0	12.7	8.0	2.2

When a MD system is getting bigger from 2PHF6 to 8PHF6, the population of β -sheet formation dramatically increases with A99DISP, A14SB and C36m FFs, while dramatically decreases with OPLS and G54a7. However, the increase or decrease of β -sheet content follows different patterns for different FFs. The boost of β -sheet is mainly from the significant increase of PS for C36m and AS for the two AMBER FFs. The decrease of β -sheet content mainly occurs to PS states for G54a7 (from 45.9% to 32.8%), while to AS states for OPLS (from 46.3% to 22.9%). Overall, A99DISP, A14SB and OPLS FFs favors the formation of antiparallel β -sheet structures, while C36m favors the formation of parallel β -sheet structures and G54a7 produces a balanced distribution of antiparallel and parallel β -sheet structures.

The evolution of parallel and antiparallel β content along simulation time is shown in Figure 5. Figure 6 demonstrates the transitions between the seven states. The state transitions of individual MD simulations (100 trajectories for 2PHF6 and 20 trajectories for 8PHF6) predicted by five FFs are shown in Figures S5 to S14 in SI. It is observed that the intensity of the state transition frequencies in 8PHF6 system is stronger than that in 2PHF6 system (Figures S5-S14). The state pairs involved in transitions include DOS-PS_i, DOS-AS_i, PS_i-PS_j and AS_i-AS_j ($i, j = 1, 2$ or 3 and $i \neq j$). However, no transition was observed between any types of antiparallel to parallel β -sheet transition or vice versa. Similar to the overall and secondary structures, the pattern of the state transitions strongly depends on the employed FF. First of all, the DOS-PS₁, PS₁-PS₂, DOS-AS₁ and AS₁-AS₂ transitions were observed by all five FFs for both 2PHF6 and 8PHF6 systems. Besides the above common transition types, DOS-PS₂, DOS-AS₂ and AS₂-AS₃ were observed in 8PHF6 system with A99DISP and in both 2PHF6 and 8PHF6 with OPLS, DOS-AS₂ and AS₂-AS₃ were found in both 2PHF6 and 8PHF6 with A14SB and C36m. G54a7 is the sole FF with which all possible transitions for both 2PHF6 and 8PHF6 systems were observed. Examining the state transitions occurred in a single simulation trajectory of 2PHF6, we found that the dominant patterns are DOS-PS_i plus PS_i-PS_j, and DOS-AS_i plus AS_i-AS_j. However, few trajectories had both the DOS-PS_i and DOS-AS_i

transition types. Contrastingly, the last pattern was frequently observed in the MD trajectories of the 8PHF6 system. This result may be explained by the fact that there are 28 possible peptide pairs in 8PHF6 system instead of one in 2PHF6 system.

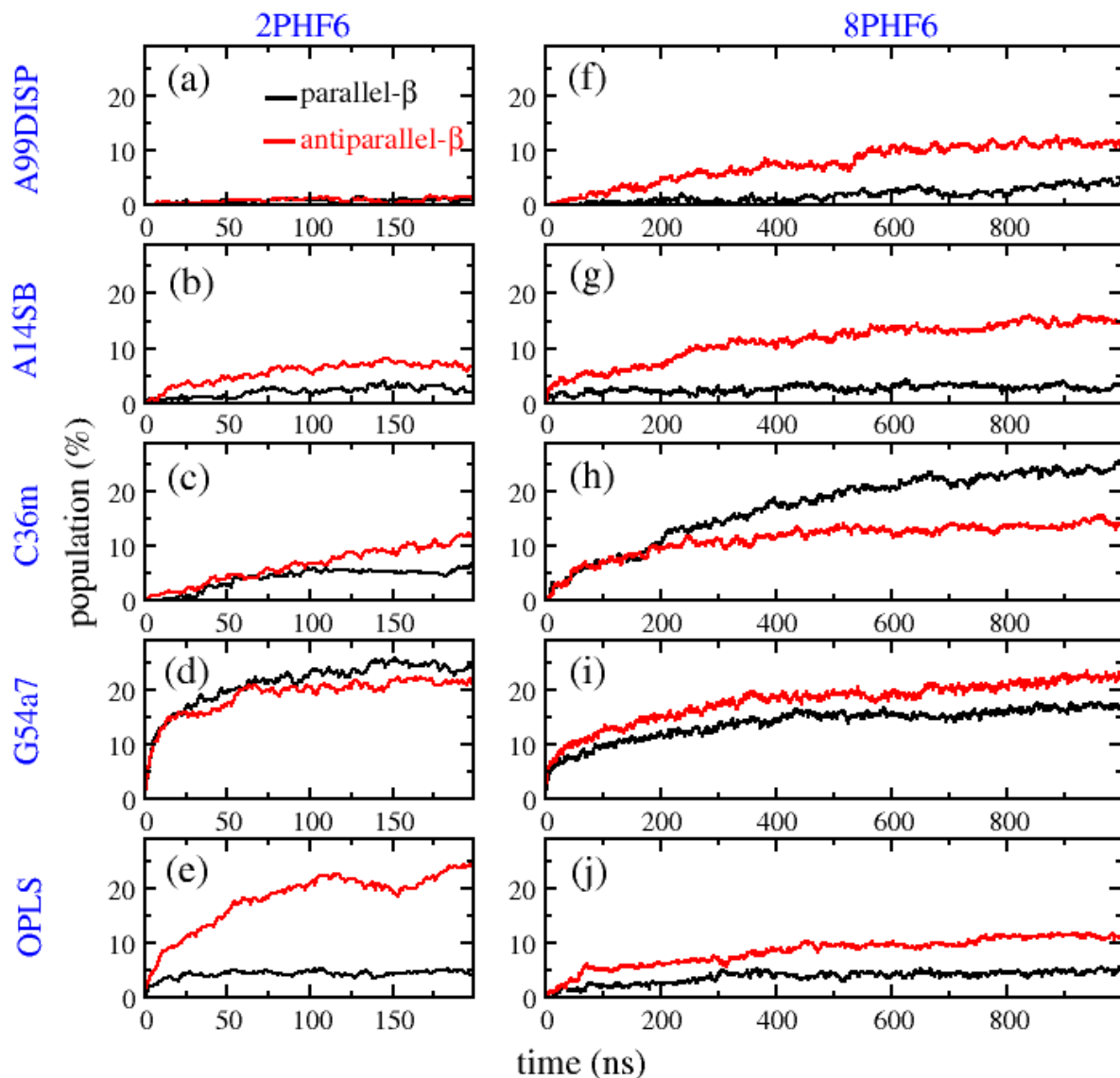


Figure 5: Evolution of the parallel β content (black lines) and antiparallel β content (red lines) of PHF6 dimers sampled using different force field for the 2PHF6 and 8PHF6 systems.

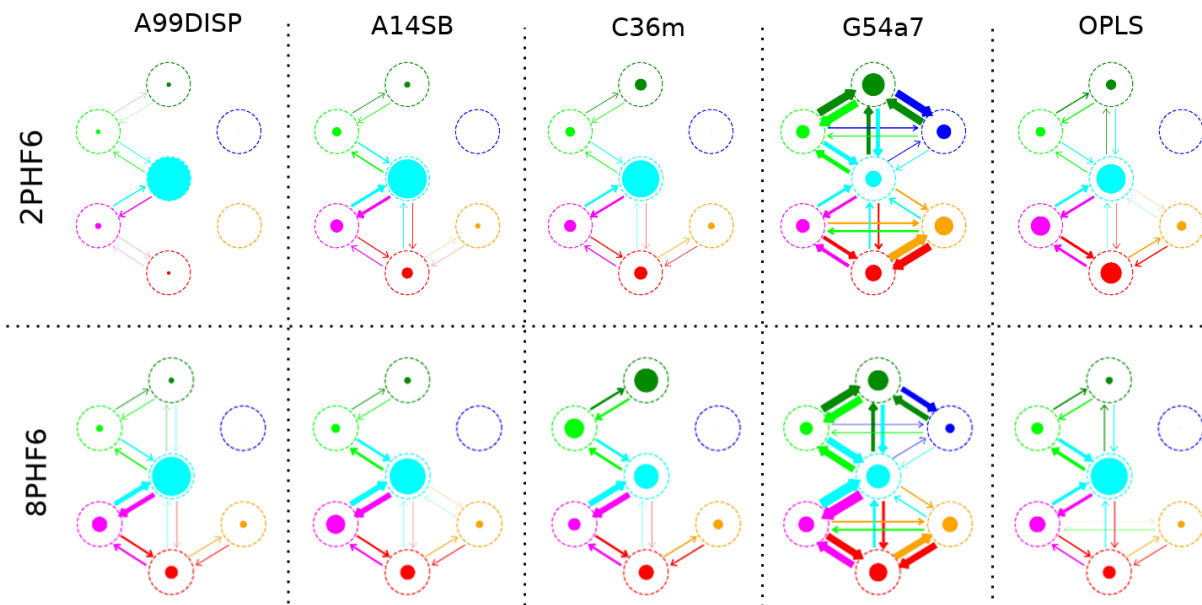


Figure 6: The populations of the seven dimeric states and the state transitions between different types of PHF6 dimers in the 2PHF6 and 8PHF6 systems. The definitions and representation of the seven dimeric states are presented in Figure 1.

Topology of β -sheet Structures in the 8PHF6 System

A β -sheet structure can be characterized by number of strands (monomeric peptides) and the relative directions of its strands. A β -sheet composed of i strands is denoted as $i\beta$ S. A 2β S can be either parallel and antiparallel depends on the directions of the two strands are the same or opposite to each other. For $i\beta$ S with more than two strands, three types are possible, either parallel, antiparallel or mixed of parallel and antiparallel. Here we examined how FFs affect the formation of $i\beta$ S in the 8PHF6 system. The populations of different types of $i\beta$ S, where i takes values of 2,3,4 and 5, are show in Figure 7, and those for $i\beta$ S with i takes values of 6, 7 or 8 are shown in Figure S15. It is observed that the larger a β -sheet structure is, the smaller its population is, and the 2β S population is the largest for all of the five FFs. The largest size of β -sheet Structures observed is 3β S for A99DISP, 4β S for OPLS, 6β S for A14SB and C36m, and 8β S for G54a7. For C36m, the parallel β -sheet structures are dominant over the antiparallel ones for all of the β -sheet sizes, while it is the opposite for other FFs. For all FFs, the mixed type of β -sheet dominates other types for $i\beta$ S ($i>2$).

Again, we have demonstrated that FFs can strongly affect the size and composition of β -sheet structures formed in the 8PHF6 system.

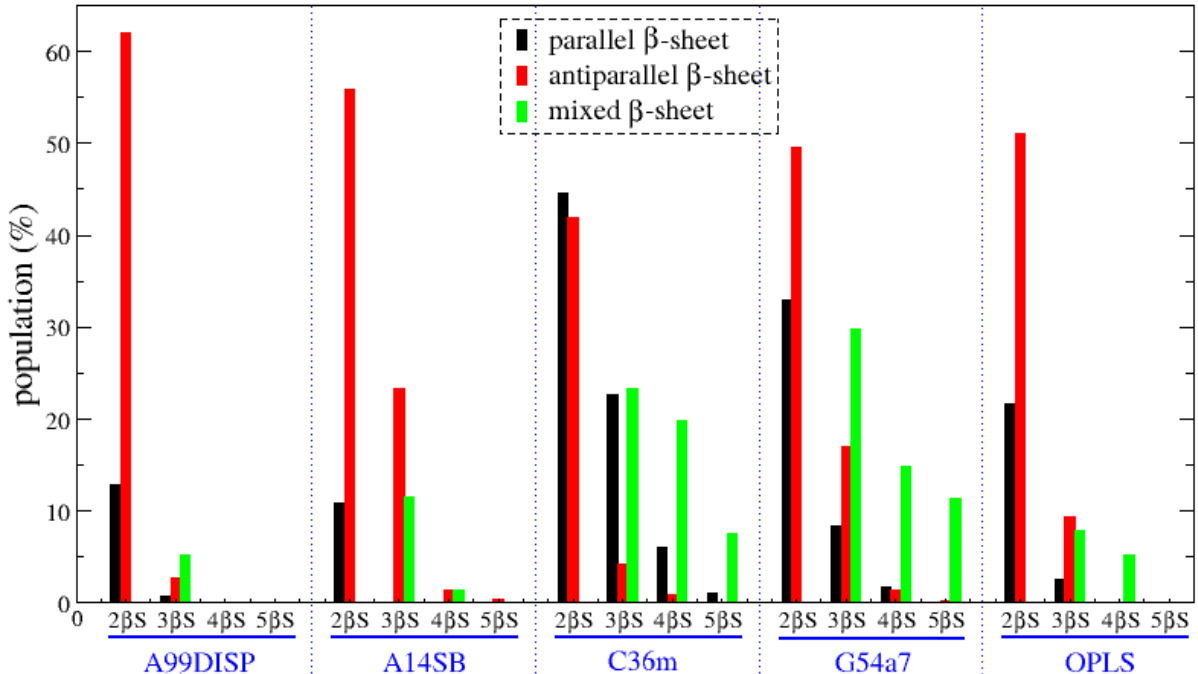


Figure 7: The populations of parallel, antiparallel and mixed type of β -sheet structures with different number of strands found in the 8PHF6 system using different force fields. A β -sheet structure containing i peptides is denoted as $i\beta$ S. For the 8PHF6 system that has 8 monomeric peptides, i takes values of 2 to 8. This figure shows the distributions of three types for $i\beta$ S ($2 \leq i \leq 5$). The distribution of three types for $i\beta$ S ($6 \leq i \leq 8$) is presented in Figure S15.

Representative Structures of 2PHF6 and 8PHF6 in the Different FFs

To find the representative structures of 2PHF6 and 8PHF6 in the FFs, we plot free energy landscape (FEL) by using SASA and P-P interaction energy as the two reaction coordinates. A MD structure will be selected as the representative structure if its reaction coordinated values is the closest to the reaction coordinated values of a local minimum in FEL. As shown in Figure 8, the FELs of 2PHF6 systems has two minima, while there is only one minimum in the FEL of 8PHF6.

