On the tracks of the aggregation mechanism of PHF6 peptide from Tau protein: molecular dynamics, energy and interaction network investigations.

Charline Fagnen[‡], Johanna Giovannini[‡], Marco Catto[†], Anne Sophie Voisin-Chiret*, and Jana Sopkova-de Oliveira Santos*

CERMN (Centre d'Etudes et de Recherche sur le Médicament de Normandie), Université de Caen Normandie, UNICAEN, 1 Boulevard Henri Becquerel, F-14032 Caen, France.

Corresponding author: jana.sopkova@unicaen.fr

ABSTRACT: The formation of neurofibrillary tangles (NFTs), composed of tau protein aggregates, is a hallmark of some neurodegenerative diseases called tauopathies. NFTs are composed of Paired Helical Filaments (PHF) of tau protein with a dominant β-sheet secondary structuration. The NFTs formation mechanism is not known yet. This study focuses on PHF6, a crucial hexapeptide responsible for tau aggregation. A 2μs molecular dynamics simulation was launched to determine the keys of the PHF6 aggregation mechanism. Hydrogen bonds, van der Waals and other non-covalent interactions as π-stacking were investigated. Parallel aggregation was slightly preferred due to its adaptability, but anti-parallel aggregation remained widely actual during the PHF6 aggregation. The analysis highlighted the leading role of hydrogen bonds identified at the atomic level for each aggregation process. The aggregation study emphasized the importance of Tyr310 during the β-sheets' complexation through π-stacking.

Introduction

The formation of neurofibrillary tangles (NFTs), composed of tau protein aggregates, is a hallmark of some neurodegenerative diseases like Alzheimer's disease, frontotemporal dementia, Pick's disease, and progressive supranuclear palsy. These diseases are named tauopathies.[1–4]

Briefly, 'tau' protein stands for 'Tubulin Associated Unit' because in its physiological form it is a microtubule-stabilizing protein through its Microtubule-Binding Domains (MTBD). This stabilizing function explains tau protein preferential brain location. Indeed, it is predominant in neurons for axonal transport.[5] When tau protein toggles from physiological to pathological shape, it means that tau protein has undergone abnormal phosphorylation, detachment from microtubule (MT) and abnormal self-aggregation.[6] This self-aggregation is what leads tau protein to form NFTs. NFTs are composed of Paired Helical Filaments (PHF) of tau protein with a dominant β-sheet secondary structuration.[7] The classical pattern of tau protein aggregate formation always begins with tau protein monomers detached from MT in a random coiled conformation. Then they go through nucleation, elongation, and end with a stable tau protein fibril.[8]

Although it remains incompletely elucidated, the severity of neurodegeneration symptoms (notably dementia and memory impairments) seems to be in high correlation with the progress of NTFs quantity in the brain.[9] Tau protein aggregation has numerous *in vitro* kinetics models to enlighten its behavior from random coil conformers to elongated aggregates,[10] but no model has yet described its initiation mechanism.

Tau protein is an intrinsically disordered protein that exists in 6 isoforms in human brain after alternative splicing of exons coming from one single gene.[11] All isoforms are composed of a MTBD that differs in terms of repeat domain number (3 to 4 domains).[12–14] MTBD is well-described for tau protein structural identification but it has also been incriminated for a central role in tau protein aggregation.[15] In fact, two hexapeptide sequences named Paired Helical Filaments (PHF6 and PHF6*), presented among the second and the third repetition domain, have been highlighted in *in vitro* experiments for their crucial role in tau protein aggregation. Both PHF6 and PHF6* are able to promote either their self-aggregation or tau protein aggregation.[16–18] They seem indeed crucial for tau protein aggregation, their deletion from tau protein preventing the whole system from aggregation.[19]

With a view to one day pharmacologically block NFTs formation, we intend to elucidate the earliest driving forces of tau protein aggregation.

Previous works have proposed different mechanism of tau protein aggregation process. Seidler *et al.* made an analysis of crystallized PHF6 and PHF6* steric interface.

They showed different modalities of aggregation process, differing from aligned to interdigitated interfaces.[18] Chen *et al.* performed *in vitro* and *in cellulo* assays to highlight how mutations on tau protein affects PHF6 aggregation process.[20] They concluded that P301L mutation has a high propensity to enhance PHF6 aggregation. Liu *et al.* have used the Markov state model to identify PHF6 aggregation pathway on a 1 μs molecular dynamic simulation of 16 PHF6 peptides. They highlighted the key residues involved in hydrophobic and hydrogen bonding interaction (I308, V309, and Y310).[21] On a larger inclusion of tau protein, Fitzpatrick *et al.* studied the conformation of PHF6 among filaments extracted from a patient's brain.[7] They highlighted different PHF6 behaviors from lone PHF6: in this kind of inclusion, PHF6 can aggregate on a self-packing model but it can also form a complementary packing interface with another peptide among tau protein sequence. So, PHF6 behavior studies rely on several works, but little is known about PHF6 aggregation initiation process at an atomistic level.

In this work, we focused on the study of PHF6 peptide ([306]VQIVYK[311]) aggregation mechanism. Here, we aim to present an inclusive representation of PHF6 possible behaviors observed during *in silico* experiments. This article focuses on the study of aggregation of the PHF6 hexapeptide[17] to understand the earliest steps of tau protein aggregation process.

RESULTS AND DISCUSSION

To study the PHF6 domain aggregation we performed a long molecular dynamics simulation of 2 μs on 62 PHF6 peptides. The PHF6 peptides were manually disposed 10 Å away from each over and surrounded by the waters and contra ions (NaCl=0.150 M). As the ultimate goal of our work is to obtain an *in silico* model able to predict the power of small synthetic compounds to disrupt tau protein-protein interaction in order to prevent its early aggregation stage, we focused here to elucidate the tau PHF6 sequence role in the aggregation process. As the *in vitro* experiences evaluating the anti-aggregating power of small compounds are most of the time carried out with the synthesized PHF6 peptide, we launched our simulation with protonated N and C-termini to be able to correlate our *in silico* predictions with the *in vitro* observations. After the system minimization and equilibration, a 2 μs production at 310 K at NPT conditions were carried out and analyzed in order to determine different aggregates which have formed. We investigated at atomic level the aggregation mechanisms and their variabilities during the simulation. For clarity, the peptide sequence [306]VQIVYK[311] was renumbered as [1]VQIVYK[6] in the manuscript.

Identification of the aggregation types. First, to determine the most stable aggregation mechanisms of the different systems observed during the simulations, the distances between the center of mass of each of the 62 peptides were computed on the last 1 μs of the simulation. For each distance, an average value was calculated to give a distance matrix (Fig. 1a). The average values greater than 40 Å were ignored, as these values are related to a non-aggregation or a late aggregation process. From this matrix, the different peptides that interact together could be identified. For example, it was possible to see that BH peptide interacts with J, Q, AG, AH, AQ, AS, AX, and BC. By connecting the commonalities, two PHF6 complexes were extracted: the complex CI (blue), the biggest one, composed of nine peptides (BH, BC, AX, AG, AQ, Q, AS, AH, J) and the complex CII (red) composed of eight peptides (V, N, D, AM, R, E, AE, and BG).

We have also identified one other trimer (orange) (F, AU, K); and three additional dimers (M/Z, AO/P, BA/BI). As the two first complexes represented the most advanced data at the aggregation level they were only retained for detailed analysis.

Timelines of the two complexes. Figure 1b illustrates complexes CI (Movie 1) and CII (Movie 2) formation timelines (the movies 3, 4 and 5 illustrate the replicas 2 and 3 results). The complex I is composed of one tetramer (BH/BC/AX/AG) called fragment 1 (CIFrag1) and one trimer (AS, AH, J) named fragment 2 (CIFrag2). As described in the timelines, the monomers individually added themselves to each other to form the fragments and the two fragments assembled into a complex at the middle of the simulation (1.02 μs). We have observed that the complexation does not prevent fragment extension: at 1.33 μs, CIFrag1 was extended by two peptides (AQ/Q). At the end of the simulation, we notice that the previously detected M/Z dimer finally binds to CIFrag2 to extend it (1.55μs). The extension is not analyzed in detail here because it occurred at the end of simulation without proof of stability, but it is still interesting to see that the two fragments can grow after the complexation.

The complex CII is composed of one pentamer (V, N, D, AM, R) named fragment 1 (CIIFrag1) and one trimer (BG, AE, E) named fragment 2 (CIIFrag2). As with the first complex, the aggregation process occurred peptide by peptide and the two fragments assembled in a complex at 1.95 μs, but CIIFrag1 pentamer lost one peptide (R) and became a tetramer during this step.

Parallel or anti-parallel aggregation? Analysis of our PHF6 simulations revealed different aggregation stages: dimerization, trimerization, tetramerization, pentamerization, fragment complexation and extension of a complexed fragments. The first question that came up, was what is the predominant peptide behavior in the different aggregates, parallel or antiparallel?

