THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

A Kinetic Approach to the Sequence−Aggregation Relationship in Disease-Related Protein Assembly

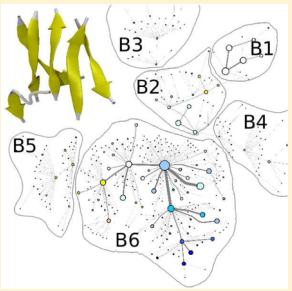
Bogdan Barz,[†] David J. Wales,[‡] and Birgit Strodel*[,†,§]

[†]Forschungszentrum Jülich GmbH Institute of Complex Systems: Structural Biochemistry (ICS-6), 52425 Jülich, Germany

[‡]University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, U.K.

[§]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

ABSTRACT: It is generally accepted that oligomers of aggregating proteins play an important role in the onset of neurodegenerative diseases. While *in silico* aggregation studies of full length amyloidogenic proteins are computationally expensive, the assembly of short protein fragments derived from these proteins with similar aggregating properties has been extensively studied. In the present work, molecular dynamics simulations are performed to follow peptide aggregation on the microsecond time scale. By defining aggregation states, we identify transition networks, disconnectivity graphs, and first passage time distributions to describe the kinetics of the assembly process. This approach unravels differences in the aggregation into hexamers of two peptides with different primary structures. The first is GNNQQNY, a hydrophilic fragment from the prion protein Sup35, and the second is KLVFFAE, a fragment from amyloid-β protein, with a hydrophobic core delimited by two charged amino acids. The assembly of GNNQQNY suggests a mechanism of monomer addition, with a bias toward parallel peptide pairs and a gradual increase in the amount of β-strand content. For KLVFFAE, a mechanism involving dimers rather than monomers is revealed, involving a generally higher β-strand content and a transition toward a larger number of antiparallel peptide pairs during the rearrangement of the hexamer. The differences observed for the aggregation of the two peptides suggests the existence of a sequence−aggregation relationship.



INTRODUCTION

Amyloid protein aggregation is one of the main factors involved in the onset of several neurodegenerative conditions, including Alzheimer's, Parkinson's, and prion diseases. The aberrant assembly of proteins such as amyloid-β (Aβ), α synuclein, and the prion protein into fibrils has been associated with the death of neuronal cells. There is increasing evidence that oligomers, present in the early stages of aggregation, are the neurotoxic species, rather than highly organized fibrils.[1−4] The prion protein and Aβ are two examples of amyloidogenic proteins whose early assembly mechanisms have been investigated using both experimental[3,5,6] and computational tools.[7−10] Two short fragments from these proteins exhibit aggregation behavior and cross-β structures similar to the ones specific to fibrils formed by the full length proteins. One is the GNNQQNY fragment from the N-terminal prion determining region of the yeast protein Sup35 (abbreviated GNN). The second is the KLVFFAE fragment from Aβ (abbreviated KLV). The motivation for studying these shorter fragments of amyloidogenic proteins is the hypothesis that their behavior captures essential features of the original proteins.[11] Moreover, these two peptides have very different primary structures: GNN is polar, while KLV has a hydrophobic core. They are therefore likely to exhibit different aggregation pathways, whose elucidation would contribute to a better understanding of the sequence−aggregation relationship.

Current stochastic models that describe the aggregation process[12] are mostly applicable to simple cases involving the attachment or detachment of a monomer to a larger aggregate, without any information about the conformational structure. A different method for describing stochastic biological processes involves kinetic transition networks[13,14] and disconnectivity graphs,[15,16] and has already been applied to the study of conformational dynamics of peptides or protein folding[17−20] and recently to aggregation.[21] In the present work, we characterize the assembly of six monomers into one hexamer for GNN and KLV, based on molecular dynamics (MD) simulations, and compare the pathways. Hexamers were chosen because previous experimental[22] and theoretical[23,24] studies suggest a classical nucleation process for the aggregation of GNN with a critical nucleus of four to six monomers.

To obtain a general picture of the complex aggregation process, we first derive coarse-grained transition networks (CGTNs) showing the transitions between oligomers of different sizes. To provide a more detailed picture, we introduce aggregation states defined on the basis of certain structural characteristics, in addition to oligomer size, and derive kinetic transition networks (KTNs) illustrating the population and transitions between these states.[13] As these KTNs are rather complex, we apply the max-flow min-cut procedure, which allows us to remove less important edges of the KTNs while preserving the maximum flow.[17,25−27] The resulting networks, which we denote mincut transition networks (MTNs), provide a concise picture of the aggregation pathway in terms of stable

Received: December 26, 2013
Revised: January 8, 2014
Published: January 8, 2014

aggregation states and transitions between them. However, they lack kinetic information, which we include by describing the free energy landscape for aggregation through disconnectivity graphs (DGs)[15,16] using the node and edge data from the MTNs. Finally, we provide information regarding the dynamics of aggregation through first passage time distributions (FPTDs) for reaching the most populated aggregation states. Although both peptides have been previously studied with both coarse-grained[23,24,28,29] and all-atom models,[14,21,30−32] we believe this is the first comparison of their aggregation behavior using a kinetic description that involves transition networks (TNs).