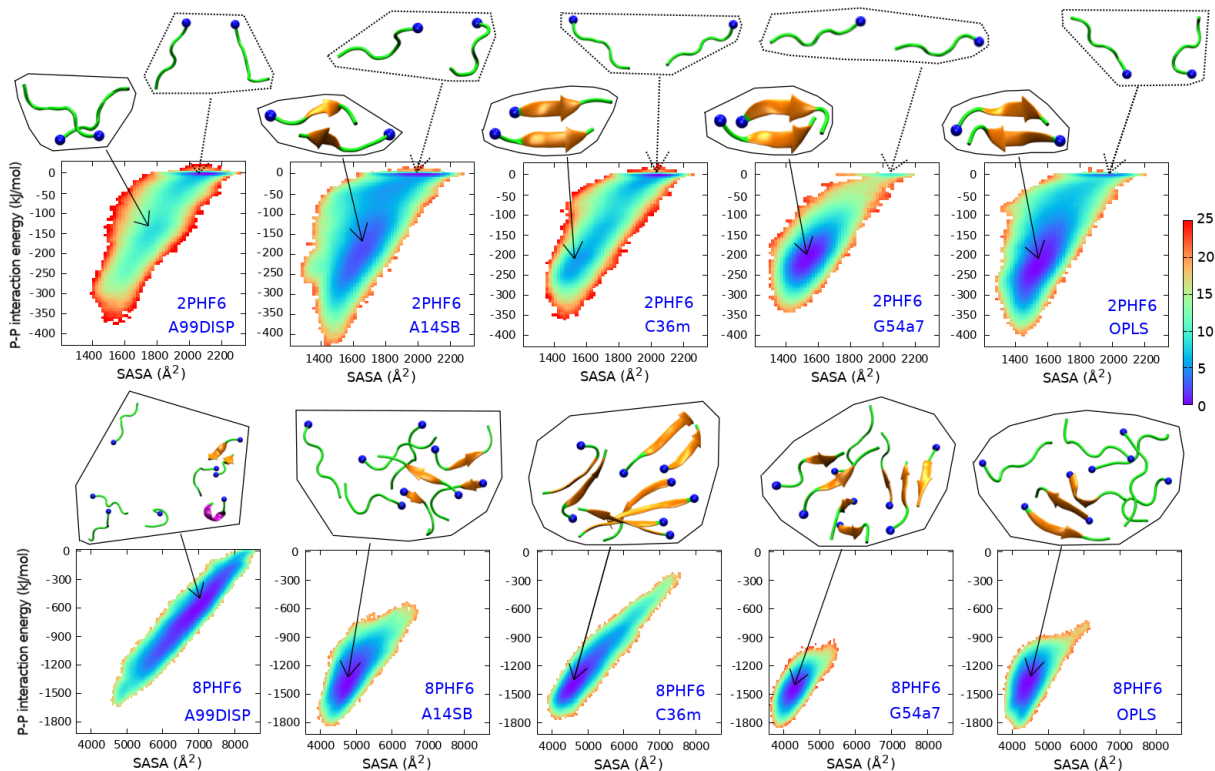


Figure 8: Free energy landscapes of 2PHF6 (upper panels) and 8PHF6 (lower panels) as a function of surface accessible solvent area (SASA) and peptide-peptide (P-P) interaction energy. Representative conformations corresponding to the free energy minima are shown as cartoons.

DISCUSSIONS

Numerous studies have been conducted to examine the applicability of mainstream biomolecular FFs for modeling IDPs. Most of those studies focused on the structures and/or structural dynamics of monomer or dimer of A β peptide (see Ref. 17 and the cited references). Only Strodel *et al.* and us recently compared the performance of all-atom FFs on the amyloid aggregation kinetics using multiple copies of A β ₁₆₋₂₂ peptide in the simulation systems.^{17,18,23} In the works by Strodel *et al.*, the simulation systems contain one or six copies of the wildtype A β ₁₆₋₂₂ and F19L and F19V/F20V mutants.^{18,23} Basing on the experimental findings,^{56,57} Strodel *et al.* first benchmarked six FFs including AMBER03WS, AMBER99SB*ILDN and AMBER99SB*ILDN with LRDI, C22*, G54a7 and OPLS.¹⁸ Afterward, they expanded the FF benchmarks to A99DISP, C36m and C36m

with increased protein-water interactions that had been developed for IDPs.^{13,14} Strodel *et al.* found that G54a7 and OPLS overestimate protein-protein interaction, while protein-water interaction is too dominant in A99DISP, with the result that the amyloid aggregation of A β ₁₆₋₂₂ peptides was inhibited in A99DISP. They concluded that A99DISP, G54a7 and OPLS FFs are not suitable for modeling IDPs. Importantly, they also showed that CHARMM FFs, C36m and C36m with increased protein-water interactions, achieved good performances in modeling amyloid aggregation, and they recommended to use these CHARMM FFs for MD simulation of IDPs. In our previous work for the benchmark of 17 all-atom FFs on the aggregation of A β ₁₆₋₂₂ dimer, we also found that GROMOS FF family and OPLS FF were not good candidates for studying amyloid aggregation.¹⁷ In addition, we pointed out that AMBER99SB-ILDN, AMBER14SB, C22*, C36 and C36m were suitable for the investigation of amyloid aggregation.¹⁷ In this study, we continued to benchmark FF performance on studying the amyloid aggregation kinetics using PHF6, a key hexapeptide of Tau protein. We chose Tau aggregation for FF benchmark study because of the following reasons: (i) similar to A β peptide, Tau is also an IDP, and its aggregation is associated with several neurodegenerative diseases;^{4,5} (ii) although experiments showed both A β ₁₆₋₂₂ and PHF6 peptides aggregated into fibrils with β -sheet structures, the β -sheet structures of PHF6 peptides is parallel, instead of antiparallel ones as A β ₁₆₋₂₂ β -sheet structures;^{40,56} (iii) the all-atom FF benchmark on the aggregation of Tau or Tau fragments has not been emerged yet. We selected A99DISP, A14SB, C36m, G54a7 and OPLS for this benchmark study as this collection covers main biomolecular FF families and many FFs are likely to achieve good performance according to our experience and Strodel's findings on studying A β . To access the applicability of the five selected FFs for studying the aggregation with MD simulations, we designed two MD systems, 2PHF6 and 8PHF6, containing two and eight copies of monomeric PHF6, respectively. For each FF, we carried out one hundred 200-ns independent MD simulations for 2PHF6 system and twenty 1000-ns independent MD simulations for 8PHF6. We then characterized the aggregation of 2PHF6 and 8PHF6 in terms of the overall structures, dimeric β -sheet formation, and topology of β -sheet structures. In the following,

we will give our recommendation on choosing proper FF models to study amyloid aggregation and provide insight on how to improve force fields to achieve better performance on studying amyloid IDPs.

Impact of the force fields on the aggregation of PHF6 peptides.

Our result showed that the aggregation mechanism of PHF6 strongly depended on the employed FF. In the case of A99DISP, the small value of N_{sc} indicates there is a weak interaction between PHF6 peptides, and the peptides tend to expose in the solvent as indicated by large SASA values (Table 1). Additionally, the peptides strongly favor coil structure. Therefore, the β -sheet Structures is hardly formed and unstable. It results that the aggregation of PHF6 peptides is inhibited in A99DISP FF. In the cases of G54a7 and OPLS FFs, an opposite scenario is observed: there is a strong interaction between the peptides and the peptides are less solvent exposed, as suggested by the large values of N_{sc} and small values of SASA. As such, the oligomerization process of PHF6 peptides is accelerated in MD simulations using those two FFs. Indeed, the population of β -sheet structures is very high for G54a7, and the transitions between disordered and weak β -sheet states, weak and medium β -sheet states, and medium and strong β -sheet states frequently occurred during MD simulations. This implies that the barriers between those states are low, and the β -sheet structures as well as fibrils of PHF6 peptides can be easily formed with G54a7 FF. In the case of A14SB and C36m, they achieved a good balance between P-P and P-S interactions, and the aggregation formation speed is between that of A99DISP and G54a7 FFs (Figure S4). FFs not only impact the aggregation kinetics, but also the β -sheet types. Only a small portion of β -sheet structures with size smaller than 4 strands were formed with A99DISP. On the contrary, the largest $i\beta S$ found in 8PHF6 is 4 for A14SB and OPLS, and 5 for C36m and G54a7 (Figure 6 and S15). More importantly, only the C36m FF produced more parallel β -sheet structures than the antiparallel ones for the 8PHF6 system.

The similarity and difference between the aggregation of 2PHF6 and 8PHF6 utilizing the same force field

In an amyloid aggregation, a dimer is formed first, then trimer or tetramer are formed from monomer and dimer or two dimers, next larger oligomeric aggregates are formed by adding a monomer to an oligomer or combining two oligomers together. At a low monomeric peptide concentration, a dimer has a lot of equilibrium time to arrange/rearrange its structure before contacting to a monomer or another oligomer. Therefore, the 2PHF6 system is suitable for investigating the equilibrium of dimer structures. On the other hand, the 8PHF6 system is good for examining the oligomerization kinetics and large oligomeric growing in real situation. In this study, we examined the secondary structures and dimeric β -sheet formation of PHF6 peptides in both 2PHF6 and 8PHF6 systems. We found the similarity between 2PHF6 and 8PHF6 systems employing the same FF in term of the pattern of transition between the seven states (Figure 6). And a significant difference between 2PHF6 and 8PHF6 systems in term of secondary structures and the population of the seven states were also observed for all of the five FFs (Tables 1 and 3). Specifically, the β -sheet content increased from 2PHF6 system to 8PHF6 system in A99DISP, A14SB and C36m, while it decreased from 2PHF6 system to 8PHF6 system in G54a7 and OPLS. It makes the order of β -sheet content from low to high for PHF6 dimer and octamer to be different. This opposite β -content changing can be explained by P-P interaction, P-S interaction, and the higher peptide concentration of 8PHF6 system compared to 2PHF6 system. In general, the increasing of peptide concentration will increase P-P interaction resulting in a faster aggregation and a higher probability of β -sheet formation. However, if P-P interaction is too high over P-S interaction, the aggregated structures will not have enough space to relax/rearrange for the growing β -sheet conformations. As a result, some peptides can form small β -sheet, but some others are stacked at a random structure binding with the formed small β -sheet. As seen in Table 2, P-P interaction energy and P-P/P-S interaction ratio in G54a7 and OPLS are greater than the corresponding ones in A99DISP, A14SB and C36m. Therefore, when it goes from 2PHF6 system to 8PHF6 system, P-P interaction may be too high over

P-S interaction in G54a7 and OPLS, but it is still reasonable in A99DISP, A14SB and C36m. In fact, it is a challenge for an experimental study to capture a dimeric structure and to characterize the oligomerization pathway of an IDP, especially for a short peptide like PHF6. Most experimental studies reported the final products of the amyloid peptide aggregation, i.e., protofibrils or fibrils that consist of many peptides. Thus, the larger number peptides in a simulation system, the better aggregation structure obtained for the comparison with experimental observation. Moreover, large-scale MD simulations of a system consisting of a large number of monomeric peptides can reveal the real aggregation kinetics which may not be obtainable by experiment.

The importance of conventional MD simulation with multiple trajectories in an aggregation investigation

Convergence of conformational sampling is an important requirement to an amyloid aggregation study using MD simulations. The methods such as replica exchange MD (REMD),⁵⁸ simulated tempering⁵⁹ and metadynamics⁶⁰ have been widely applied to enhance conformational sampling in many amyloid aggregation studies. However, these techniques do not allow for tracking the evolution of the aggregation directly, therefore the aggregation pathway and kinetics which are very important features of amyloid aggregation cannot be obtained. On the other hand, these aggregation characterizations can be easily tracked in a conventional MD simulation. The limitation of the conventional MD simulation is that the system may be trapped in one of many local minima, leading to only single pathway and kinetics observed, thus the aggregation picture will not be fully described. To overcome this limitation, multiple long MD simulations starting from different initial structures should be used. To compare the convergence of conformational sampling between the multirun conventional MD (CMD) simulations and REMD simulation, we have also performed REMD simulation for 2PHF6 systems using 16 replicas with the temperature from 305 K to 383 K (the details of REMD simulation are described in SI). As shown in Figure S16, there is a high agreement between the CMD and REMD in term of the distributions of SASA, Rg, and DOS, PS and APS structures. This result further supported that our multirun CMD simulations achieved a good

convergence of conformational sampling. Moreover, if the sampling frequency and time window for the snapshot collection from CMD and REMD simulations are the same, the number of snapshots collected from the multirun CMD is much larger than that collected from REMD at the same simulation temperature. For 2PHF6 system, the number from CMD is a hundred times larger. Therefore, the multirun CMD simulation provided a more detailed picture of amyloid aggregation than REMD simulation. Importantly, our result showed that the pathway of the dimeric β -sheet formation is different from one MD trajectory to another. None of a single trajectory can be applied to depict the whole aggregation picture. The aggregation pathway can only be adequately revealed by analysis of multiple independent MD trajectories as we did for the two MD systems (100 trajectories for 2PHF6 and 20 trajectories for 8PHF6).

C36m is a top candidate for MD simulations of amyloid aggregation

It was revealed by experiment that PHF6 peptides aggregate into fibrils with parallel β -sheet structures.⁴¹ Among the five FFs, only C36m sampled a dominant population of parallel β -sheet structures over that of the antiparallel β -sheet structures in MD simulations of 8PHF6 (46.9% vs 23.1%). Additionally, C36m provided the strongest CH- π interaction in PHF6, which has been observed in a previous NMR study.⁵⁴ In other words, C36m is the best choice for the investigation of PHF6 aggregation by MD simulation. C36m also achieved a good performance in our previous FF benchmark on amyloid aggregation for A β ₁₆₋₂₂ peptide.¹⁷ In addition, Strodel. *et al.* found that C36m obtained the best agreement with experimental observation on the aggregation of A β ₁₆₋₂₂ peptide and its mutations.^{18,23} Taken together, we recommend utilizing C36m to study amyloid aggregation if there is no benchmark created for the amyloid peptide. Of course, other FFs may be applied to a special situation, such as G54a7 may be applied to screen inhibitors of amyloid aggregation, or a specific amyloid peptide, such as A14SB is a good candidate for studying A β .