Figure 1c reports the percentage of parallel and antiparallel structures in dimers, trimers, tetramers, pentamers presented throughout the simulation. We observed that 54.50% of the formed dimers present a parallel interface, slightly higher than anti-parallel. In the trimer distribution, most trimers (57.19%) have one parallel and one anti-parallel interfaces, while two anti-parallel or two parallel

interfaces account for 30.53% and 12.28% of the trimers, respectively. This aggregation initiation does not describe any clear preference between parallel and anti-parallel interfaces. In the tetramer distribution, three major parallel – one parallel interfaces (26.84%), the third conformation is parallel and two anti-parallel interfaces (22.67%). At this aggregation level for the first time, the parallel preference is clearly displayed: 70.09% of the conformations account for 92.76% of cases. The first conformation consists of two parallel and one anti-parallel interfaces (43.25%), the second conformation is one parallel – one anti- structures have two parallel interfaces out of the three. Figure 1c shows that two types of pentamers represented 91.28% of the aggregated pentamers during the simulations. The first conformation has two parallel and two anti-parallel interfaces (81.18%) and the second conformation has three anti-parallel and one parallel interfaces (10.1%). So, the pentamerization involves both parallel and anti-parallel aggregated species. The hexamers are not shown in Figure 1c because only one conformation represented 97.29% of the hexamers. This hexamer is described by two parallel, two anti-parallel, and one parallel interfaces. The hexamer showed a new a preference for parallel aggregation, but it was able to aggregate only thanks to anti-parallel aggregation. In conclusion, from our results, both mechanisms appear to be required for aggregation of the PHF6 peptides.
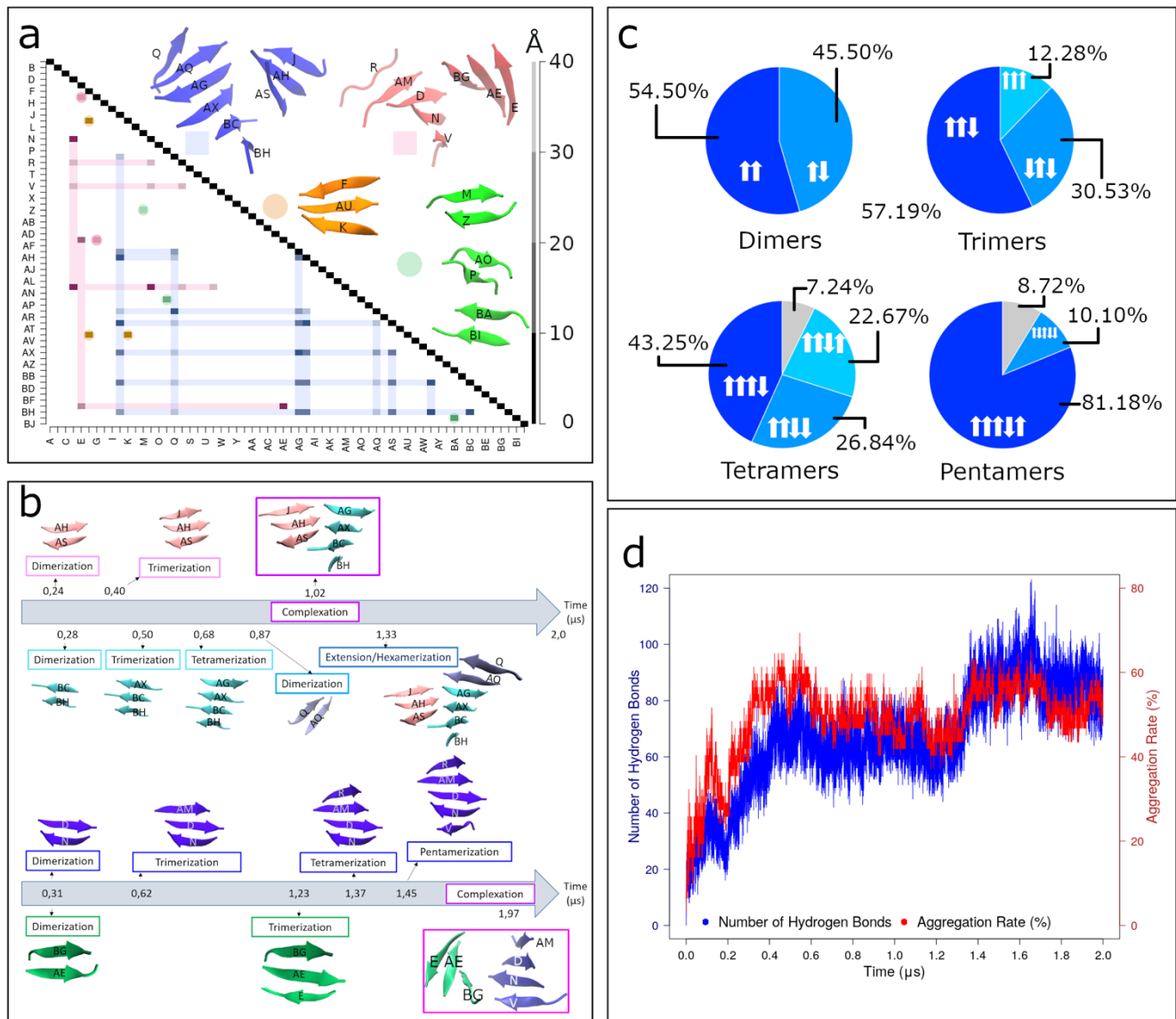
Figure 1. PHF6 aggregation process. a) Distances Matrix computed between the peptides' center of mass. The presented values are the distance averages between the two centers of mass on the last microsecond of simulations. The more the dot is dark, the more the two centers of mass are close. Average values greater than 40 Å were represented in white. The blue and red networks link the dots of the CI and CII complex, respectively. Orange and green circles describe the trimer and dimers, respectively. b) Timelines of the formation of the CI and CII complexes. Concerning CI, the fragments 1, 2 and the extension are respectively represented in cyan, coral and violet. Concerning CII, the fragments 1 and 2 are respectively represented in blue and green. c) Percentages of the different types of β-strands observed in dimers, trimers, tetramers and pentamers. d) Double ordinate axis chart comparing the total number of Hbonds created by all peptides with the progress of the aggregation rate (%), for each frame.

Parallel Aggregation. All along the two complexes formation, seven parallel interfaces were observed: three in dimers, three in trimers and one in tetramer. The three parallel dimers are: BH-BC that initiates the CIFrag1 aggregation, AQ-Q that is the dimer that extended the complex CI and BG-AE that starts the CIIFrag2 aggregation. Among the trimers, the first trimer (BH-BC-AX) is a fully parallel trimer located in the CIFrag1; the second and the third came from an anti-parallel dimerization but their trimerization interface is parallel: AS-AH-J (CIFrag2) and N-D-

AM (CIIFrag1). The last case is the tetramer N-D-AM-R (CI-IFrag1) that is formed from an anti-parallel dimer through a parallel trimerization and its tetramerization is parallel. The * notation is attributed to the last peptide that aggregates, or in the case of dimer to the hexapeptide that hosts the following peptide in the next step of aggregation.

Overall, three different mechanisms of parallel aggregation (PA) were detected: i) aggregation initiation and prolongation by parallel interfaces without constraint, ii)

4

aggregation prolongation following a previous parallel interface, iii) aggregation prolongation taking into account π-stacking between the β-sheets of the complex.

Mechanism 1 of parallel aggregation: Preferential Mechanism. This first PA mechanism is the preferred mechanism at the initiation of a parallel dimerization (3/3 dimers). It is also the most frequent (5/7) of the overall PA mechanisms; it is not only used for the aggregation initiation (dimerization) but also during the prolongation (post-dimerization) depending on the involved residues' availability. This mechanism was observed in the three parallel dimers BH-BC, AQ-Q, BG-AE and in the trimerization interfaces AS-AH/J and N-D/AM and is illustrated here by the BH-BC dimerization and AH-J prolongation (Fig. 2a). The figure describing the other parallel interfaces is available in Supplementary Information (Fig. S1).

Looking at the representative structures of these systems (Fig. 2a) and their interaction networks (Fig. 2b), four common hydrogen bonds (Hbonds) were detected: Q2-Q2*, V4-Q2*, V4-V4* and K6-V4*. The Hbond Q2-Q2* was detected in BH-BC, BG-AE dimers but was absent in the AQ-Q one (Fig. S1). Nevertheless, detected van der Waals contacts in the interaction network varied among the three interfaces. To detect the main interactions occurring between the two interfaces during the dimerization, total and van der Waals interaction energies were computed.

Eight main interactions, detected in the total interaction energy analysis (with average value $|E_{int}| > 3.00$ kcal/mol), involved five of six residues (Q2, I3, V4, Y5, K6) of the first monomer and four of six residues (V1*,Q2*, I3*, V4*) of the second monomer (Fig. 2c): Q2-V1* (-3.0 ± 1.6 kcal/mol), Q2–Q2* (-3.3 ± 1.7 kcal/mol), I3-Q2* (-3.9 ± 0.7 kcal/mol), V4-Q2* (-3.9 ± 0.7 kcal/mol), V4-I3* (-3.9 ± 0.6 kcal/mol), V4-V4* (-4.0 ± 0.7 kcal/mol), Y5-V4* (-4.0 ± 0.9 kcal/mol), and K6-V4* (-4.4 ± 1.1 kcal/mol). From this analysis, among the eight highlighted strong interaction, four are directly lying with the Hbond's formation. The residues not participating to the interaction between the two monomers are located on the dimer's extremities: Y5 and K6 of the first monomer and V1* of the second one. On the other hand, the two central residues V4, V4* of each monomer contribute to five of the eight interactions and are responsible for 36.3% of the total interaction energy between monomers.

To explain four remaining strong interaction energies (I3-Q2*, V4-I3*, Y5-V4* and K6-Y5*), not correlated with a Hbond's formation between the residues, an investigation of the van der Waals interactions (Fig. 2d) was carried out. This analysis revealed four strong contacts: I3-Q2*, V4-I3*, Y5-V4*, K6-Y5* with average values of -2.2 ± 0.8 kcal/mol, -2.2 ± 0.7 kcal/mol, -2.5 ± 0.8 kcal/mol, -2.2 ± 0.9 kcal/mol, respectively. The first three detected van der Waals interactions coincide with the three already distinguished in the total interaction energy analysis. The last one between Y5* and K6 represents a van der Waals contact at the dimer extremity, but its total interaction remains limited. Therefore, van der Waals interaction between the monomers is mainly led by I3, V4, and Y5 that contribute at 21.0%, 20.9%, and 17.1%, respectively.

Analyses of the Hbonds all along the dimers trajectories were carried out to correlate the energetic level with the Hbond's presence (Fig. 2e). The four common Hbonds could be divided into two groups depending on their volume occupancy values. The first group gathers the Hbonds with high scores values: V4(HN) – Q2*(O) Hbond showing the highest volume occupancy's scores (average value: 0.934) and V4(O) – V4*(HN) one (average value: 0.925). Both Hbonds are located on each representative structure (Fig. 2a) at each step of the aggregation process (Fig. S2). It can be concluded that the first bond to be created is V4(HN) – Q2*(O), then V4(O) – V4*(HN). Their consistency and their volume occupancy highlight their crucial role in the parallel dimer's aggregation.