MATERIALS AND METHODS

MD Simulations. For both GNN and KLV, we performed five independent all-atom MD simulations of six peptides inserted in a cubic box with side length 10 nm filled with explicit SPC water,[33] resulting in a concentration of about 10 mM. Each simulation was 500 ns long, producing a total of 2.5 μs per sequence. The simulations were performed with the Gromacs 4.5.5 parallel software package[34] and the GROMOS96 43A1 force field.[35] Hydrogen atoms were treated as virtual interaction sites, permitting an integration time step of 5 fs while maintaining energy conservation.[36] The system was maintained at 310 K and 1 bar via velocity rescaling with a stochastic term algorithm[37] and the Berendsen barostat,[38] respectively. Electrostatic interactions were described using the particle mesh Ewald algorithm[39,40] with a cutoff of 0.9 nm. van der Waals interactions were cut off at 1.4 nm with a switching distance of 0.9 nm. During each MD simulation, structures were saved every 5 ps, resulting in 500 000 snapshots per peptide.

Definition of Aggregation States. The aggregation state is defined as a number with five digits, $N1N2N3N4N5$, where each digit corresponds to a structural feature or to an oligomeric state. $N1$ represents the oligomeric size, identified using a cutoff distance of 0.5 nm between any two atoms belonging to different peptides. $N2$ is the number of peptides connected by at least one hydrogen bond. Hydrogen bonds are defined on the basis of distance and angle cutoffs of 0.41 nm and 40°, respectively. $N3$ and $N4$ are the numbers of antiparallel and parallel peptides within the oligomer. Two chains are antiparallel or parallel if they are connected by at least two hydrogen bonds involving two unique amino acids from each peptide. $N5$ is the average number of amino acids in β-strand conformation per peptide in the oligomer. An amino acid is defined to be in a β-strand conformation if the dihedral angles ϕ and ψ (in degrees) of the backbone are contained in the polygon with vertices (−180, 180), (−180, 126), (−162, 126), (−162, 108), (−144, 108), (−144, 90), (−50, 90), and (−50, 180).[41]

Transition Matrix. To calculate the transition matrix that includes all pairwise transitions between aggregation states, we first identified all the states and the number of transitions between them along the 2.5 μs trajectory using a lag time of 5 ps. Using these states and transitions, we built an $N \times N$ matrix, where N is the number of states encountered. Each element of the matrix represents the population of a particular transition between two states. After normalizing the rows, each element represents the probability for a particular transition to occur. The transition matrix did not change significantly for longer lag times.

Kinetic Transition Networks. The row normalized transition matrices were converted into KTNs using the software Visone[42] and the minimum stress algorithm. In the KTN plots, the nodes represent aggregation states, the area of each node is proportional to the population of the state, and the color of the node indicates the average number of amino acids in the β-strand conformation ($N5$). The color of the network edges corresponds to the value of the transition probability. The CGTNs were constructed using oligomeric states with the thickness of the edges encoding the total number of transitions between pairs of states. On the basis of KTNs, we have derived networks that preserve the maximum flow using the mincut procedure, which we denote as MTNs.[17,25−27] In the case of MTNs, the thickness of the edges corresponds to the number of transitions and the color of the nodes to the quantity ($N3-N4$).

Representative States and Conformations. For each KTN, we selected representative aggregation states and illustrate representative conformations in Figure 2. To select representative states, we first sorted by decreasing population and then identified the states with the highest population for each $N1$ value (or oligomer). For each state, we selected a structure that reflects characteristics of the aggregation state. The snapshots were generated with the software package VMD.[43]

Disconnectivity Graphs. The number of times state i was visited during the simulations was used to define the partition function as $Z_i = \sum_j n_{ij}$, where n_{ij} is the number of transitions from state j to state i.[17,26] The free energy (FE) of state i was derived from Z_i as $F_i = -kT \ln Z_i$, where k is the Boltzmann constant and T is the temperature. To calculate the FE barrier between two states, we performed the following steps: (1) We derived the undirected graph with edge capacities $c_{ij} \sim (n_{ij} + n_{ji})/2$. (2) We applied the mincut algorithm to the new undirected network to obtain the new transition capacities $c'_{ij} = c'_{ji}$, which reflect the maximum flow through the network. (3) We derived the partition function $Z_{ij} = c'_{ij}(h/kT)(1/t_s)$, where h is the Planck constant, T is 310 K, and t_s is the sampling time interval of 5 ps. (4) We estimated the FE barriers as $F_{ij} \sim -kT \ln(Z_{ij})$.[26] Using the FE minima and barriers, we constructed FE disconnectivity graphs based on the method developed by Becker and Karplus.[15]

First Passage Time Distributions. The first passage time is defined as the time interval needed to reach a certain aggregation state along the MD trajectory. For a particular aggregation state, we calculated all the first passage times for each of the five trajectories and then constructed a histogram. FPTDs starting from specific states were calculated by considering only the time intervals originating at positions where the specific states occurred.