Different FFs have been employed in previous computational studies to model the aggregation and aggregates structure of Tau protein or its fragments.²⁷⁻³⁷ In the following content, we would like

to briefly review the previous computational studies focusing on PHF6 aggregation. In 2008, using an implicit solvent all-atom minimalistic model and extensive Monte Carlo simulation, Li *et al.* investigated the aggregation of the tau fragment Ac-VQIVYK-NH₂ with considered systems of 12, 24 and 36 chains.²⁷ They showed that the peptides aggregated into β -sheet structures with mixed parallel/antiparallel β -strand organization, which is observed by all the five FFs considered in this benchmark. Interestingly, they found that the fraction of parallel β -sheet structure increases with aggregate size, and they proposed the reorganization of the β -sheets into parallel structure is an important rate-limiting step in the formation of PHF6 fibrils. However, their data also showed that the fraction of anti-parallel β -sheet is larger than that of the parallel β -sheet when the size of aggregates is smaller than 16 peptides. Moreover, although the force field used in Li *et al.*'s work is all-atom and it was stated to go with an implicit solvent, there is not any potential term of water-protein interaction. Therefore, their simulations are likely peptides in a vacuum condition. In 2017, Smit and his co-workers employed a coarse-grained force field with implicit solvent for REMD simulations to investigate the aggregation of PHF6 and PHF6* (VQIINK) peptides.³¹ Their simulations showed that while both fragments form disordered aggregates, only PHF6 is able to form parallel β -sheet fibrils.³¹ At that time, this result was consistent with the experimental evidence, in which the fibril structure of PHF6 was observed, but not for PHF6*. However, a year later, the fibril structure of PHF6* had been discovered in an *in vitro* study.⁶¹ In 2019, Liu *et al.* used AMBER99SB force field⁶² and TIP3P water for MD simulations to investigate the aggregation of PHF6 peptides.³³ They found that PHF6 can spontaneously aggregate to form multimers enriched with β -sheet structure, and the β -sheets prefer to exist in a parallel way. Their result indicates AMBER99SB may be also a good force field candidate for MD simulations of PHF6 aggregation. Note that Liu *et al.* performed only a single MD run, leading to their conformational sampling is not good as ours as we performed multiple MD runs. Additionally, AMBER99SB is not good as C36m in modeling the aggregation of A β peptides.^{17,18,23} Overall, C36 is better than AMBER99SB in simulation study of amyloid aggregation. Recently, Arya *et al.* studied impact of terminal capping

on PHF6 aggregation in a joint experimental/MD simulation Study.³⁵ In their simulations, OPLS and TIP3P were applied for protein and water, respectively. Although the fractions of parallel and anti-parallel β -strand was not estimated, the representative structures (Figure 5 in Ref.35) from the MD simulations and inter-residue hydrogen bond map (Figure 6 in Ref.35) suggested that the β -sheet formed by no-capping PHF6 is antiparallel and the β -sheet formed by Ac-VQIVYK-NH₂ is mixed parallel/antiparallel. Notably, in the representative structures (Figure 5 in Ref.35), the maximum size of β -sheets was four peptides, much smaller than 25 peptides in their simulation system, and the monomers and small β sheets crossed to others. It reflected the overestimation of protein-protein interaction in the simulations.

Opinions on developing force fields for studying for amyloid aggregation

Recent FF developments mostly focused on the correction for the torsional angles of backbone and side chain, and some other intramolecular terms to improve the prediction of protein secondary structural propensities as well as the representation of protein folding conformations. A99DISP was introduced on the basis of AMBER99SB-ILDN²² with torsion optimization targeting (AAQAA)₃ fraction helicity and polyalanine scalar couplings.¹⁴ A14SB is a revision of the AMBER99SB⁶² with modified torsions of backbone and side chains, producing better measurements comparing with experiments.¹¹ C36m was developed from C36 with a refined backbone CMAP potential derived from reweighting calculation and a better description of specific salt bridge interactions. C36m improves the conformational sampling for intrinsically disordered peptides and proteins.¹³ OPLS took the bond stretching and angle bending terms mostly from AMBER94⁶³ except for alkanes, for which the parameters were taken from CHARMM22.⁸ All torsional and nonbonded parameters of OPLS were reoptimized to reproduce conformational energetics, gas-phase intermolecular energetics, and thermodynamic properties of pure liquids. G54a7 was based on G53a6 with new ϕ/ψ

torsional angle terms and a modified N-H, C-O repulsive term to correct the G53A6 helical propensities. Those FF developments have been partly successful with respect to their targets when they reproduced more accurately protein structures observed in experiments. However, it is not enough for the case of protein aggregation, particularly amyloid aggregation, in which intermolecular interactions such as the P-P and P-S interactions and their ratio also play an important role, as E_{P-P} , E_{P-S} and E_{P-P}/E_{P-S} directly influence the aggregation speed and growth of β -sheet conformations. As discussed above, both the dominance of P-S interaction to P-P interaction in A99DISP and the overestimation of P-P interaction in G54a7 and OPLS can lead to unexpected result in amyloid aggregation simulation. Therefore, P-P/P-S interaction balance should be considered in FF developments.

Besides the reasonable aggregation rate and growth of β -sheet structures, the proper aggregated structures formed during MD simulation is also an important factor to benchmark a force field for studying amyloid aggregations. For the case of PHF6 peptide simulation, parallel β -sheet structure should be the dominant population as suggested by experiment. Among the five FFs benchmarked in this work, the dominant of parallel β -sheet structure over anti-parallel β -sheet structure was only happened in C36m. Because the parallel/antiparallel β -sheet formation is mostly controlled by P-P interaction, we further analyzed P-P interaction of dimeric β -sheet structures in term of electrostatic and Lennard-Jones interaction energies to figure out the reason of the formation preference. The electrostatic, Lennard-Jones and total P-P interaction energies of dimeric β -sheet structures are shown in Table 4. As seen, P-P interaction energy of the parallel β -sheet structure is lower than that one of antiparallel β -sheet structure for G54a7 and C36m, while the trend is opposite for the other three FFs. This suggests that C36m and G54a7 have higher tendency to form β -sheet structures than the other three force fields. This is consistent with aforementioned observation on the population analysis of β -sheet structures for the five FFs. Further examination on the energy components of the interaction energy, it was found that the electrostatic interaction energy is lower than that of the

Lennard-Jones interaction energy for G54a7 and especially C36m (Table 4). This result indicates that a good balance of the electrostatic and van der Waals interactions should be taken into account in future FF developments for studying amyloid peptides. In fact, most today's main stream force fields inherited partial atomic charges from old FFs developed a long time ago. For example, A14SB still used the partial atomic charges of A94. Recently, our group developed a new charge model called ABCG2 for the development of a next generation general AMBER force field-GAFF3. ABCG2 was developed to optimize a set of bond charge correction (BCC) parameters to reproduce the experimental solvation free energies of small molecules.⁶⁴ Comparing to the original BCC parameters, ABCG2 significantly improved the performance of GAFF2 in solvation free energy calculation for diverse solutes in various organic solvents across a range of different dielectric constants. The new ABCG2 charge model may be applied to develop a special force field for studying IDPs including amyloid peptides and proteins. Finally, we suggest FF developments for IDPs should include amyloid aggregation benchmark for some well-known amyloid peptides and multiple monomers should be included in a simulation system.

Table 4: Peptide-peptide interaction energies (kJ/mol) in dimeric β -sheet structures in the different force fields. The data includes total, electrostatic (elec) and Lennard-John (lj) energies. For each structural type, the data is averaged from the window time $[t_{eq}, t_{full}]$, all trajectories and both 2PHF6 and 8PHF6.

FFs	PS				APS				PS-APS
	Total	elec	lj	elec-lj	total	elec	lj	elec-lj	
A99DISP	-224	-107	-117	10	-236	-115	-121	6	12
A14SB	-217	-100	-117	17	-227	-111	-116	5	10
C36m	-225	-118	-107	-11	-206	-106	-100	-6	-19
G54a7	-199	-105	-94	-6	-188	-97	-91	-7	-11
OPLS	-214	-95	-119	24	-225	-102	-123	21	11

CONCLUSION

We have investigated how well five popular biomolecular FFs perform in studying the oligomerization process of PHF6 peptide, a key segment of Tau protein. Our result showed that P-P interaction is overestimated by G54a7 and OPLS, leading to the formation of a large portion of antiparallel β -sheet structures, while this interaction is underestimated by A99DIPS, making the FF inhibits PHF6 aggregation. The solvent exposure and P-P interactions were well-balanced by A14SB and C36m FFs, so that they are good candidates to characterize the aggregation kinetics of PHF6 peptide. However, only C36m can sample a dominant population of parallel β -sheet structures over antiparallel β -sheet structure, in agreement with experimental findings. In addition, C36m also provided the highest frequency of CH- π interaction in PHF6 peptide, that has been experimentally observed. Taken the previous studies on A β ₁₆₋₂₂ peptide into consideration, C36m is the only FF suitable to study both A β and Tau aggregation.

ASSOCIATED CONTENT

Supporting information includes the distribution of SASA_{pp}, R_g , and N_{sc} in two time-windows (Figure S1); the evolution of R_g , SASA, N_{sc} , N_{hb} (Figure S2); the distribution of R_g per peptide (Figure S3); the evolution of secondary structures (Figure S4) of 2PHF6 and 8PHF6 along the MD simulation time; the population of the seven states and the state transitions of PHF6 dimer in the 2PHF6 system (Figures S5-S9) and in 8PHF6 system (Figures S10-S14) with different force fields; the populations of parallel, antiparallel and mixed type of β -sheet structures with varied sizes formed in the 8PHF6 system with different force fields (Figure S15); the comparison between conventional MD simulation with multiple trajectories and REMD simulation for 2PHF6 systems (Figure S16).

AUTHOR INFORMATION

Corresponding Authors

Email: junmei.wang@pitt.edu

Author Contributions

V.H.M and J.W designed the project. V.H.M performed the simulations and analyzed data. V.H.M, X.H, J.G and J.W discussed and wrote the paper.

Notes

The authors declare that they have no competing financial interests

ACKNOWLEDGMENTS

This work was supported by the following funds from the National Institutes of Health (NIH) and the National Science Foundation (NSF): NIH R01GM079383, NIH P30DA035778, and NSF 1955260. The authors also thank the computing resources provided by the Center for Research Computing (CRC) at University of Pittsburgh.

References

- (1) Pearson, H. A.; Peers, C. Physiological roles for amyloid β peptides. *J. Physiol.* **2006**, 575, 5-10.
- (2) Giuffrida, M. L.; Caraci, F.; Pignataro, B.; Cataldo, S.; Bona, P. D.; Bruno, V.; Molinaro, G.; Pappalardo, G.; Messina, A.; Palmigiano, A.; Garozzo, D.; Nicoletti, F.; Rizzarelli, E.; Copani, A. β -Amyloid Monomers Are Neuroprotective. *J. Neurosci.* **2009**, 29, 10582–10587.
- (3) Kadavath, H.; Hofele, R. V.; Biernat, J.; Kumar, S.; Tepper, K.; Urlaub, H.; Mandelkow, E.; Zweckstetter, M. Tau stabilizes microtubules by binding at the interface between tubulin heterodimers. *Proc. Natl. Acad. Sci.* **2015**, 112, 7501–7506.
- (4) Shafiei, S. S.; Guerrero-Munoz, M. J.; Castillo-Carranza, D. L. Tau Oligomers: Cytotoxicity, Propagation, and Mitochondrial Damage. *Front Aging Neurosci.* **2017**; 9: 83 2017, 9, 83.

- (5) Cline, E. N.; Maíra Assunção Bicca, K. L. V.; Klein, W. L. The Amyloid-Oligomer Hypothesis: Beginning of the Third Decade. *J. Alzheimers Dis.* **2018**, *64*, S567–S610.
- (6) Bloom, G. S. Amyloid- β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* **2014**, *71*, 505–8.
- (7) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27–68.
- (8) Jensen, K. P.; Jorgensen, W. L. Halide, Ammonium, and Alkali Metal Ion Parameters for Modeling Aqueous Solutions. *J. Chem. Theory Comput.* **2006**, *2*, 1499–1509.
- (9) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Jr., R. A. D.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B.* **2010**, *114*, 2549–64
- (10) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Winger, S. R.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–56.
- (11) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (12) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (13) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; Jr, A. D. M. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods.* **2017**, *14*, 71–73.
- (14) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci.* **2018**, *115*, E4758–E4766.
- (15) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.
- (16) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154.
- (17) Man, V. H.; He, X.; Derreumaux, P.; Ji, B.; Xie, X.-Q.; Nguyen, P. H.; Wang, J. Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of A β _{16–22} Dimer. *J. Chem. Theory Comput.* **2019**, *15*, 1440–1452.

- (18) Ismail, M. C.-P. A. E.; Strodel, B. On the Applicability of Force Fields to Study the Aggregation of Amyloidogenic Peptides Using Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 6063–6075.
- (19) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–L49.
- (20) Kollman, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc Chem Res.* **1996**, *29*, 461–469.
- (21) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Alexander D. MacKerell, J. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (22) Lindorff-Larsen, K. and Piana, S. and Palmo, K. and Maragakis, P. and Klepeis, J.L. and Dror, R.O. and Shaw, D.E., Improved side-chain torsion potentials for the AMBER99SB protein force field. *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 1950.
- (23) Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*(12), 6462–6475.
- (24) Nielsen, J.; Bjerring, M.; Jeppesen, M.; Pedersen, R.; Pedersen, J.; Hein, K.; Vosegaard, T.; Skrydstrup, T.; Otzen, D.; Nielsen, N. Unique Identification of Supramolecular Structures in Amyloid Fibrils by Solid State NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 2118–21.
- (25) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riek, C.; Grothe, R.; Eisenberg, D. Structure of the cross- β spine of amyloid-like fibrils. *Nature* **2005**, *435*, 773–778.
- (26) Goedert, M.; Eisenberg, D. S.; Crowther, R. A. Propagation of Tau Aggregates and Neurodegeneration. *Annu. Rev. Neurosci.* **2017**, *40*, 189–210.
- (27) Li, D.-W.; Mohanty, S.; Irbäck, A.; Huo, S. Formation and Growth of Oligomers: A Monte Carlo Study of an Amyloid Tau Fragment. *PLoS Comput. Biol.* **2008**, *12*, e1000238.
- (28) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O’Hern, C. S.; Rhoades, E. The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophys. J.* **2012**, *103*, 1940–1949.