The second group includes the Hbonds with intermediate occupancy volume values: K6(HN) – V4*(O) and Q2(O) – Q2*(HN). The lowest rates concerned Q2(O) – Q2*(HN) Hbond with the average score of 0.796; this bond was detected at each aggregation state except for the extension following the complexation (Fig. S1-2). We concluded that it is the last one to be created and is less crucial to the aggregation. The volume occupancy values of K6(HN) – V4*(O) were higher than Q2(O) – Q2*(HN) ones and it was detected at each aggregation step, so this is the third Hbond to be created.

In conclusion, this first PA mechanism takes place at the center of the interface where V4 from both monomers of this interface plays a key role. These results are coherent with the Hbonds previously found: among the four detected Hbonds, V4 contributed to three of them. The Hbonds' formation order was deduced as follows: 1) V4(HN) – Q2*(O), 2) V4 (O) – V4* (HN), 3) K6(HN) – V4*(O), 4) Q2(O) – Q2* (HN). The first three seem essential in the aggregation of a parallel dimer. The MM-PBSA analysis per residue calculated along the trajectory for mechanism 1 highlighted as the major contributing residues to the dimer binding free energy: Q2*, V4* and V4 in initiation and in prolongation (Fig 3), the residues participating to the H-bonds (Fig 2). Interestingly, the Alascan calculation on the representative structure, allowing us estimate the side chains' contribution to the dimer stabilisation, showed that the stabilizing residues differ between the initiation and prolongation mechanism. The major stabilizing effect is due to the Q2*, I3*, V4*, K6 in the initiation and it is of a hydrophobic origin (Fig S). These residues coincide with the strong non-polar binding free-energies (Fig 3). While in the prolongation mechanism V4 and V4* played a destabilising role. The van der Waals energies showed a one residue interaction shift between the two hexapeptides correlated with a one-residue-alignment shift of peptides.

Mechanism 2 of parallel aggregation: Taking into account the previous parallel aggregation. This second PA mechanism was detected in a full parallel trimer on the trimerization interface BG-BH/AX (Fig. 2a), K/BE/BD (Fig. S10) and BE/Z (Fig. S12). To accommodate in parallel the third peptide, BH and AX had to align according to the availability of the Hbond's donor/acceptor on BH peptide. The previous parallel dimerization between BG and BH monopolized the key residues involved in the first mechanism (Fig. 2a-b) and so the AX peptide had to adjust along other available residues. The Hbonds between the main chain atoms stabilizing this interface in this mechanism are: $I_3(O) - Y_5^*(HN)$, $I_3(HN) - I_3^*(O)$, $V_1(O) - I_3^*(HN)$ and $Y_5(HN) - Y_5^*(O)$ (Fig. 2a).

The profile of main interactions detected for this parallel interface BH-BC/AX is also different from the previous one. From total interaction energy analysis, seven main interactions were detected (Fig. 2c): V_1-I_3^* (-3.1 ± 1.6 kcal/mol), Q_2-I_3^* (-4.3 ± 1.0 kcal/mol), I_3-I_3^* (-4.2 ± 0.8 kcal/mol), I_3-V_4^* (-3.8 ± 0.7 kcal/mol), I_3-Y_5^* (-4.4 ± 0.8 kcal/mol), V_4-Y_5^* (-3.0 ± 0.9 kcal/mol), and Y_5-Y_5^* (-3.5 ± 2.1 kcal/mol). The I_3-Y_5^*, I_3-I_3^*, V_1-I_3^* and Y_5-Y_5^* reflect the Hbond formation in this mechanism and Q_2-I_3^*, I_3-V_4^* and V_4-Y_5^* are correlated with the van der Waals
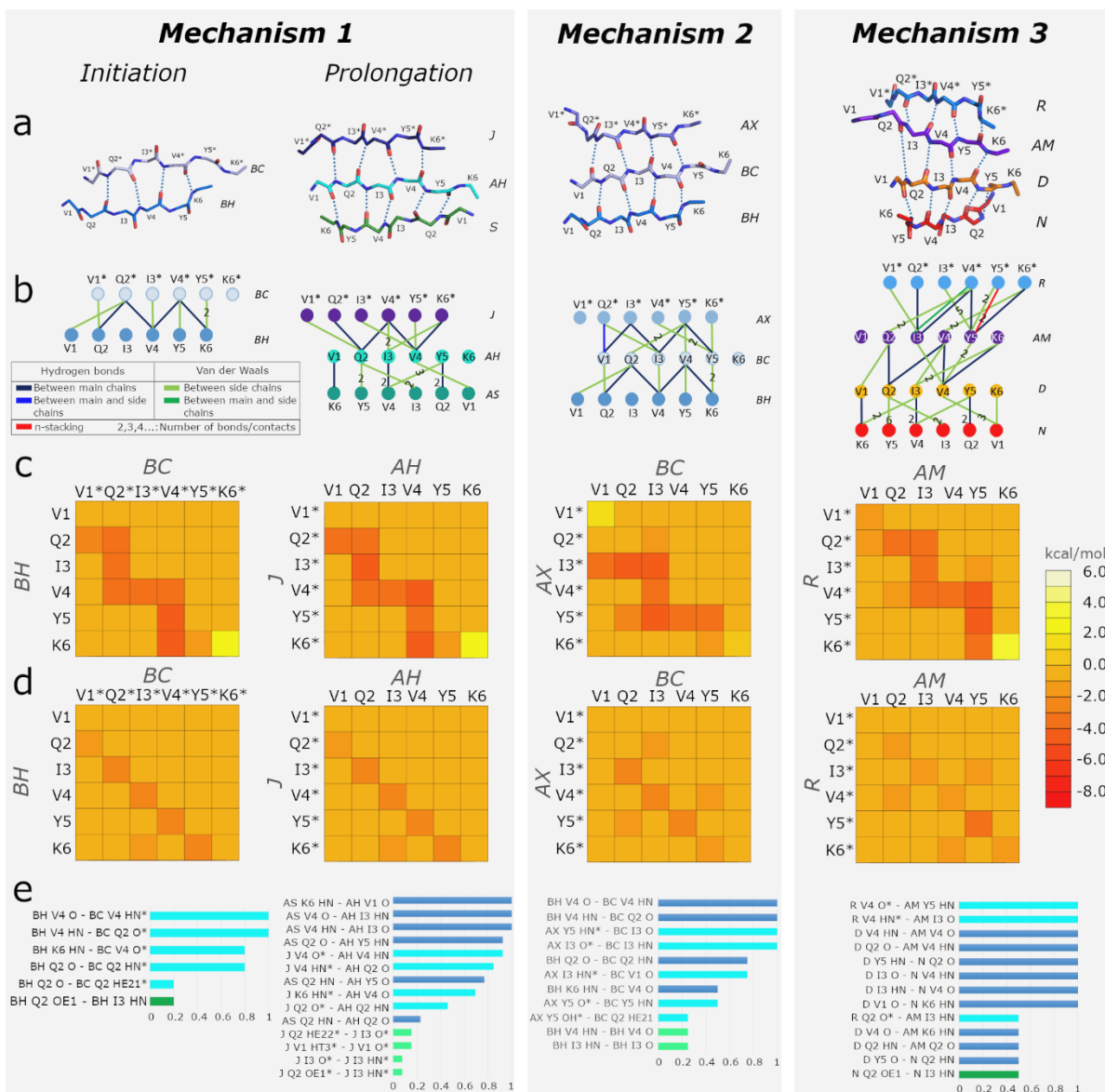
6

Figure 2. Three parallel mechanisms. a) Representative structures for each mechanism. b) Interaction network of the representative structures. Blue, green and red lines represent Hbonds between main chains, van der Waals contacts, and π-stacking respectively. The number on/near the line represents the number of this interaction type. c) Interaction energy matrix per residue. d) Van der Waals interaction energy matrix per residue. e) Volume occupancy histograms of the detected Hbonds (Cyan and dark blue bars annotate to the inter-monomer Hbonds of new and previous interfaces, respectively; green ones the intra-monomer HBonds).

contacts (Fig. 2d), they represented the highest detected van der Waals energies: Q2-I3* (-2.3 ± 0.8 kcal/mol), I3-V4*(-2.1 ± 0.7 kcal/mol), V4-Y5* (-2.0 ± 0.6 kcal/mol).

Hbonds study was carried out to explain the importance of the detected Hbonds for this interface. The four detected Hbonds between the main chain atoms corresponded to the highest interaction energies seen in the previous analysis: I3(O) – Y5*(HN), I3(HN) – I3*(O), V1(O) – I3*(HN) and Y5(HN) – Y5*(O). An additional low frequency (0.25) Hbond between side chain atoms Q2(HE2) and Y5*(OH) was also detected. The four main Hbonds

were found on the interaction network of the trimer representative structure with a frequency greater than 0.5 (Fig. 2b) and they can be classified into two groups: i) high volume occupancies and ii) intermediate. The first group comprises three Hbonds: I3(HN) – I3*(O), I3(O) – Y5*(HN) and V1(O) – I3*(HN) with volume occupancies of 1.000, 1.000 and 0.750, respectively. Moreover, the interaction networks extracted on representative structures at each CIFrag1 aggregation step involving BG-BH/AX trimer always showed these three Hbonds (Fig. 2b, S2). Their high occupancy levels and their constancy demonstrate therefore their crucial role in aggregation.
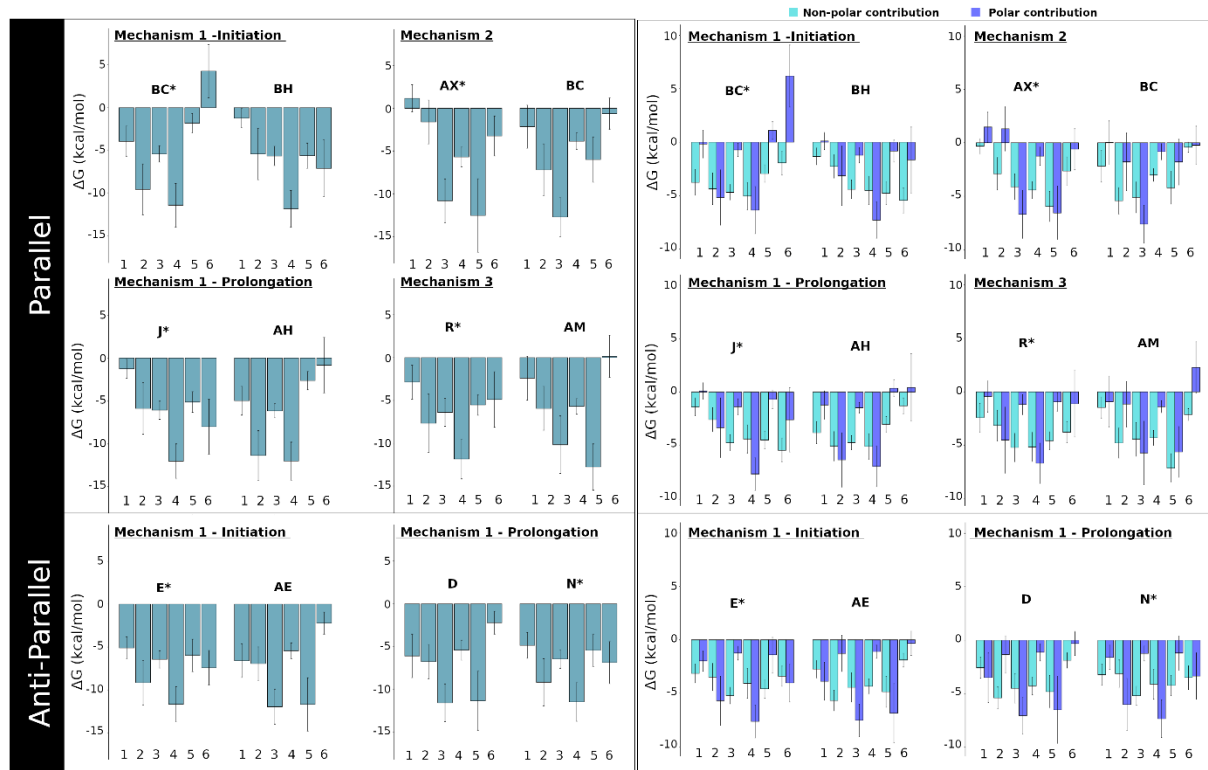
7

Figure 3. MM-PBSA analysis per residue on the representative parallel and antiparallel dimers along trajectory. On the right panel total free-energy contribution per residue, on the left decomposition to the polar and non-polar parts of the free-energy.

The second group is composed of only Y5(HN) – Y5*(O) Hbond, which was seen for the half of the trimerization trajectory (volume occupancy 0.5). Interestingly, its frequency increased along the tetramerization trajectory (0.714, Fig. S2) demonstrating stabilization of this Hbond during the following aggregation step.

The formation order of these Hbonds is: firstly, I3(HN) – I3* (O), I3(O) – Y5* (HN), and V1(O) - I3*(HN) and finally, Y5(HN) – Y5*(O). Only the first Hbond group seems to be fundamental in this aggregation mechanism, and I3 or I3* residue are involved in each of these Hbonds. Interestingly, the observed Hbond network in this shifted parallel interface was different from the one observed in mechanism 1. The MM-PBSA analysis per residue along the trajectory for this dimer confirms this result as they brought up as the major contributors to binding free energy, I3*, Y5* and I3 (Fig 3), residues involved in the formed Hbonds. More, the Alascan calculation on the representative structure revealed as the highly stabilizing residue Y5*, Q2, I3, V4 (Fig S9) and their stabilizing role is of hydrophobic character (Fig 3).

This second mechanism describes a process where new central residues are involved: the two ended residues V1 and K6 are only responsible for 12.56% of the total interaction energy and are involved in any Hbond, whereas Q2, I3, V4, and Y5 are responsible for 20.35%, 23.16%, 18.96%, and 24.97% of interaction energy, respectively. The main difference between this interface and the previous one is in the central residue sharing. Indeed, in this interface the crucial role in the aggregation is carried out by I3 and not by V4. This is explained by the fact that antedated dimer parallel interactions have already engaged the donor and acceptor of V4 residue and so they are not available anymore for a new aggregation step.

Mechanism 3 of parallel aggregation: Taking into account the π-stacking. The last PA mechanism was observed in the tetramer interface N-D-AM/R (Fig. 2a-b). This tetramer was built from an anti-parallel dimer and a parallel peptide addition in the trimerization step. During its aggregation, the two Y5 side chains interacted reciprocally through a π-stacking which impacted the neighboring interactions. So, calculated interaction network on the representative structure of this parallel tetramerization N-D-AM/R interface detected only three Hbonds between the main chain atoms: Y5(HN) – V4*(O), I3(O) – V4*(HN), and I3(HN) – Q2*(O) (Fig. 2b).

Interaction energy studies of the N-D-AM/R interface identified seven main interactions (Fig. 2c): Q2-Q2* (-3.7 ± 0.0 kcal/mol), I3-Q2* (-3.3 ± 0.2 kcal/mol), I3-I3* (-3.4 ± 0.5 kcal/mol), I3-V4* (-3.5 ± 0.9 kcal/mol), V4-V4* (-3.5 ± 0.1 kcal/mol), Y5-V4* (-4.3 ± 0.5 kcal/mol), and Y5-Y5* (-4.5 ± 0.0 kcal/mol). The three interactions: I3-Q2*, I3-V4*, and Y5-V4* are correlated with the presence of Hbonds. The van der Waals contacts are at the origin of the remaining interaction energies. In this interface the face-to-face resi-

8

dues (Q2-Q2*, I3-I3*, V4-V4*, Y5-Y5*) interacted reciprocally and are responsible for the detected interaction energies.

Y5 is involved in 26.72% of the total interaction energy, but two central residues, I3 and V4, showed a high contribution to the interaction between the two monomers: they respectively represent 23.54% and 24.61%. Here, the highest detected van der Waals interaction energy is between Y5, due to the perfect alignment between the two monomers that leads to the π-stacking formation between the two residues.

All along the tetramer N-D-AM-R trajectory, these three Hbonds Y5(HN) – V4*(O), I3(O) – V4*(HN), and I3(HN) – Q2*(O) are conserved (Fig. 4a) with volume occupancies of 1.00, 1.000, and 0.500, respectively. All these Hbonds are observed on the representative structure at each aggregation step (Fig. 2b and Fig. S2 CII Tetramer). In the interaction network of the representative structure of the tetramer a fourth Hbond was detected: Y5(O) - K6*(HN), but this bond is not found on all the trajectory, so its importance is lower.

The creation order between Y5(HN) – V4*(O) or I3(O) – V4*(HN) cannot be determined, but their occupancy volume highlights their fundamental role in the aggregation. The Hbond I3(HN) – Q2*(O) is the last to occur. These interactions are different from the previous cases because of the π-stacking interactions between the two Y5 (Fig. 2b). The MM-PBSA analysis per residue along all trajectory annotated for this interface as the major contributing residues, Q2*, V4*, I3 and Y5 (Fig 3). The Alascan calculation showed that the stabilizing role is played by I3*, V4*, Y5*, K6* and Y5 (Fig. S9) and that is of hydrophobic character (Fig 3). The Y5 non-polar contribution was the highest one detected among all residues for this interface which highlights the importance of Y5 in this interface (Fig 3). The study of the parallel interfaces of the dimer and the trimer showed different interaction types; again, this interface shows its adaptation capacity with a new set of interactions that permits the tetramerization through a parallel interface.

Anti-parallel Aggregation. As previously mentioned, two complexes were formed during the 2 μs simulations. In these complexes four cases of anti-parallel aggregation (APA) were detected: two during the dimerization, one during the trimerization and the last during the tetramerization. The two dimers are: AS-AH and N-D, they respectively initiate CIFrag2 and CIIFrag1 formation. The trimer BG-AE-E (CIIFrag2) and tetramer BH-BC-AX-AG (CIFrag1) are the extensions of a parallel dimer and of a full parallel trimer.

APA is very different from the parallel one. As previously seen, the PA mechanisms varied in function of the available residues and the π-stacking of Y5. Contrariwise, in all antiparallel interfaces listed in the complexes, the aggregates were formed with a unique mechanism. Indeed, the detected mechanism was the same for initiation and

for prolongation that are respectively illustrated here by the dimer N-D* and the trimer BG-AE/E* (Fig. 4a), where asterisk* symbolizes the last aggregating peptide or in di-
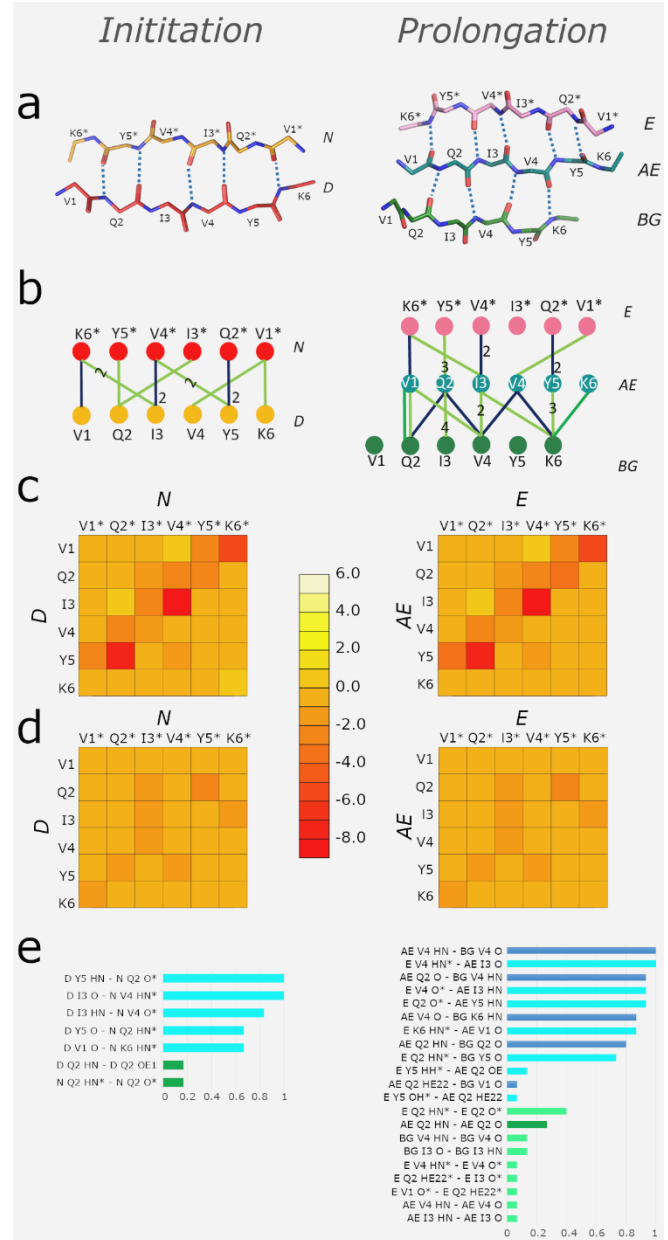