- (29) Luo, Y.; Ma, B.; Nussinov, R.; Wei, G. Structural Insight into Tau Protein's Paradox of Intrinsically Disordered Behavior, Self-Acetylation Activity, and Aggregation. *J. Phys. Chem. Lett.* **2014**, *5*, 3026–3031.
- (30) Ganguly, P.; Do, T. D.; Larini, L.; LaPointe, N. E.; Sercel, A. J.; Shade, M. F.; Feinstein, S. C.; Bowers, M. T.; Shea, J.-E. Tau Assembly: The Dominant Role of PHF6 (VQIVYK) in Microtubule Binding Region Repeat R3. *J. Phys. Chem. B.* **2015**, *119*, 4582–4593.
- (31) Smit, F. X.; Luiken, J. A.; Bolhuis, P. G. Primary Fibril Nucleation of Aggregation Prone Tau Fragments PHF6 and PHF6*. *J. Phys. Chem. B.* **2017**, *121*, 3250–3261.
- (32) Eschmann, N. A.; Georgieva, E. R.; Ganguly, P.; Borbat, P. P.; Rappaport, M. D.; Akdogan, Y.; Freed, J. H.; Shea, J.-E.; Han, S. Signature of an aggregation-prone conformation of tau. *Sci. Rep.* **2017**, *7*, 44739.
- (33) Liu, H.; Zhong, H.; Liu, X.; Zhou, S.; Tan, S.; Liu, H.; Yao, X. 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.* **2019**, *10*, 4810–4823.
- (34) Chen, D.; Drombosky, K. W.; Hou, Z.; Sari, L.; Kashmer, O. M.; Ryder, B. D.; Perez, V. A.; Woodard, D. R.; Lin, M. M.; Diamond, M. I.; Joachimiak, L. A. Tau local structure shields an amyloid-forming motif and controls aggregation propensity. *Nat Commun.* **2019**, *10*, 2493.
- (35) Arya, S.; Ganguly, P.; Arsiccio, A.; Claud, S. L.; Trapp, B.; Schonfeld, G. E.; Liu, X.; Cantrell, K. L.; Shea, J.-E.; Bowers, M. T. Terminal Capping of an Amyloidogenic Tau Fragment Modulates Its Fibrillation Propensity. *J. Phys. Chem. B.* **2020**, *124*, 8772–8783.
- (36) Derreumaux, P.; Man, V. H.; Wang, J.; Nguyen, P. H. Tau R3-R4 Domain Dimer of the Wild Type and Phosphorylated Ser356 Sequences. I. In Solution by Atomistic Simulations. *J. Phys. Chem. B.* **2020**, *124*, 2975–2983.
- (37) Yang, J.; Agnihotri, M. V.; Huseby, C. J.; Kuret, J.; Singer, S. J. A theoretical study of polymorphism in VQIVYK fibrils. *Biophys. J.* **2021**, *120*, 1396–1416.
- (38) Spillantini, M. G.; Goedert, M. Tau Pathology and Neurodegeneration. *Lancet Neurol.* **2013**, *12*, 609–22.
- (39) Stohr, J. et al. A 31-residue Peptide induces Aggregation of tau's Microtubule-Binding Region in Cells. *Nat. Chem.* **2017**, *9*, 874–881.
- (40) Mohamed, N.-V.; Herrou, T.; Plouffe, V.; Piperno, N.; Leclerc, N. Spreading of Tau Pathology in Alzheimer's Disease by Cell-to-Cell Transmission. *Eur J. Neurosci.* **2013**, *37*, 1939–1948.

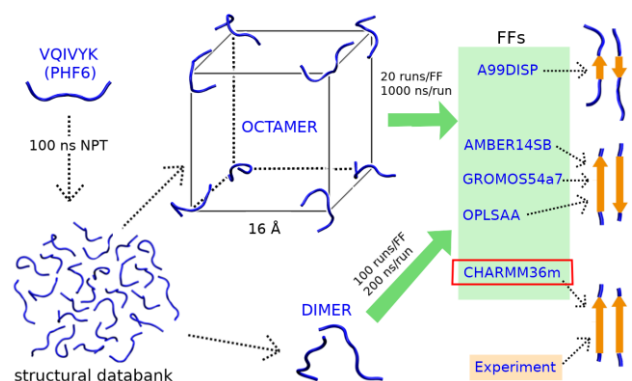
- (41) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riek, C.; Eisenberg, D. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature*. **2007**, *447*, 453–457.
- (42) Ma, B.; Nussinov, R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol.* **2006**, *10*, 445–452.
- (43) Chebaro, Y.; Mousseau, N.; Derreumaux, P. Structures and Thermodynamics of Alzheimer's Amyloid- β A β_{16-22} Monomer and Dimer by Replica Exchange Molecular Dynamics Simulations: Implication for Full-Length A β Fibrillation. *J Phys Chem B*. **2009**, *113*, 7668–7675.
- (44) Case, D. A. et al. "AMBER 19"; University of California: San Francisco, **2019**.
- (45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *779*, 926–935.
- (46) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (47) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (48) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (49) Hockney, R. W.; Goel, S. P.; Eastwood, J. Quit high resolution computer models of plasma. *J. Comp. Phys.* **1974**, *14*, 148–158.
- (50) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **1997**, *18*, 1463–1472.
- (51) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (52) Frishman, D.; Argos, P. Knowledge-based secondary structure assignment. *Proteins: Structure, Function, and Genetics*. **1995**, *23*, 566 – 579.
- (53) Heinig, M.; Frishman, D. STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **2004**, *32*, W500–2.

- (54) Sogawa, K.; Minoura, K.; In, Y.; Ishida, T.; Taniguchi, T.; Tomoo, K. CH- π interaction in VQIVYK sequence elucidated by NMR spectroscopy is essential for PHF formation of tau. *Biopolymers*. **2014**, *102*, 288–95.
- (55) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. C-H $\cdots\pi$ -Interactions in Proteins. *J. Mol. Biol.* **2001**, *307*, 357–377.
- (56) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid Fibril Formation by A β_{16-22} , a Seven-Residue Fragment of the Alzheimer's β -Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry*. **2000**, *39*, 13748–13759.
- (57) Senguen, F. T.; Doran, T. M.; Anderson, E. A.; Nilsson, B. L. Clarifying the influence of core amino acid hydrophobicity, secondary structure propensity, and molecular volume on amyloid- β 16-22 self-assembly. *Mol. Biosyst.* **2011**, *7*, 497–510.
- (58) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141.
- (59) Zhang, T.; Nguyen, P. H.; Nasica-Labouze, J.; Mu, Y.; Derreumaux, P. Folding Atomistic Proteins in Explicit Solvent Using Simulated Tempering. *J. Phys. Chem. B*. **2015**, *119*, 6941–6951.
- (60) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.
- (61) Seidler, P. M.; Boyer, D. R.; Rodriguez, J. A.; Sawaya, M. R.; Cascio, D.; Murray, K.; Gonen, T.; Eisenberg, D. S. Structure-based inhibitors of tau aggregation. *Nat. Chem.* **2018**, *10*, 170–176.
- (62) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. **2006**, *65*, 712–725.
- (63) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (64) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153*, 114502.

For Table of Contents Use Only

Effects of All-atom Force Fields on Amyloid Peptide Assembly: The Case of PHF6 Peptide of Tau Protein

Viet Hoang Man, Xibing He, Jie Gao, and Junmei Wang



Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of PHF6 Peptide of Tau Protein

(SUPPORTING INFORMATION)

Viet Hoang Man,¹ Xibing He,¹ Jie Gao,² and Junmei Wang^{1*}

¹ *Department of Pharmaceutical Sciences and Computational Chemical Genomics
Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261, USA.*

² *Department of Neuroscience, The Ohio State University Wexner Medical Center,
Columbus, OH 43210, USA.*

E-mail: junmei.wang@pitt.edu

Replica exchange simulation

For each 2PHF6 system, the replica exchange simulation (REMD) was performed using 16 replicas, which covers a temperature range from 305 K to 383 K. The initial structures of REMD were the initial structures of the first 16 of 100 MD trajectories. The temperatures of replicas were determined by using the method proposed by Patriksson and van der Spoel (Phys. Chem. Chem. Phys., 2008, 10, 2073-2077). The requested acceptance ratio was around 20 %. Exchanges between replicas were attempted every 2 ps, large enough compared to the coupling time of the heat bath. Each replica was run for 200 ns, and the data were collected every 10 ps. The first 20 ns of all trajectories were excluded for analysis. All other simulation parameters were set the same as those ones of the conventional MD simulation (see the main tex).

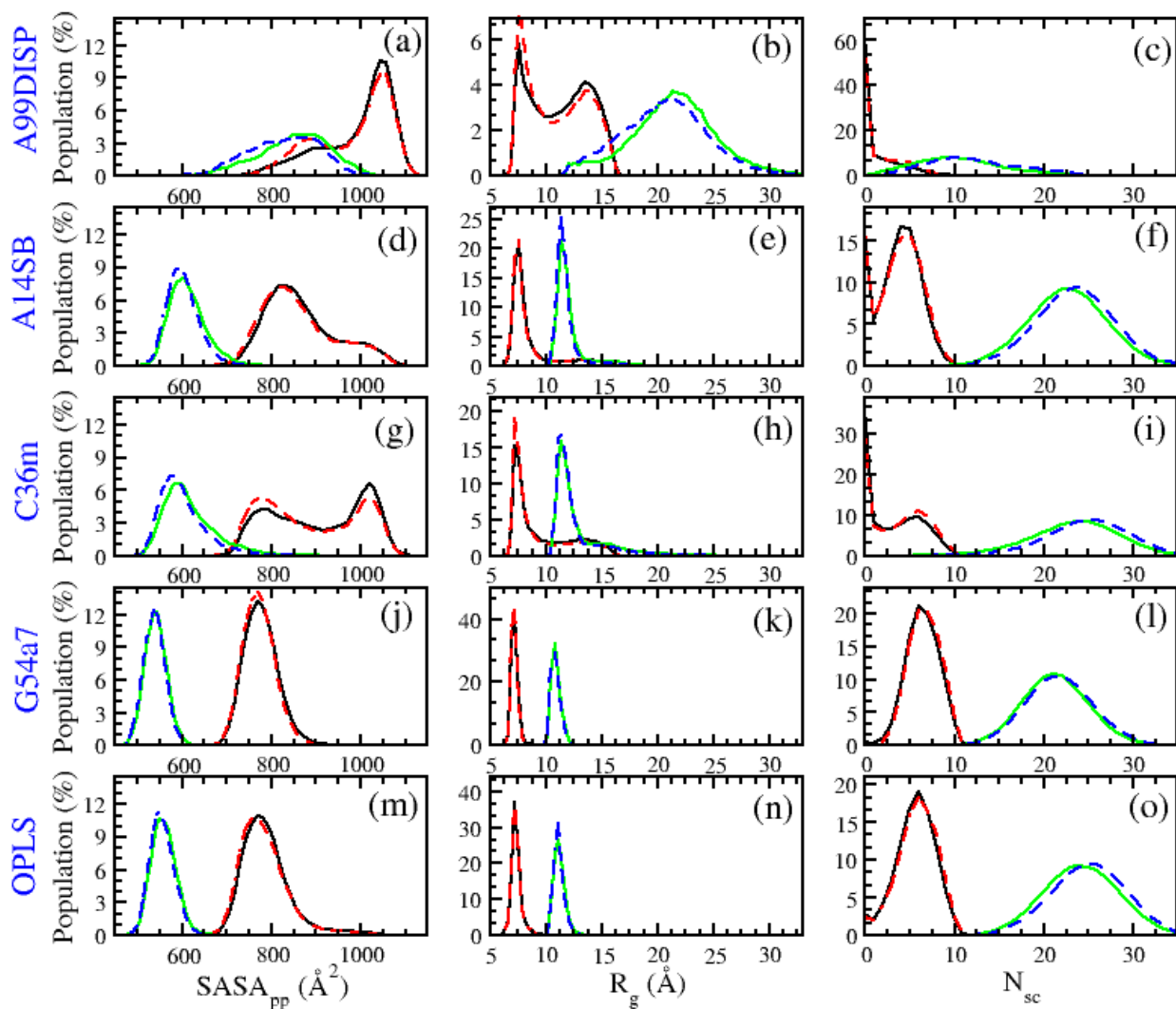


Figure S1: The distribution of surface accessible solvent area per peptide ($SASA_{pp}$), radii of gyration (R_g) and number of intermolecular side chain-side chain contact (N_{sc}) of 2PHF6 and 8PHF6 in corresponded systems. The data of 2PHF6 systems were calculated from the simulation periods, from 50 ns to 125 ns (solid black lines) and from 125 ns to 200 ns (dash red lines), of 100 trajectories. The data of 8PHF6 systems were calculated from the simulation periods, from 200 ns to 600 ns (solid green lines) and from 600 ns to 1000 ns (dash blue lines), of 20 trajectories.