Figure 4. Anti-parallel mechanism. a) Representative structures. b) Interaction network of the representative structures. Blue and green lines represent Hbonds between main chains and van der Waals contacts, respectively. The number on/near the line represents the interaction type number. c) Interaction energies matrix per residue. d) Van der Waals interaction energy matrix per residue. e) Volume occupancy histograms of the detected Hbonds (cyan and dark blue bars annotate the inter-monomer Hbonds of previous and new interfaces, respec-

mers the peptide hosting last peptide in the next aggregation step. In the case of N-D* and BG-AE/E*, D* and E*

were the peptides engaged in the next step of the aggregation. However, the interaction network in N-D* was inversed (Fig. 4a), the dimer N peptide acted like peptide E* and the D* peptide like AE. For more clarity, the asterisk* symbol in the following description of the anti-parallel mechanism will be relative to the BG-AE/E* and N-D* was modified to N*-D.

The interaction network analyses of the representative structures showed that the anti-parallel interfaces are stabilized by five Hbonds occurring between main chain atoms (Fig. 4b): I3(O)-V4*(HN), Y5(HN)-Q2*(O), I3(HN)-V4*(O), Y5(O)-Q2*(HN) and V1(O)-K6*(HN). Five detected principal interaction energies (Fig. 3c) greater than 3 kcal/mol are Y5-V1*, Y5-Q2*, I3-V4*, Q2-Y5* and V1-K6*. Compared to the PA, there are fewer interactions, but these interactions are stronger. Indeed, the five detected interactions showed an average total interaction energies of -3.04 ± 1.18 kcal/mol, -7.36 ± 1.62 kcal/mol, -8.18 ± 1.05 kcal/mol, -2.98 ± 1.48 kcal/mol, and -5.23 ± 1.64 kcal/mol, respectively. Comparing with the interaction network of the representative structures (Fig. 4b), Q2-Y5* is the only interaction that cannot be explained by the Hbond formation but its origin is in van der Waals contacts (Fig. 4d). In conclusion, the same residues from both peptides participate in all anti-parallel interface described here: Q2, I3, V4 and Y5 and they fairly share the central role in the aggregation; they represent 18.22%, 19.56%, 19.07%, and 20.24% of interaction energies, respectively.

The average occupancy volumes (on all four anti-parallel interfaces) of the five principal Hbonds (Fig. 4e): I3(O)-V4*(HN), Y5(HN)-Q2*(O), I3(HN)-V4*(O), Y5(O)-Q2*(HN) and V1(O)-K6*(HN) were 0.964, 0.931, 0.858, 0.660 and 0.765, respectively. Moreover, three other Hbonds between side-chains were detected, two in the trimer (Y5*(HH) – Q2 (OE) and Y5*(OH) – Q2 HE22) and one in the tetramer (Q2*(HE21) - Y5 (OH)) but they are not taken into account because of their weak volume occupancy score (<0.2). To finish, for all antiparallel interfaces sixteen intra-chains Hbonds were also detected with volume occupancy score <0.5.

The five inter-chain Hbonds were present on the representative structures of each aggregation step (Fig. 4b and S2-3) except Y5(O)-Q2*(HN) in AS-AH dimer (Fig. S3). They can be divided into two groups. I3(O)-V4*(HN) and Y5(HN)-Q2*(O) presented maximal volume occupancies close to 1.000 and represent the first group. These two Hbonds were present at each aggregation step (Fig. S2). It

the same time. However, the score of I3(HN)-V4*(O) increased all along the aggregation until reaching 1.000. Consequently, this Hbond happens before V1(O)-K6*(HN) that remains stable.

In conclusion in APA, the two first Hbonds to be created are I3(O)-V4*(HN) and Y5(HN)-Q2*(O), the third is I3(HN)-V4*(O), due to its stabilization during the aggregation process, the fourth V1(O)-K6*(HN) and the last Y5(O)-Q2*(HN). These five Hbonds can be considered essential

is not possible to clearly distinguish the formation order in this first group, but thanks to their volume occupancy, it can be concluded that they are the two first to be created. The second group brings together intermediate volume occupancy Hbonds with I3(HN)-V4*(O), Y5(O)-Q2*(HN) and V1(O)-K6*(HN). By looking at the volume occupancy of the Hbonds during the next steps of the aggregation, it
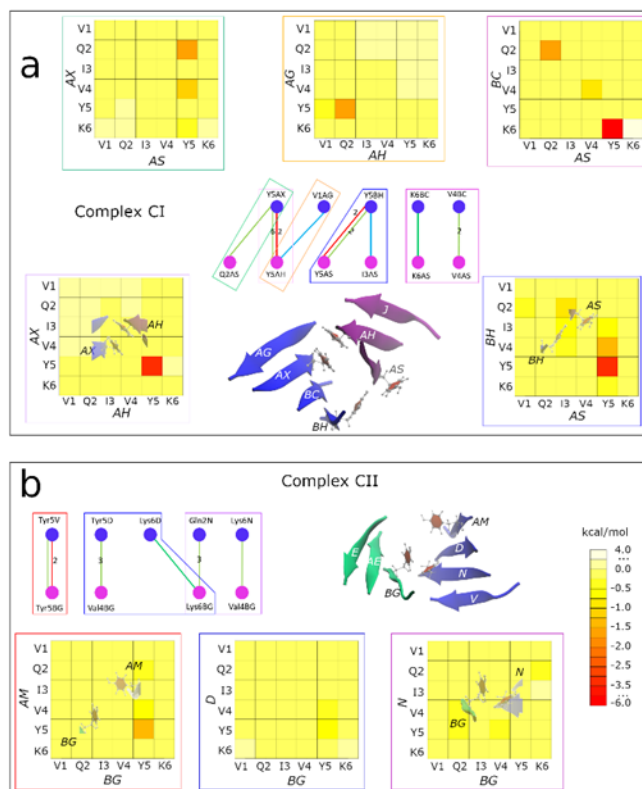

Figure 5. Complexation. Interaction network and structures of the representative structures of CI complex (a) and CII one (b). In the interaction networks: dark green, green and red lines represent the van der Waals contacts between main and side chains, the van der Waals contacts between side chains and π-stacking, respectively. The number indicated on/near the line represents the number of this type of interaction. In the squares are the interaction energy matrix per residue for the peptides involved in the complexation.

can be noticed that Y5(O)-Q2*(HN) is not very stable along the trajectory. In conclusion, it is the last Hbond to be created and it seems not be crucial in the aggregation process. The two Hbonds I3(HN)-V4*(O) and V1(O)-K6*(HN) were created at

in the APA formation and stabilization. The MM-PBSA analysis along the trajectory confirmed that the residues involved in H-bonds contribute mostly to the dimer binding free energy, I3, Y5, V4* and Q2* (Fig 3). The Alascan revealed that the stabilizing role is played by Q2*, V4*, Y5*, K6* and V4 in the initiation while in the prolongation by K6*, Y5*, I3* and V4 (Fig S9).

Complexation. A complexation is a process that describes the aggregation of two fragments. In fact, two β-

sheets - located in two different planes - aggregate face to face. Figures 4a-b describe the analysis of the complexation through the CI and CII complexes, respectively. The interaction network between fragments calculated from the representative structure of two complexes are showed in figures 5a-b.

First, it should be noted that the CI complex (Movie 1) is formed earlier than the CII one, CII was created very shortly before the end of the simulations (at 1.97 μs). Furthermore, another important difference between the two complexes is their starting point. The CI complex came from the aggregation of a tetramer and a trimer, while the CII complex from the aggregation of a pentamer with a trimer. Interestingly, the growing from tetramer to pentamer of CIIFrag1 has happened on the other side than the previous growing tetramer (Fig. 1b bottom). Moreover, during the CII complexation one hexapeptide of pentamer was ejected: the hexapeptide at the aggregate extremity which was added last during the tetramerization (Movie 2). During the analysis of this tetramer carried out previously (Fig. 2a mechanism 3), a π-stacking interaction between Y5 of neighboring peptides was detected.

In the CI complex twelve van der Waals contacts, four π-stacking and two Hbonds have been revealed. While in the CII complex only van der Waals (nine contacts) and π-stacking (two) interactions were observed. The comparison of interaction networks of the two complexes shows that only one van der Waals interaction (K6/K6) is common between them. The inconstancy of the other van der Waals interaction shows that they are not the key element in the complexation process.

The same conclusion can be drawn about the Hbonds. Two Hbonds were detected between the two fragments in the CI complex and none in the CII complex. Furthermore, volume occupancy analysis of the CI complex Hbonds did not confirm their existence more than the half of the trajectory. Indeed, the most stable Hbond (between a main and a side chain atom) was found between two Y5 (AS/BH) with a volume occupancy of 0.500. Interestingly, between these two Y5, π-stacking interactions were also identified; a second π-stacking interaction was seen between two other Y5 residues (AX/AH). The CII complex also presents this kind of Y5 π-stacking between BG and V chains. This π-stacking between the two Y5 during complexation was also seen in the replica 2 between M and AI, then AB/A in Fig S10. The representative structures highlighted also the involvement of π-stacking interactions and van der Waals K6-K6 interactions. To evaluate these two phenomena all along the complexes' trajectories, interaction energies indicating the contribution of each residue were computed. The results are present around the complexes' interaction network in a colored squared relative to the studied interface (Fig. 5). Then, for each interface the interaction energy shared by chains from different fragments was computed (matrices in colored squares in Fig. 5). The green, orange and purple squares of the figure 5a concern the interfaces

in CI complex where no π-stacking was identified. The detected energetic contribution was very different in the three interfaces, but two residues stand out: K6 and Y5. K6 is very ambivalent: it presents a strong interaction energy with Y5, but its interaction with K6 of the other chain is completely energetically unfavorable. This can be observed through positive interaction energies of +1.5 ± 1.9 kcal/mol between K6 of AS and K6 of AX, +3.2 ± 1.61 kcal/mol between K6 of AS/K6 of BC while the interaction energy between Y5 of AS and K6 of BC is negative (favorable) of -5.1 ± 2.8 kcal/mol (Fig. 4a). In conclusion, on all five interfaces in CI complex, K6 shows a rather overall unfavorable interaction energy. The same phenomenon was observed also in the CII complex, where K6 showed a new an unfavorable average interaction energy on all interfaces (Fig. 5b).

On the opposite, no ambiguity was detected for Y5. In the interfaces where π-stacking was observed, Y5 represented 97.47% (CI: AS/AX), 52.64% (CI: AH/AS), 73.57% (CII: BG/V) of the total interaction energy between the two chains. Even if the network interaction analysis did not identify π-stacking between D and BG hexapetides in the CII, Y5 represented 74.50% of the interaction between these two chains. This result was corroborated by the hydrophobic contact analysis revelling Y5-Y5 interaction as crucial in complex formation (Fig. S10 A-B). Moreover, on the atomic structures of the CII complex (Fig. 5b), the possibility of π-stacking is corroborated by their proximity. Interestingly, the Y5 of V stacked with Y5 of T in tetramer and during the complexation Y5 of V changed to interact now strongly with Y5 of B. In the CII complex, Y5 residues are responsible for 76.54% of all interaction energy between the two fragments, among this interaction energy: 59.92% is relative to mutual Y5-Y5 interactions. This allows to think that the π-stacking between two fragments is the more stabilizing phenomenon and that π-stacking plays a key role in the complexation.

CONCLUSIONS

This study highlights two aggregation types: PA and APA mechanisms. Here, seven interfaces have been investigated, revealing four different sets of interactions: three from the PA and one from the APA.

The same four Hbonds were found on all AP interfaces: I3(O) - V4(HN), V4(O) – I3(HN), Q2(O) – Y5(HN) and V1(O) – K6(HN). The other detected Hbond, Q2(HN) – Y5(OH) was not present for all interfaces and is not considered as essential in the aggregation, unlike the first four. This aggregation process is based on interactions between four residues: I3-V4 (3/3 cases) and Q2-Y5 (2/3 cases). Unfortunately, the formation order of each Hbond is not easy to deduce. However, the first Hbond formed was always ones between I3-V4. Consequently, I3(O) – V4(HN) or

$V_4(O) - I_3(HN)$ is the starting point of an APA. This aggregation process needs to create the exact interactions between the residues along the anti-parallel interfaces and it is not able to adapt. Liu[21] reported three Hbonds during anti-parallel dimerization of tau hexapeptide in their study: $V_1(O) - V_4(HN)$, $Q_2(O) - I_3(HN)$ and $V_4(O) - V_1(NH)$. The difference with our results most probably comes from the different treatment of the terminal ends; in our system the N and C-terminal were protonated whereas Liu and co-workers during their simulations capped the terminal charges. Recently, Arya *et al.*[22] in their study have shown the PHF6 aggregation is highly dependent on the charges on its termini.

The PA observed in simulations revealed a very different behaviour compared to APA: not one but three different mechanisms were detected. The PA mechanism changed depending on the environment. Indeed, when two monomers assembled to form a dimer, they had no limit to perform interactions. However, when a monomer bound to a pre-existing aggregate, it must take into account residue's availability. Consequently, a preferred mechanism with four Hbonds was identified during dimer formation and one trimerization (from an anti-parallel dimer allowing these interactions): $V_4(HN)-Q_2(O)$, $V_4(O)-V_4(HN)$, $K_6(HN)-V_4(O)$ and $Q_2(O)-Q_2(HN)$, following their formation order. The last Hbond is not considered crucial in aggregation, unlike the three others. The importance of V_4 is obvious as it is involved in the three crucial Hbonds. By using replica exchange molecular dynamics Ganguly and coll.[23] studied the Hbonds between two parallels monomers during the dimerization. Their work did not provide details at the atomic level but identified the interacting residues revealing four significant interactions: V_4-Q_2, K_6-V_4, Q_2-Q_2, and V_4-V_4, the same that we found in our study. Two other Hbonds were detected but at low levels (K_6-Q_2, K_6-K_6); they are not expected to be crucial in aggregation.

The second PA mechanism was observed during trimerization and was principally composed of the following Hbonds: $I_3(O)-I_3(HN)$, $Y_5(HN)-I_3(O)$, $I_3(HN)-V_1(O)$ and $Y_5(O)-Y_5(HN)$. The three first Hbonds initiate the aggregation, and only then the fourth $Y_5(O)-Y_5(HN)$ is formed. Nevertheless, the last Hbond is not crucial in the aggregation process. Unlike the first preferential mechanism, V_4 cannot be involved in aggregation because it already interacts with another monomer in the previously formed parallel interface. As V_4 is not available, the aggregation process can rely on the I_3 and Y_5 residues, involved in three and two crucial interactions, respectively.

The third mechanism of PA occurs during the tetramerization and is composed of three Hbonds: $V_4(O)-Y_5(HN)$, $V_4(HN)-I_3(O)$ and $Q_2(O)-I_3(HN)$. The formation order cannot be clearly defined for the starting point, the two first Hbonds are created at the beginning of the aggregation, followed by the last one. V_4 plays a new crucial role, but it interacts with $Y_5(HN)$ and $I_3(O)$ instead of $V_4(HN)$ and $Q_2(O)$ as in mechanism 1. This change indicates a shift in the monomers' alignment. So, the PA shows the possibility of face-to-face aggregation but also a one-residue shift.

No meaningful hydrophobic contacts were found for all these different mechanisms (Fig. S9 D), thus hydrophobic contacts do not have a key role in the aggregation of dimers or trimers.

These results explain why the parallel interface is preferred in tau protein aggregation, even if it is energetically less favorable (see Table S1). The PA mechanism can adapt considering the previous interactions. It is not the case for APA, which requires particular residues to occur without adaptation possibilities. It is well described by the distributions of parallel and anti-parallel interfaces. Indeed, dimers and trimers show a fairly equal division. Dimers describe 54.5% of the parallel interface, while the most frequent trimer (57.19%) conformation is composed of one parallel and one anti-parallel interfaces. This equal distribution can be explained by the fact that it is the beginning of the aggregation: the required conditions to build an anti-parallel interface are verified. However, with the aggregate elongation, these conditions are more challenging to obtain, which explains why 97.29% of the hexamers are composed of three parallel interfaces and two anti-parallel interfaces. The aggregation of the two complexes studied here describes seven parallel interfaces and five anti-parallel interfaces. Considering these observations, it is essential to note that tau protein aggregation does not rely on a single mechanism but on a set of them. Tau protein aggregation is achieved by PA but also needs the anti-parallel mechanism to occur. Consequently, it is crucial to consider all mechanisms to design an efficient aggregation disruptor.

The tau protein aggregates in two steps. The first is the aggregate forming action to form a β-sheet. The second is the face to face β-sheets' complexation. In the simulations, the complexation did not occur between aggregates smaller than a trimer. No stable Hbond or van der Walls interaction were found to explain the complexation of two aggregates. However, we noted that each complex forms at least one π-π interaction between the Y_5 (Y_{310}) aromatic cycles. Y_5 was responsible for 61.26% and 76.54% of the interaction energy between the two fragments of the CI and CII complexes, respectively: it is crucial in the complexation. Nonetheless, hydrophobic contacts are not responsible for the initiation of the aggregation but helped to stabilize the complex. Indeed, thanks to fluorescence, circular dichroism and electron microscopy, Nishura[24] highlighted experimentally a crucial role of Y_5 during PHF6 aggregation.

Consequently, to prevent the complexation phase, an antiaggregating agent should have an aromatic moiety able to disrupt π-π interactions stabilizing the complexes. Wan[25] investigated the impact of norepinephrine (which has an aromatic moiety) on the tau aggregation by molecular dynamics, and norepinephrine successfully disrupted

a part of the tau protein. The same phenomenon was observed with the fisetin (with two aromatic cycles)[26] which prevented β-sheet formation, both experimentally and theoretically. Other molecules have showed also experimentally inhibition of the tau aggregation as proanthocyanidin,[27] purpurin,[28] naphthoquinone-tryptophan[29] and naphthoquinone-dopamine hybrids,[30] 1-(phenylsulfonyl)-1H-indole-based multifunctional ligands.[31] All these compounds had aromatic cycles in their structures.