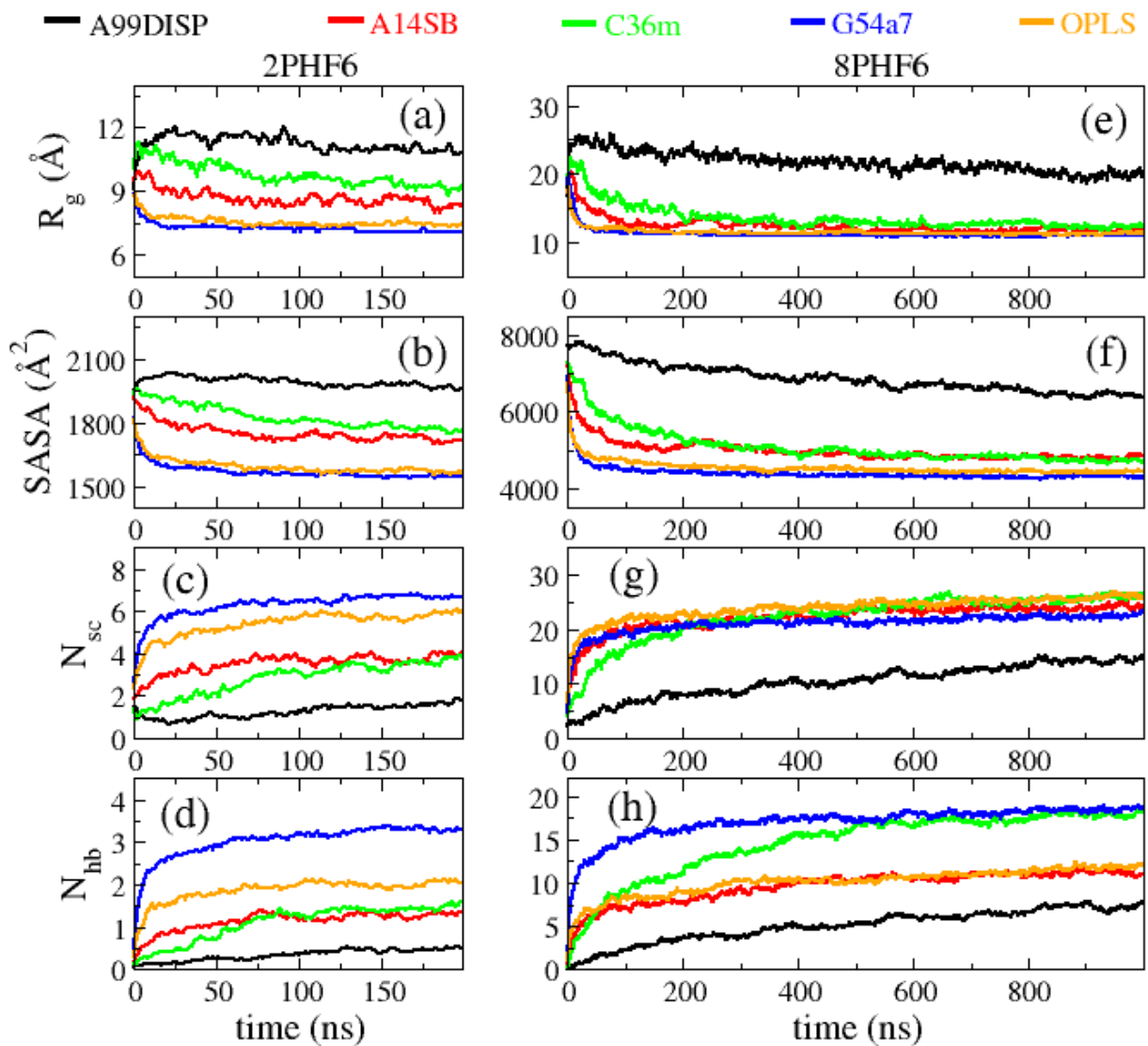


Figure S2: Evolution of the numbers of intermolecular side chain-side chain contacts (N_{sc}), the numbers of intermolecular of main chain hydrogen bonds (N_{hb}), surface accessible solvent area (SASA) and radii of gyration (R_g) of 2PHF6 and 8PHF6 along the MD simulation time using different force fields.

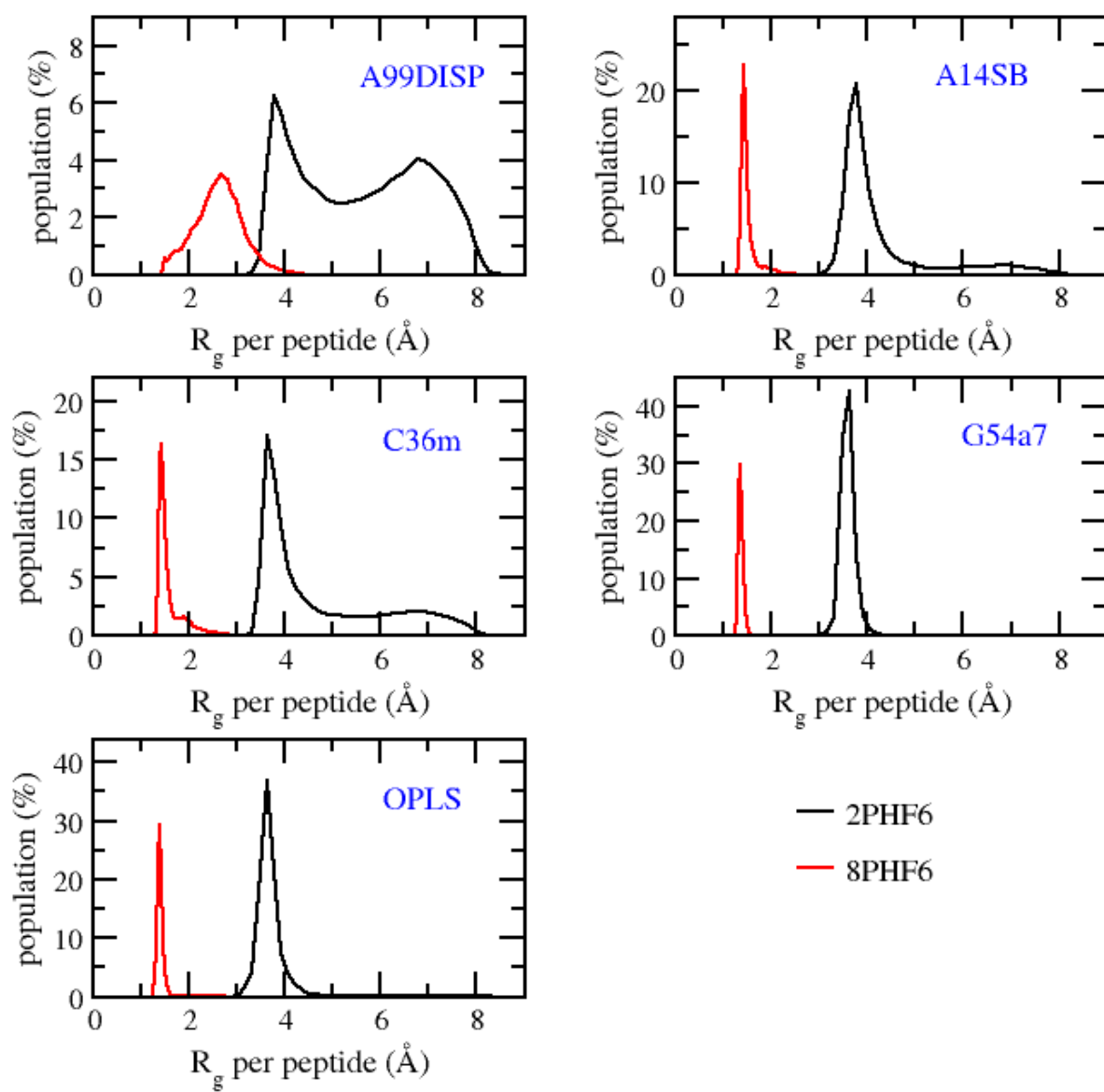


Figure S3: The distribution of gyration radius per peptide of 2PHF6 and 8PHF6 in corresponded systems.

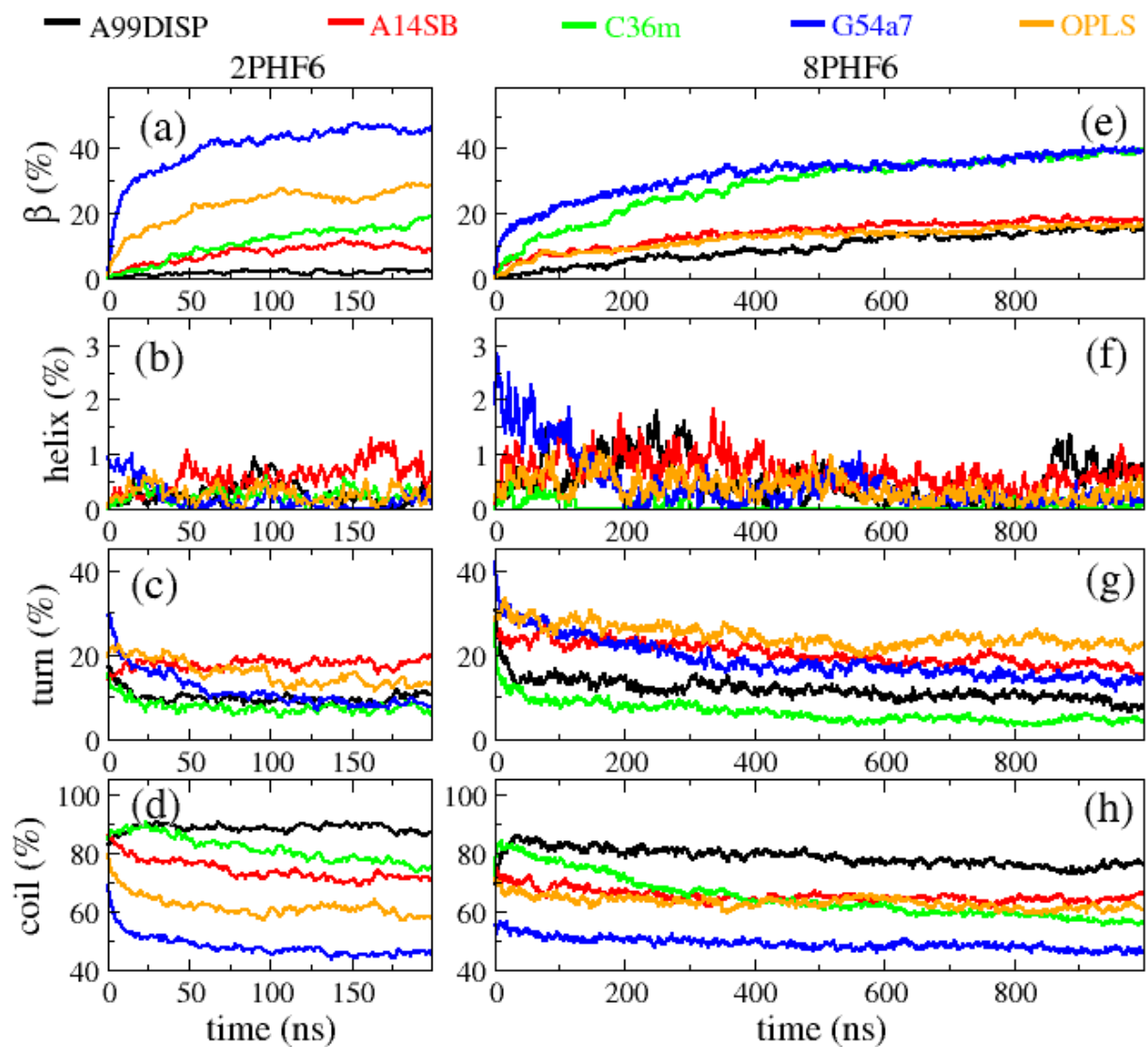


Figure S4: Evolution of the secondary structures of PHF6 peptides in different systems along the MD simulation time using different force fields

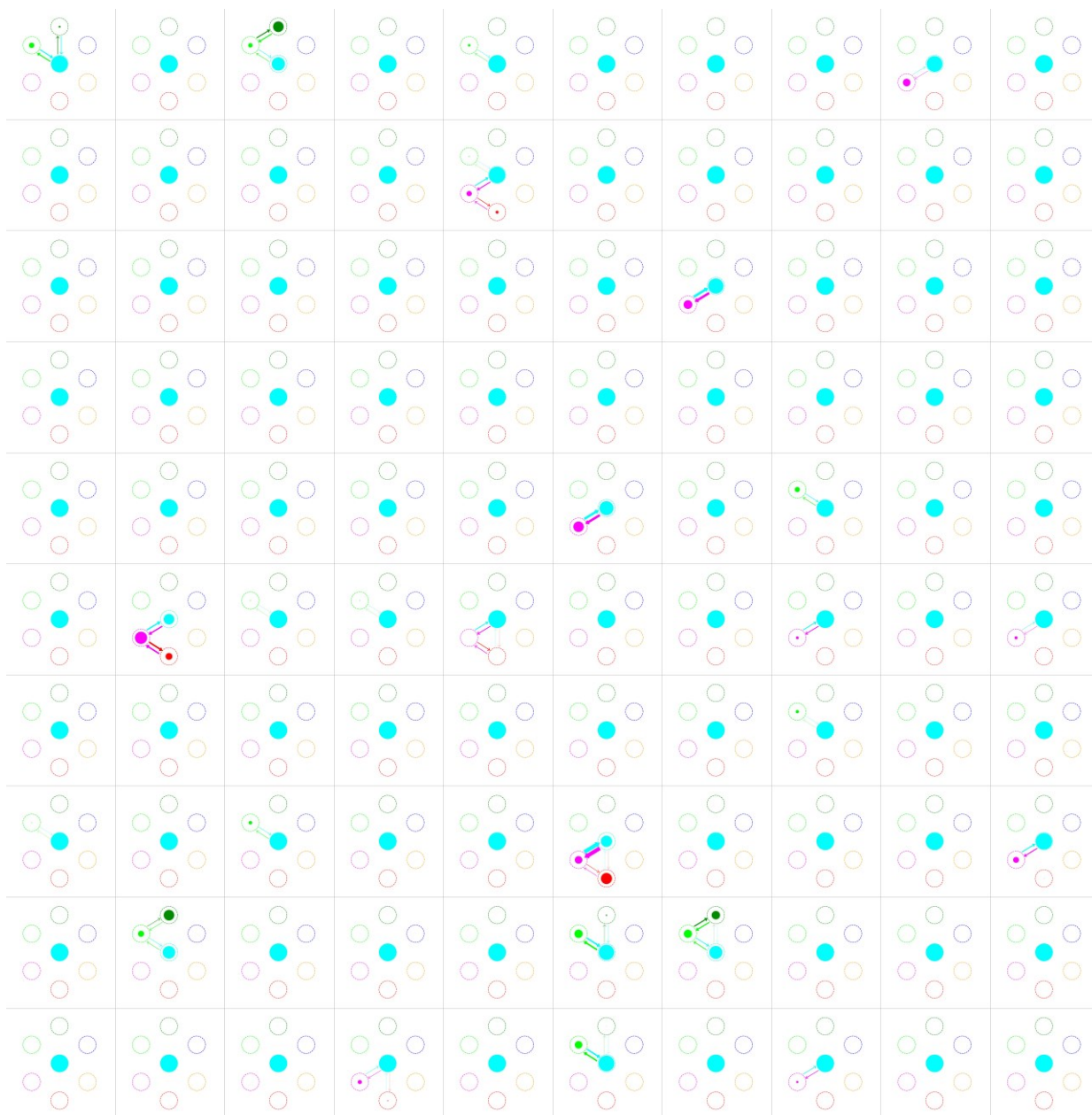


Figure S5: The populations of the seven states and the state transitions of the PHF6 dimer in the 2PHF6 system with A99SB-DISP force field. The analyses were performed for 100 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