METHODS

Molecular Modelling and Dynamics. Modelling of PHF6 was realized from the crystallographic structure 2ON9[32] at the resolution of 1.51Å. This peptide was repeated 62 times and each of these 62 peptides were manually disposed at least 10 Å away from each other. We performed 2.00 μs molecular dynamics simulations on this system. All of the molecular dynamics simulations were performed using NAMD 2.12[33] with the Charmm36 force field[34]. To simulate the aqueous solvent environment, each system was surrounded by a cubic box (dimensions: 144.539 Å x 144.539 Å x 144.539 Å) of TIP3P explicit water model[35], and 0.150 M NaCl was added to the system using the CHARMM-GUI solvator[36]. We protonated the termini ends in the Charmm-gui preparation process. The Van der Waals interactions were truncated using a force switching function between 10 and 12 Å, and the Particle Mesh Ewald (PME)[37] method was used to compute the long-range electrostatic interactions. The SHAKE[38] algorithm was applied to restrain all bonds involving hydrogen atoms. The systems underwent energy minimization for 10,000 steps. The systems were afterward heated up to 310K, and the system was temperature-equilibrated during 50 ps via heating reassignment under NVT (canonical ensemble) conditions. Finally, the system ran freely for 2.00 μs under NPT (isothermal-isobaric ensemble) conditions. Langevin dynamics with a damping coefficient of 1 ps-1 was used to maintain the system temperature, and the Nosé-Hoover Langevin[39,40] piston method was used to control the pressure at 1 atm.

The end of the stabilization of the system was detected to 104 ns. To check the timeline of the aggregation process, 100 ns molecular dynamics simulations were computed from the end of stabilization of the system using the same parameters than the ones used for the trajectory of 2 μs but with different velocities. Analysis of this trajectory are provided in Fig S9. The present work was performed using computing resources of CRIANN (criann.fr, Normandy, France).

Analysis. The Charmm c40b2 version was used for all analysis requiring.

Detection of the aggregation rate. To detect the aggregation rate of the 62 peptides, a Charmm home-script was written. The first step consists in the detection of the neighboring peptides for each peptide. First, for each peptide, all the peptides surrounding it in a radius of 4.00Å are listed. Then, the two monomers can aggregate in different ways: ten cases were taken into consideration. Fig. S6 describes the ten different aggregation types. To be considered as aggregated, the distances between the Cα atoms of the penultimate residues before each extremity (red in Fig. S6) must both be less than 6.00Å. Finally, the lists containing common peptides are concatenated. This approach allows to know exactly the aggregation rate and more accurately the number of monomers, dimers, trimers, tetramers, pentamers and hexamers by frame (Fig. 1c-d).

Detection of the aggregation type. The previous script concerning the aggregation rate gives the number of dimers, trimers, tetramers, pentamers and hexamers, but also their exact chain composition. The chains organization is established by measuring the distance between each chain: the longest distance gives the names of the two peptides located at the extremity of the aggregates. The next step is to find the two extremities' neighbors and continue until the last peptide is located. Finally, the order of the chains in the aggregate is known, it is possible to evaluate the parallel or anti-parallel feature of the different interfaces and obtain a final score of parallel and anti-parallel interfaces.

Hydrogen bonds analysis. The number of Hbonds by frame (Fig. 1d) was made with the MDAnalysis python library[41,42]. The identification of Hbonds with the exact involved atoms and their volume occupancies (Fig. 2, 4) were computed with a Charmm script involving the coor hbond command.

Distances between two centers of mass. The coordinates of the centers of mass are evaluated from the gyration radius computed with coor gyr command of Charmm. Then, the distances between the two center of mass are calculated with a Charmm home-script.

Energies. Three types are computed all along this work: the total potential energy of the system considered (to detect the representative structures of the system), the total interaction energy and the Van der Waals interaction energy to evaluate the contribution of each residue in the interactions and the role of Van der Waals contacts. These energies were computed with the inte command of Charmm for each aggregation step (Fig. S4-5).

MM-PBSA and Computational AlaScan. Residue-wise decomposition of the binding free energy as well as AlaScan prediction was performed within the MM-PBSA approach using CHARMM home-made scripts (Fig 3, Fig S8 and S9).

Ab initio analysis. energy calculations reported in the present study were carried out using the GAUSSIAN 98 software[43] at HF/3-21G level.

Determination of the representative structures. Determining the representative structures of the different dimers, trimers, tetramers, pentamers and complexation follows three steps. The first is the selection of the frames that represent the system considered at a specific state. In order to select them, the distances between the centers of mass of the last peptide that bound to the system and the

other peptides of the system are computed. For example, the interval of frames concerning a trimer is from the peptide binding (to form a trimer) until the binding of a new peptide (tetramerization). The detection of the frames for each state is described in the Fig. S7. The second step consists in the total potential energy calculations for each system at each frame considered. The third step selects the structure at the frame where its potential energy is the lowest, this is the representative structure. Twenty and eighteen representative structures are determined for the complexes I and II, respectively.

Interaction Network. The interaction network is obtained by submitting the representative structures previously determined to the webserver RING[44]. This last gives access to structural data like the Hbonds, the Van der Waals interactions and the π-stacking involved in the aggregation process. Their processing and their visualization were realized with Cytoscape[45]. This study has been done on all the representative structures at each aggregation step (Fig. S2).

Hydrophobic contacts. Hydrophobic contacts were computed with PyContact[46] tool (version 1.0.4). PyContact was applied on the trajectories of each system (dimer, trimer, tetramer, complex). Detailed results available in the Supplementary (S10).

ASSOCIATED CONTENT

Supporting Information (Fig. S1) Parallel mechanisms. (Fig. S2) Interaction network of each interface during the different steps of the aggregation. (Fig. S3) Anti-parallel mechanism. (Fig. S4) Total interaction energy matrix between each residue for all the interfaces of the two complexes during the different steps of the aggregation. (Fig. S5) Van der Waals energy matrix between each residue for all the interfaces of the two complexes during the different steps of the aggregation. (Fig. S6) Different types of parallel and anti-parallel interfaces. (Fig. S7) Timelines of the different steps of the aggregation. (Fig S8) (Fig. S9) Computational Alanine Scan on the dimer representative structures. (Fig. S10) Hydrophobic contacts. (Fig. S11) Replica 2 Results. (Fig. S12) Replica 3 Results. (Fig. S13) Replica 4 Results. (Fig. S14) Evolution of the MMPBSA per residue along the trajectory of the interface BH/BC. (Fig. S15) Evolution of the MMPBSA per residue along the trajectory of the interface BC/AX. (Fig. S16) Evolution of the MMPBSA per residue along the trajectory of the interface AH/J. (Fig. S17) Evolution of the MMPBSA per residue along the trajectory of the interface R/AM. (Fig. S18) Evolution of the MMPBSA per residue along the trajectory of the interface AE/E (Fig. S19)) Evolution of the MMPBSA per residue along the trajectory of the interface N/D. (Table S1) Energy characteristic of the parallel and anti-parallel representative dimers calculated using the Molecular Mechanics approaches (MM-PBSA) as well as the ab initio ones. (Movie 1) Formation of the complex CI. (Movie 2) Formation of the complex CII. (Movie 3) First complexation in replica 2 simulations. (Movie 4) Second complexation in replica 2 simulations. (Movie 5) Complexation in replica 3 simulations.

AUTHOR INFORMATION

Corresponding Author

Jana Sopkova-de Oliveira Santos - CERMN (Centre d'Etudes et de Recherche sur le Médicament de Normandie), Université de Caen Normandie, UNICAEN, 1 Boulevard Henri Becquerel, F-14032 Caen, France ; https://orcid.org/ 0000-0002-4829-8120 ; jana.sopkova@unicaen.fr ; +33 2 31 56 68 21

Anne Sophie Voisin-Chiret - CERMN (Centre d'Etudes et de Recherche sur le Médicament de Normandie), Université de Caen Normandie, UNICAEN, 1 Boulevard Henri Becquerel, F-14032 Caen, France ; https://orcid.org/ 0000-0001-5564-2244 ; anne-sophie.voisin@unicaen.fr ; +33 2 31 56 68 04

Present Addresses

†Marco Catto - Department of Pharmacy–Pharmaceutical Sciences, University of Bari Aldo Moro, Via E. Orabona 4, 70125 Bari, Italy.

Author Contributions

JSOS, ASVC and MC conceived and designed the research. JG performed molecular dynamics. CF analyzes molecular dynamics simulations. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

ABBREVIATIONS

NFTs, Neurofibrillary Tangles; PHF, Paired Helical Filaments; MTBD, Microtubule-Binding Domains; MT, microtubule; CI, Complex I; CII, Complex II; PA, parallel aggregation; APA, anti-parallel aggregation; PME, Particle Mesh Ewald.

REFERENCES

(1)   Avila, J. Tau Aggregation into Fibrillar Polymers: Taupathies. FEBS Letters 2000, 476 (1–2), 89–92. https://doi.org/10.1016/S0014-5793(00)01676-8.

(2)   Kovacs, G. G. Chapter 25 - Tauopathies. In Handbook of Clinical Neurology; Kovacs, G. G., Alafuzoff, I., Eds.; Neuropathology; Elsevier, 2018; Vol. 145, pp 355–368. https://doi.org/10.1016/B978-0-12-802395-2.00025-0.

(3)   Silva, M. C.; Haggarty, S. J. Tauopathies: Deciphering Disease Mechanisms to Develop Effective Therapies. Int J Mol Sci 2020, 21 (23), E8948. https://doi.org/10.3390/ijms21238948.

(4)   Tan, C.-C.; Zhang, X.-Y.; Tan, L.; Yu, J.-T. Tauopathies: Mechanisms and Therapeutic Strategies. J Alzheimers Dis 2018, 61 (2), 487–508. https://doi.org/10.3233/JAD-170187.

(5)   Guo, T.; Noble, W.; Hanger, D. P. Roles of Tau Protein in Health and Disease. Acta Neuropathol 2017, 133 (5), 665–704. https://doi.org/10.1007/s00401-017-1707-9.

(6)   Wang, Y.; Mandelkow, E. Tau in Physiology and Pathology. Nat Rev Neurosci 2016, 17 (1), 5–21. https://doi.org/10.1038/nrn.2015.1.

(7)   Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. Nature 2017, 547 (7662), 185–190. https://doi.org/10.1038/nature23002.