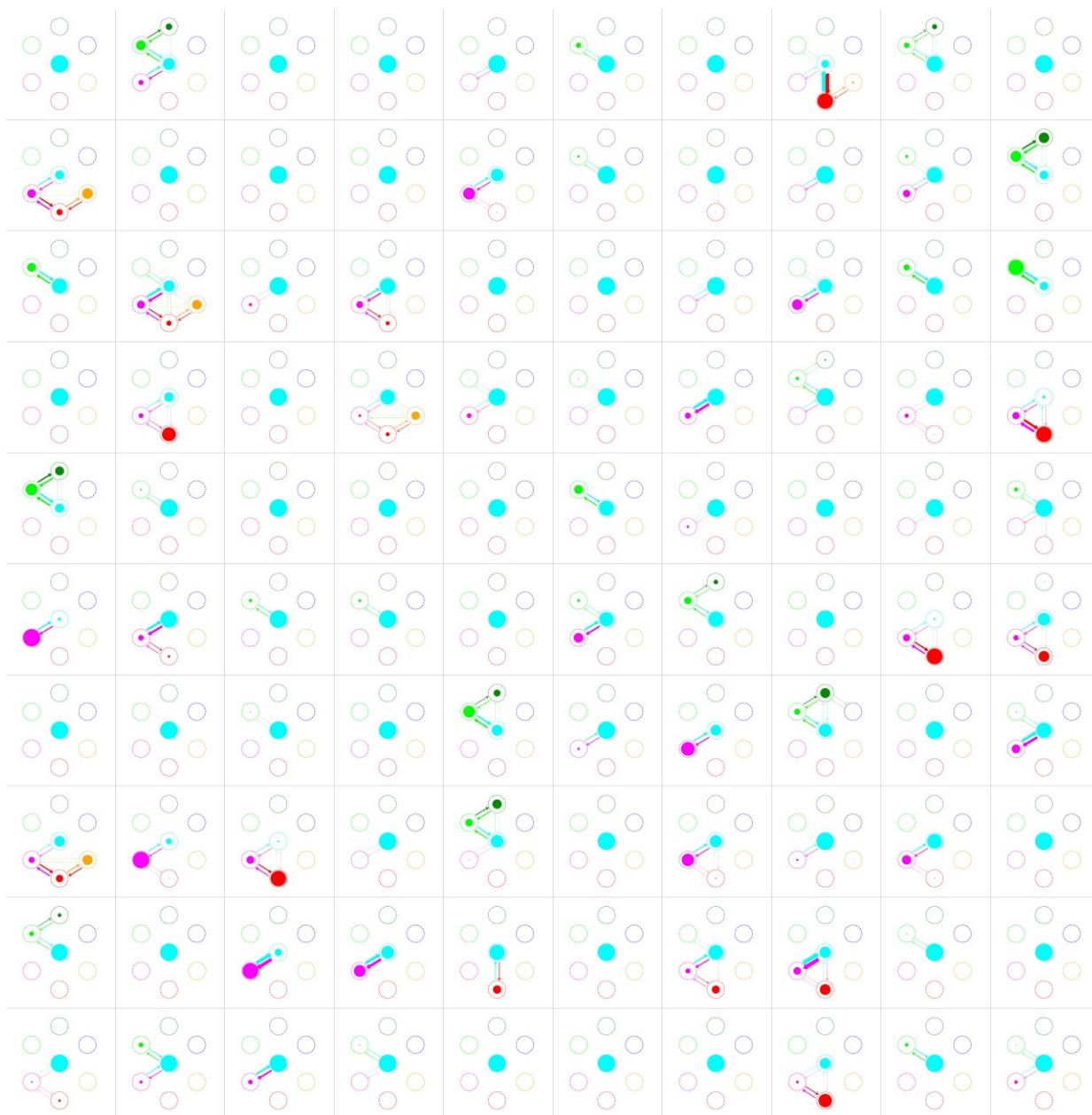


Figure S6: The populations of the seven states and the state transitions of the PHF6 dimer in the 2PHF6 system with A14SB force field. The analyses were performed for 100 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

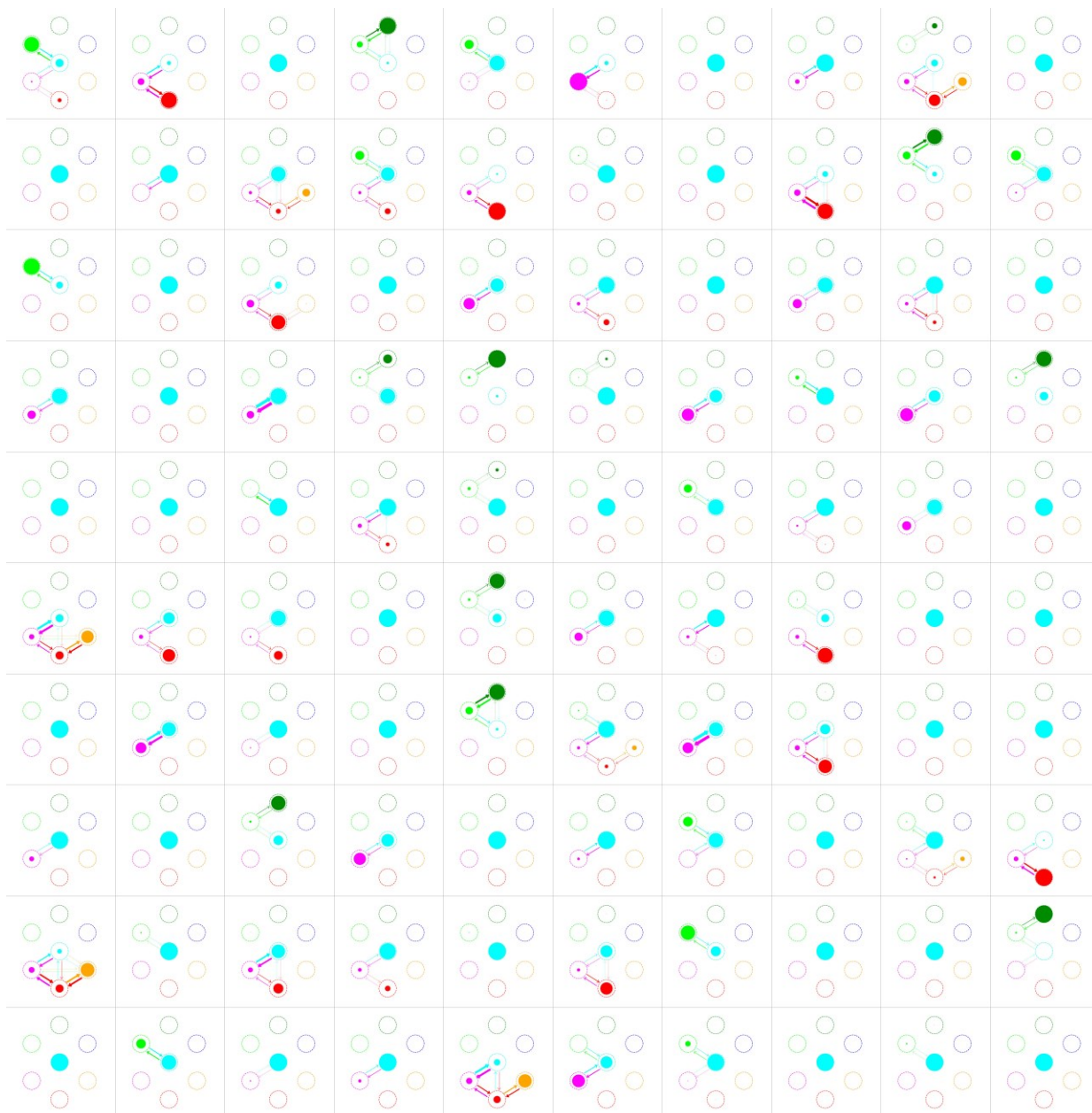


Figure S7: The populations of the seven states and the state transitions of the PHF6 dimer in the 2PHF6 system with C36m force field. The analyses were performed for 100 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

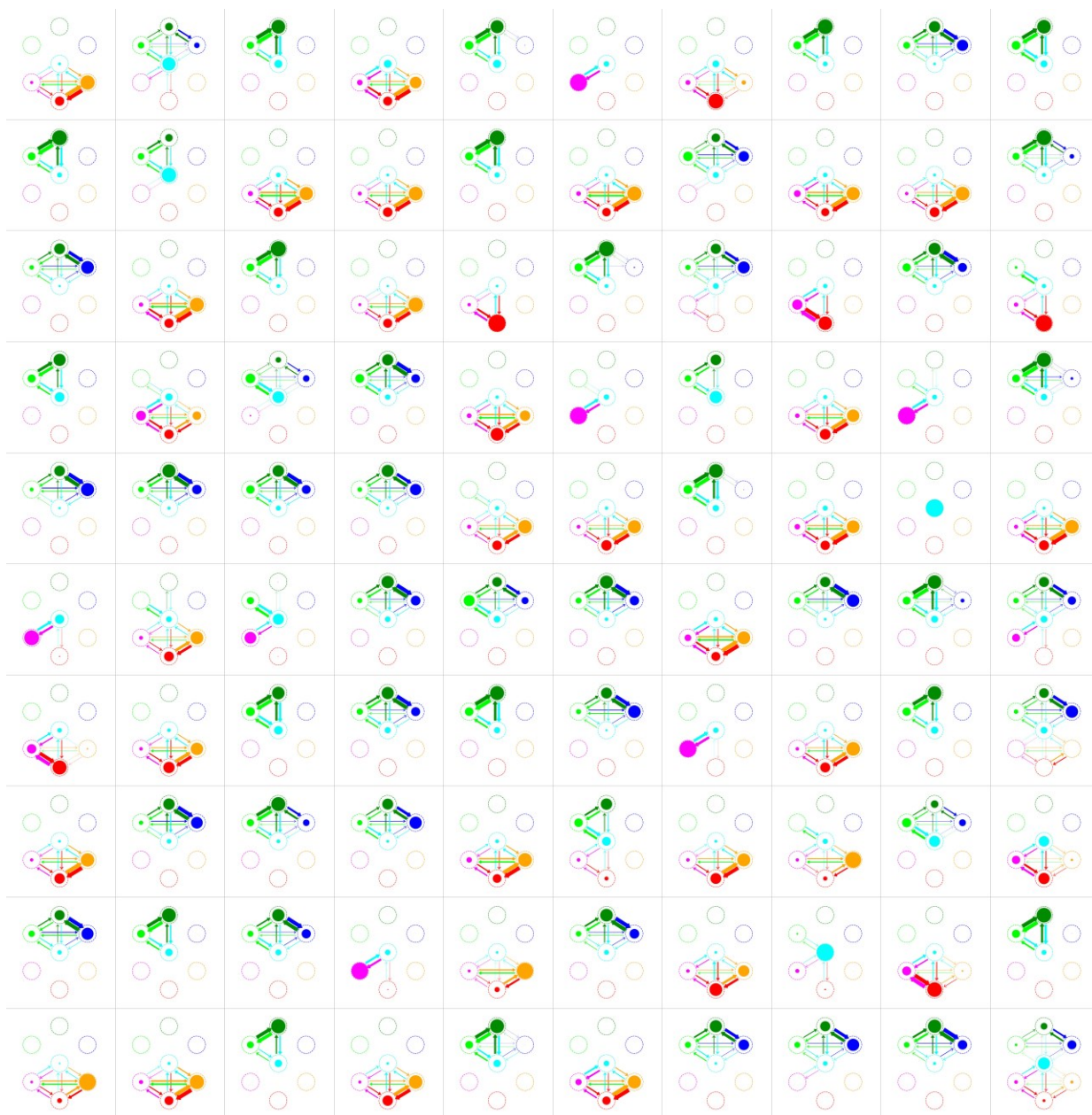


Figure S8: The populations of the seven states and the state transitions of the PHF6 dimer in the 2PHF6 system with G54a7 force field. The analyses were performed for 100 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

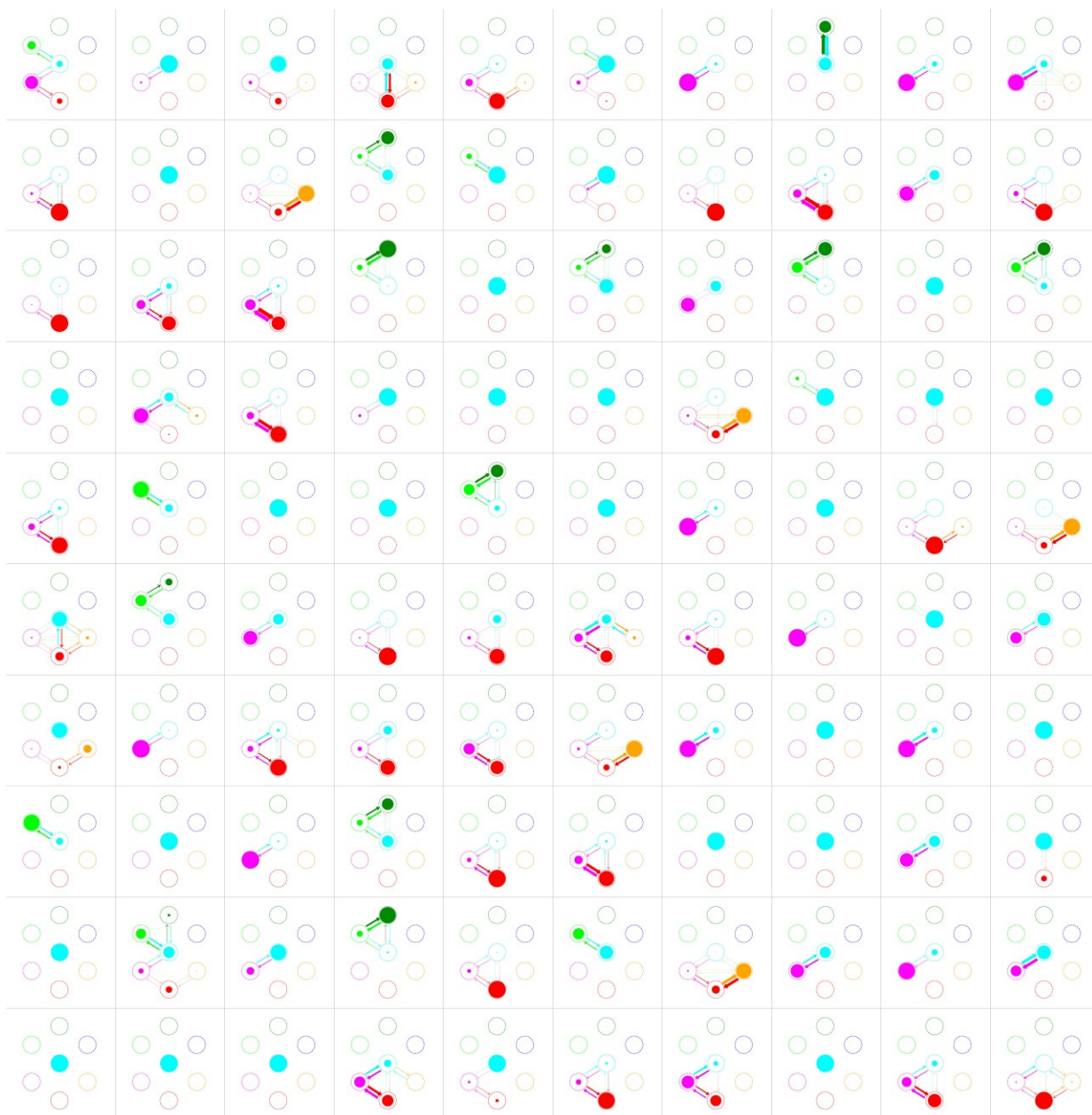


Figure S9: The populations of the seven states and the state transitions of the PHF6 dimer in the 2PHF6 system with OPLS force field. The analyses were performed for 100 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

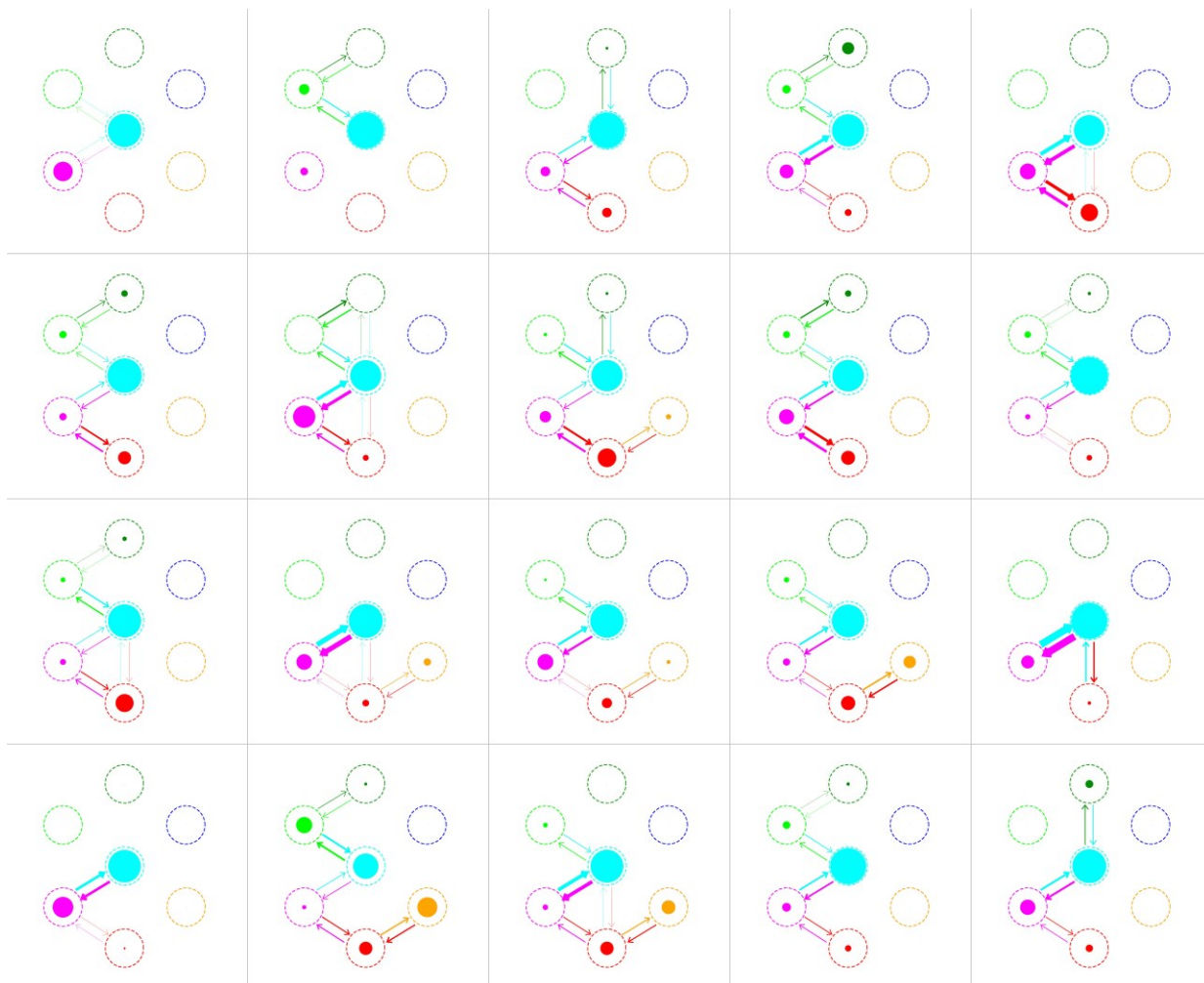


Figure S10: The populations of the seven states and the state transitions of the PHF6 dimer in the 8PHF6 system with A99SB-DISP force field. The analyses were performed for 20 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

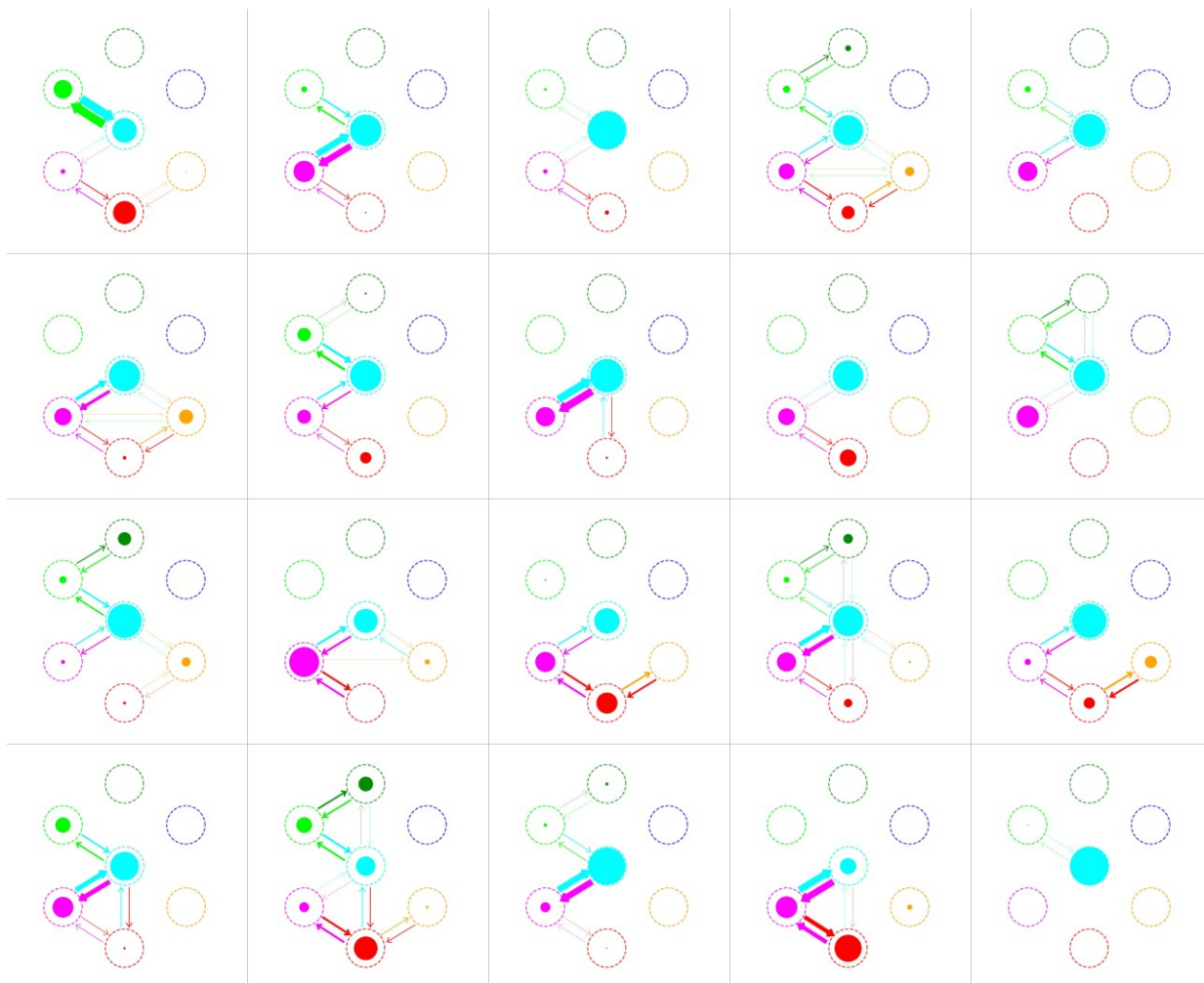


Figure S11: The populations of the seven states and the state transitions of the PHF6 dimer in the 8PHF6 system with A14SB force field. The analyses were performed for 20 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

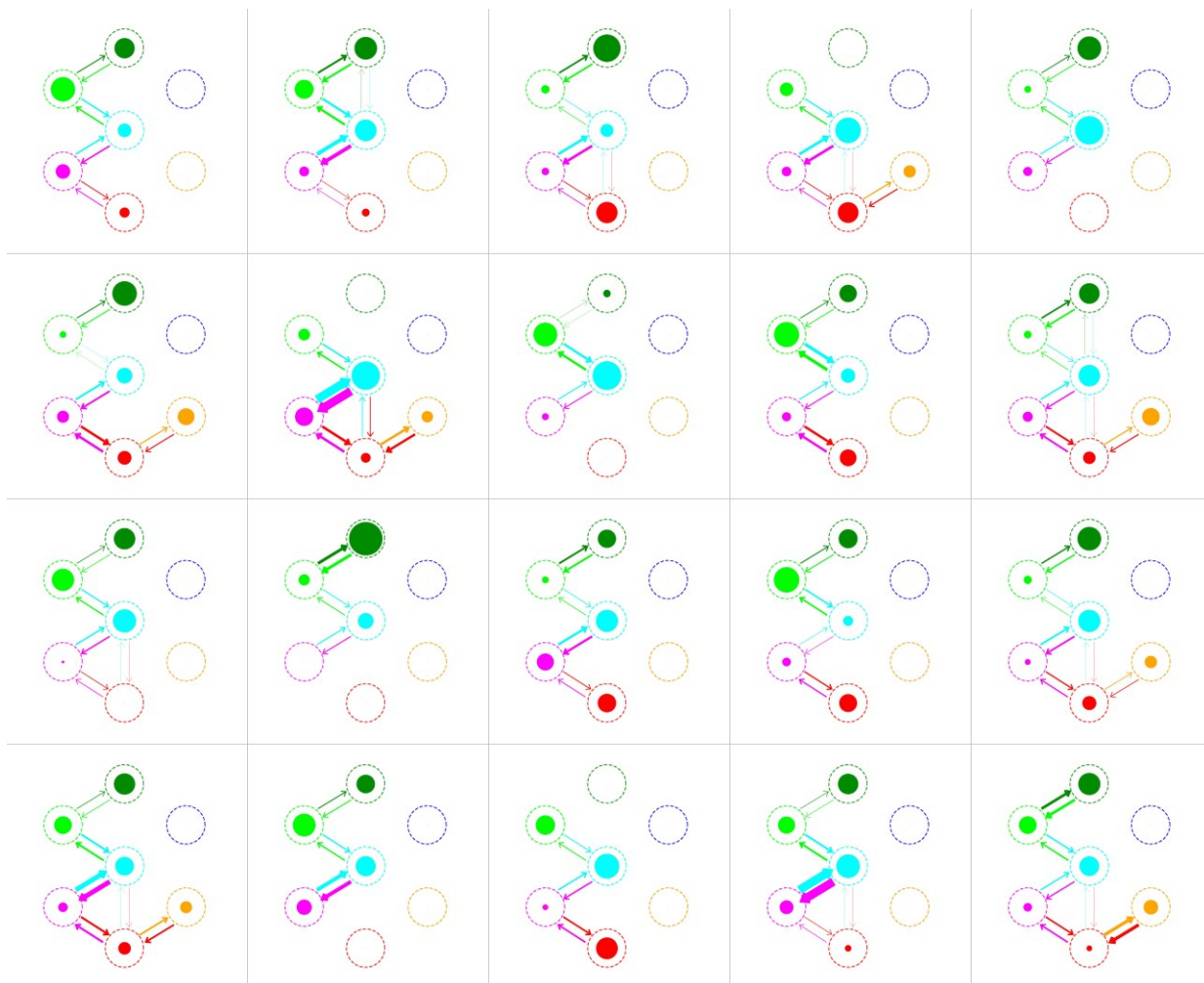


Figure S12: The populations of the seven states and the state transitions of the PHF6 dimer in the 8PHF6 system with C36m force field. The analyses were performed for 20 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