(8)   Nizynski, B.; Dzwolak, W.; Nieznanski, K. Amyloidogenesis of Tau Protein. Protein Sci 2017, 26 (11), 2126–2150. https://doi.org/10.1002/pro.3275.

(9)   Naseri, N. N.; Wang, H.; Guo, J.; Sharma, M.; Luo, W. The Complexity of Tau in Alzheimer's Disease. Neurosci Lett 2019, 705, 183–194. https://doi.org/10.1016/j.neulet.2019.04.022.

(10)   Qi, H.; Despres, C.; Prabakaran, S.; Cantrelle, F.-X.; Chambraud, B.; Gunawardena, J.; Lippens, G.; Smet-Nocca, C.; Landrieu, I. The Study of Posttranslational Modifications of Tau Protein by Nuclear Magnetic Resonance Spectroscopy: Phosphorylation of Tau Protein by ERK2 Recombinant Kinase and Rat Brain Extract, and Acetylation by Recombinant Creb-Binding Protein. Methods Mol Biol 2017, 1523, 179–213. https://doi.org/10.1007/978-1-4939-6598-4_11.

(11)   Goedert, M.; Spillantini, M. G.; Jakes, R.; Rutherford, D.; Crowther, R. A. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. Neuron 1989, 3 (4), 519–526. https://doi.org/10.1016/0896-6273(89)90210-9.

(12)   Crowther, T.; Goedert, M.; Wischik, C. M. The Repeat Region of Microtubule-Associated Protein Tau Forms Part of the Core of the Paired Helical Filament of Alzheimer's Disease. Annals of Medicine 1989, 21 (2), 127–132. https://doi.org/10.3109/07853898909149199.

(13)   Kadavath, H.; Jaremko, M.; Jaremko, Ł.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. Folding of the Tau Protein on Microtubules. Angewandte Chemie International Edition 2015, 54 (35), 10347–10351. https://doi.org/10.1002/anie.201501714.

(14)   Rösler, T. W.; Tayaranian Marvian, A.; Brendel, M.; Nykänen, N.-P.; Höllerhage, M.; Schwarz, S. C.; Hopfner, F.; Koeglsperger, T.; Respondek, G.; Schweyer, K.; Levin, J.; Villemagne, V. L.; Barthel, H.; Sabri, O.; Müller, U.; Meissner, W. G.; Kovacs, G. G.; Höglinger, G. U. Four-Repeat Tauopathies. Prog Neurobiol 2019, 180, 101644. https://doi.org/10.1016/j.pneurobio.2019.101644.

(15)   Tomoo, K.; Yao, T.-M.; Minoura, K.; Hiraoka, S.; Sumida, M.; Taniguchi, T.; Ishida, T. Possible Role of Each Repeat Structure of the Microtubule-Binding Domain of the Tau Protein in in Vitro Aggregation. J Biochem 2005, 138 (4), 413–423. https://doi.org/10.1093/jb/mvi142.

(16)   Kolarova, M.; García-Sierra, F.; Bartos, A.; Ricny, J.; Ripova, D. Structure and Pathology of Tau Protein in Alzheimer Disease. Int J Alzheimers Dis 2012, 2012, 731526. https://doi.org/10.1155/2012/731526.

(17)   von Bergen, M.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E.-M.; Mandelkow, E. Assembly of τ Protein into Alzheimer Paired Helical Filaments Depends on a Local Sequence Motif (306VQIVYK311) Forming β Structure. PNAS 2000, 97 (10), 5129–5134.

(18)   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 (2), 170–176. https://doi.org/10.1038/nchem.2889.

(19)   Li, W.; Lee, V. M.-Y. Characterization of Two VQIXXK Motifs for Tau Fibrillization in Vitro. Biochemistry 2006, 45 (51), 15692–15701. https://doi.org/10.1021/bi061422+.

(20)   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 (1), 2493. https://doi.org/10.1038/s41467-019-10355-1.

(21)   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 (12), 4810–4823. https://doi.org/10.1021/acschemneuro.9b00488.

(22)   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 Amyloidogenix Tau Fragment Modulates Its Fibrillation Propensity. . J Phys Chem B 2020, 124, 8772-8783. https://doi.org/10.1021/acs.jpcb.0c05768

(23)   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 (13), 4582–4593. https://doi.org/10.1021/acs.jpcb.5b00175.

(24)   Nishiura, C.; Takeuchi, K.; Minoura, K.; Sumida, M.; Taniguchi, T.; Tomoo, K.; Ishida, T. Importance of Tyr310 Residue in the Third Repeat of Microtubule Binding Domain for Filament Formation of Tau Protein. J Biochem 2010, 147 (3), 405–414. https://doi.org/10.1093/jb/mvp181.

(25)   Wan, J.; Gong, Y.; Xu, Z.; Dong, X.; Wei, G.; Zhang, Q. Molecular Dynamics Simulations Reveal the Destabilization Mechanism of Alzheimer's Disease-Related Tau R3-R4 Protofilament by Norepinephrine. Biophys Chem 2021, 271, 106541. https://doi.org/10.1016/j.bpc.2021.106541.

(26)   Xiao, S.; Lu, Y.; Wu, Q.; Yang, J.; Chen, J.; Zhong, S.; Eliezer, D.; Tan, Q.; Wu, C. Fisetin Inhibits Tau Aggregation by Interacting with the Protein and Preventing the Formation of β-Strands. Int J Biol Macromol 2021, 178, 381–393. https://doi.org/10.1016/j.ijbiomac.2021.02.210.

(27)   Li, Q.; Xiong, C.; Liu, H.; Ge, H.; Yao, X.; Liu, H. Computational Insights Into the Inhibition Mechanism of Proanthocyanidin B2 on Tau Hexapeptide (PHF6) Oligomer. Front Chem 2021, 9, 666043. https://doi.org/10.3389/fchem.2021.666043.

(28)   Viswanathan, G. K.; Shwartz, D.; Losev, Y.; Arad, E.; Shemesh, C.; Pichinuk, E.; Engel, H.; Raveh, A.; Jelinek, R.; Cooper, I.; Gosselet, F.; Gazit, E.; Segal, D. Purpurin Modulates Tau-Derived VQIVYK Fibrillization and Ameliorates Alzheimer's Disease-like Symptoms in Animal Model. Cell Mol Life Sci 2020, 77 (14), 2795–2813. https://doi.org/10.1007/s00018-019-03312-0.

(29)   Frenkel-Pinter, M.; Tal, S.; Scherzer-Attali, R.; Abu-Hussien, M.; Alyagor, I.; Eisenbaum, T.; Gazit, E.; Segal, D. Naphthoquinone-Tryptophan Hybrid Inhibits Aggregation of the Tau-

Derived Peptide PHF6 and Reduces Neurotoxicity. J Alzheimers Dis 2016, 51 (1), 165–178. https://doi.org/10.3233/JAD-150927.

(30)   Paul, A.; Viswanathan, G. K.; Huber, A.; Arad, E.; Engel, H.; Jelinek, R.; Gazit, E.; Segal, D. Inhibition of Tau Amyloid Formation and Disruption of Its Preformed Fibrils by Naphthoquinone-Dopamine Hybrid. FEBS J 2021, 288 (14), 4267–4290. https://doi.org/10.1111/febs.15741.

(31)   Wichur, T.; Pasieka, A.; Godyń, J.; Panek, D.; Góral, I.; Latacz, G.; Honkisz-Orzechowska, E.; Bucki, A.; Siwek, A.; Głuch-Lutwin, M.; Knez, D.; Brazzolotto, X.; Gobec, S.; Kołaczkowski, M.; Sabate, R.; Malawska, B.; Więckowska, A. Discovery of 1-(Phenylsulfonyl)-1H-Indole-Based Multifunctional Ligands Targeting Cholinesterases and 5-HT6 Receptor with Anti-Aggregation Properties against Amyloid-Beta and Tau. Eur J Med Chem 2021, 225, 113783. https://doi.org/10.1016/j.ejmech.2021.113783.

(32)   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.; Riekel, C.; Eisenberg, D. Atomic Structures of Amyloid Cross-Beta Spines Reveal Varied Steric Zippers. Nature 2007, 447 (7143), 453–457. https://doi.org/10.1038/nature05695.

(33)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. Journal of Computational Chemistry 2005, 26 (16), 1781–1802. https://doi.org/10.1002/jcc.20289.

(34)   Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. Journal of Computational Chemistry 2013, 34 (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(35)   Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J. Phys. Chem. A 2001, 105 (43), 9954–9960. https://doi.org/10.1021/jp003020w.

(36)   Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 2016, 12 (1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935.

(37)   Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. J. Chem. Phys. 1993, 98 (12), 10089–10092. https://doi.org/10.1063/1.464397.

(38)   Forester, T. R.; Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. Journal of Computational Chemistry 1998, 19 (1), 102–111. https://doi.org/10.1002/(SICI)1096-987X(19980115)19:1<102::AID-JCC9>3.0.CO;2-T.

(39)   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. The Journal of Chemical Physics 1998, 103 (11), 4613. https://doi.org/10.1063/1.470648.

(40)   Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. J. Chem. Phys. 1994, 101 (5), 4177–4189. https://doi.org/10.1063/1.467468.

(41)   Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. J Comput Chem 2011, 32 (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(42)   Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M. N.; Seyler, S.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations; 2016. https://doi.org/10.25080/MAJORA-629E541A-00E.

(43)   Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.

(44)   Piovesan, D.; Minervini, G.; Tosatto, S. C. E. The RING 2.0 Web Server for High Quality Residue Interaction Networks. Nucleic Acids Research 2016, 44 (W1), W367–W374. https://doi.org/10.1093/nar/gkw315.

(45)   Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003, 13 (11), 2498–2504. https://doi.org/10.1101/gr.1239303

(46)   Scheurer, M.; Rodenkirch, P.; Siggel, M.; Bernardi, R. C.; Schulten, K.; Tajkhorshid, E.; Rudack, T. PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations. *Biophysical Journal* **2018**, *114* (3), 577–583. https://doi.org/10.1016/j.bpj.2017.12.003.

For Table of Contents Only