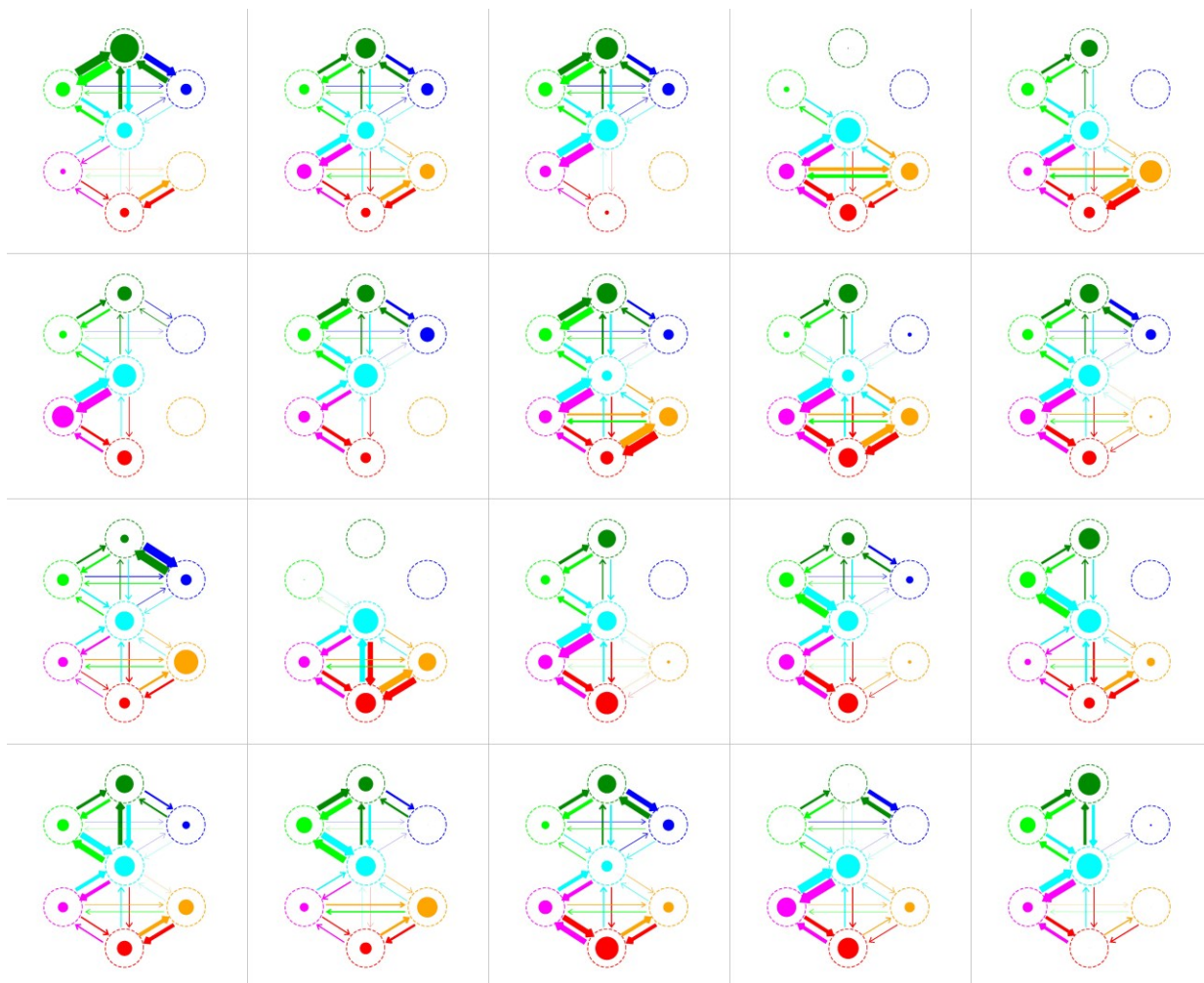


Figure S13: The populations of the seven states and the state transitions of the PHF6 dimer in the 8PHF6 system with G54a7 force field. The analyses were performed for 20 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

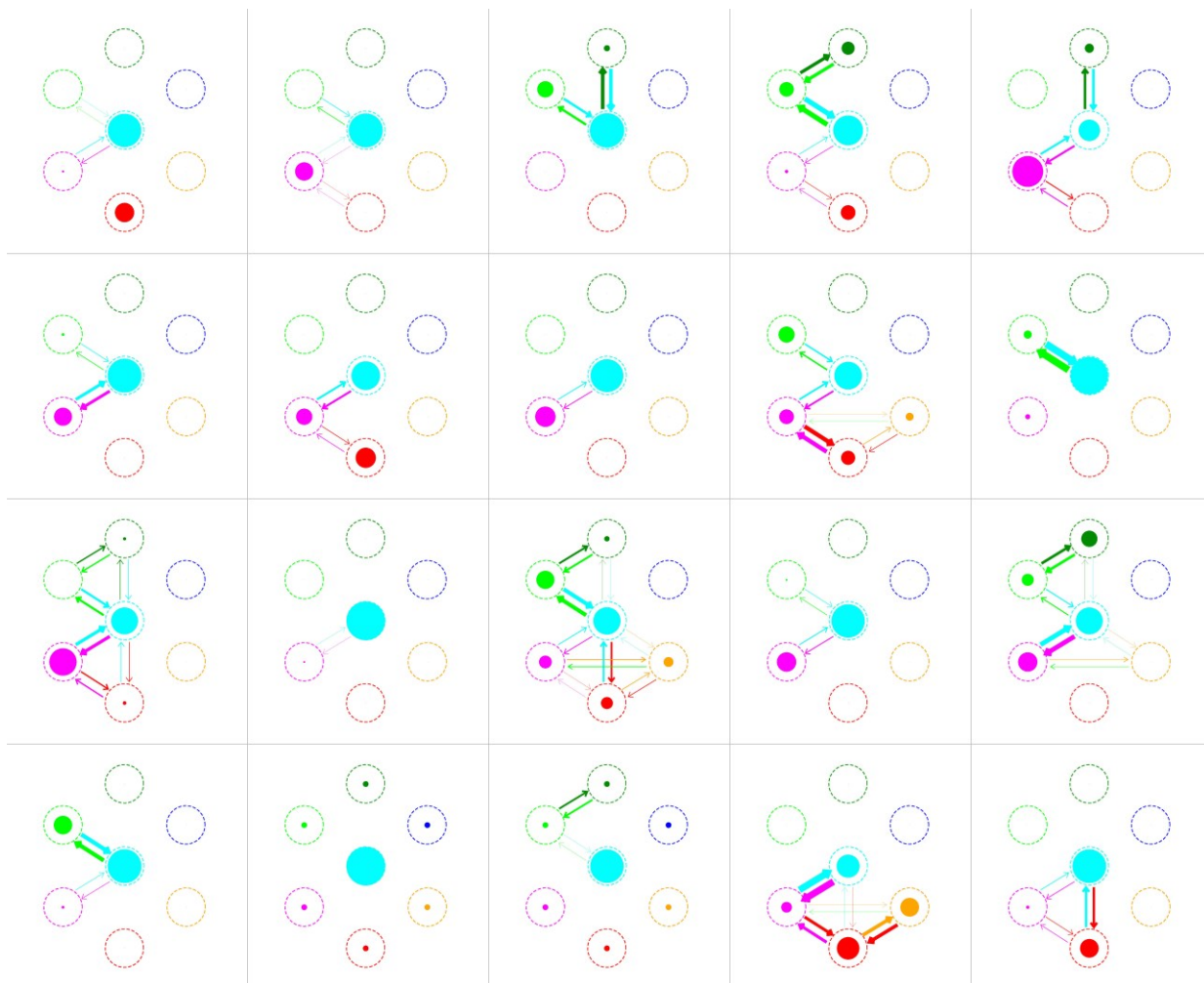


Figure S14: The populations of the seven states and the state transitions of the PHF6 dimer in the 8PHF6 system with OPLS force field. The analyses were performed for 20 MD trajectories and each was shown in a rectangle block. The definitions and representations of the seven states are presented in Figure 1.

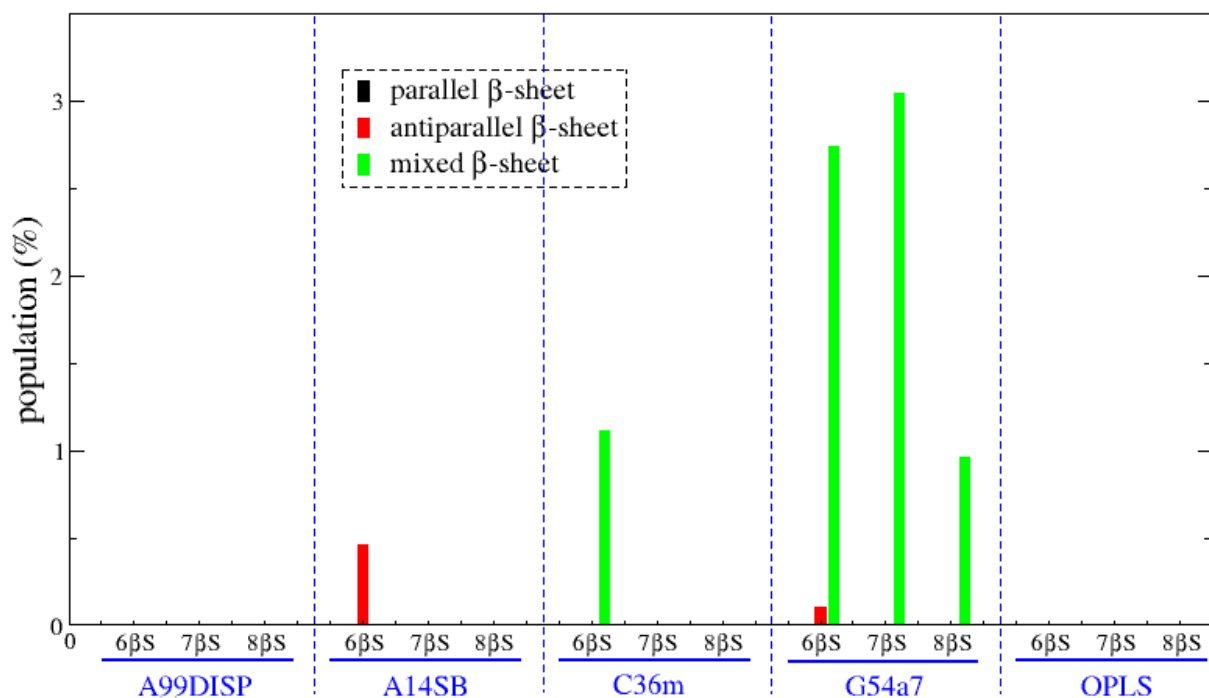


Figure S15: The populations of parallel, antiparallel and mixed type of β -sheet structures with different number of strands found sizes formed in the 8PHF6 system in using different force fields. A β -sheet structure containing i peptides is denoted as $i\beta$ S. For the 8PHF6 system that has having 8 monomeric peptides, i takes values of 2 to 8. This figure shows the distributions of three types for $i\beta$ S ($6 \leq i \leq 8$).

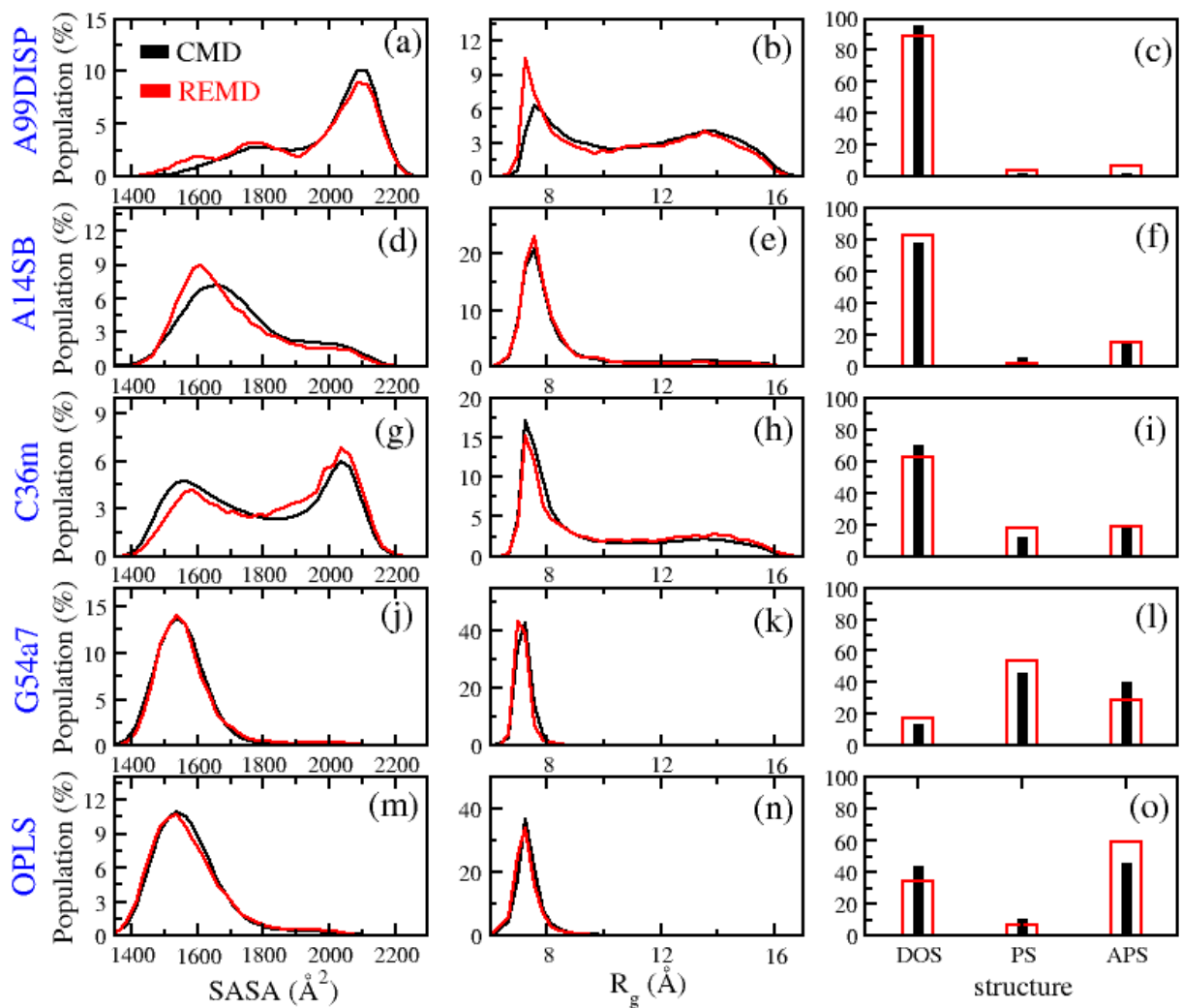


Figure S16: The comparison between conventional MD (CMD) simulation with 100 trajectories (black) and REMD simulation (red) for 2PHF6 systems. The data is shown for the population of SASA, gyration radius, and DOS, PS and APS structures. The REMD conformation data were collected when the replica temperature is 310 K, at which CMD were performed. The agreement between CMD and REMD confirms the convergence of conformational sampling of the multirun CMD simulation.