RESULTS AND DISCUSSION

Coarse-Grained Transition Networks. To obtain a general picture of the assembly process, we have derived CGTNs that describe the dynamics of the system at the oligomer level (Figure 1). In the case of GNN (Figure 1a), hexamers are the dominant oligomers at 310 K, with a population of 77.2%, followed by monomers (11.9%), pentamers (4.8%), dimers (3%), trimers (2.3%), and tetramers (0.7%). For KLV (Figure 1b), the hexamers are also the most populated oligomeric state (65.4%), followed by monomers (15.2%), dimers (11.0%), tetramers (3.9%), pentamers (2.9%), and trimers (1.6%). The high hexamer population indicates fast initial assembly of small oligomers into a hexamer that undergoes slower rearrangements. This result may reflect the

dx.doi.org/10.1021/jp412648u | J. Phys. Chem. B 2014, 118, 1003−1011

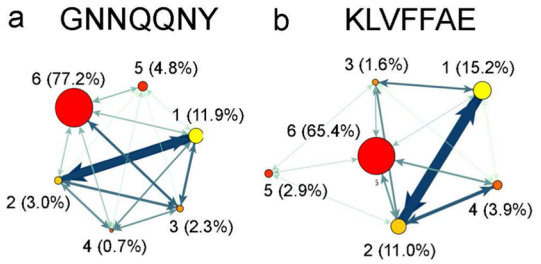


Figure 1. Coarse grained transition networks for GNN (a) and KLV (b). The oligomeric states are shown as nodes connected by edges that represent the total number of transitions between two oligomeric states. The area of each node corresponds to the population, while the color corresponds to the oligomeric state. The color and thickness of the edges change from light green to dark blue and thin to thick, respectively, and correspond to changes from fewest to most transitions between oligomeric states.

concentration (10 mM) used in our simulations. Studies using higher concentrations of around 100 mM report a spontaneous collapse into one disordered aggregate.[44] The fact that GNN has a larger population of hexamers (77.2%) than KLV (65.4%) indicates faster assembly, which might result from the presence of a tyrosine residue.[30]

From the number of transitions between oligomeric states, encoded in the thickness of the edges between the nodes, it is clear that the largest number of transitions occurs between monomers and dimers in both systems. An important observation in the case of KLV is the lack of transitions from monomers to pentamers. This result indicates that pentamers are formed mainly through the combination of dimers and trimers. In the case of GNN, monomers contribute to all oligomeric states, suggesting a growth/breakage of the oligomer by the addition/removal of a monomer. This mechanism is supported by recent coarse-grained studies,[23,29] which emphasize the role of monomer addition/loss up to oligomers of size 5 or 6 at 300 K. Beyond that size, the main aggregation mechanism involved oligomers rather than monomers. The CGTNs described above indicate significant differences between the aggregation of the two peptides. To obtain a better picture of the assembly process, we now consider more detailed kinetic transition networks.

Kinetic Transition Networks. KTNs for GNN and KLV, based on the aggregation state definition (*N*1*N*2*N*3*N*4*N*5) from the Materials and Methods section, are shown in Figure 2a and b, respectively. In both cases, there are three distinct interconnected clusters: C1 for states with *N*1 (oligomer size) values of 1−4, C2 for states with *N*1 = 5, and C3 for *N*1 = 6. For KLV, the clusters are more interconnected than for GNN, which suggests a more dynamic assembly process, with larger oligomers breaking into smaller ones more frequently. The coloring of the nodes is based on the average number of β-strand amino acids per peptide, encoded in *N*5. This information reveals an overall higher β-strand content in KLV relative to GNN, even for the states with low *N*1 values, indicating a higher degree of organization during the early assembly stages for KLV. It is important to note the central location of the states with *N*1 = 1 (e.g., 10001) in the case of GNN but not for KLV. This result agrees with the CGTNs, where we observed that monomers contribute to all oligomeric states for GNN but not for KLV.

While both C1 clusters include states with different *N*1 values, the structure for GNN is more compact and has fewer transitions to cluster C3 than for KLV. Representative aggregation states for GNN are 10001 with a population of 4.6%, 22003 (0.4%), 33002 (0.3%), and 44203 (0.1%). Representative conformations for each state are shown in Figure 2a. They generally have a small number of parallel or antiparallel peptide pairs and an average number of β-strand amino acids below three. In the case of KLV, representative aggregation states from cluster C1 are 10003 (2.5%), 22104 (3.0%), 33113 (0.4%), and 44214 (1.7%). For most of these states, the number of parallel and antiparallel peptide pairs is larger than for GNN and the number of antiparallel pairs is

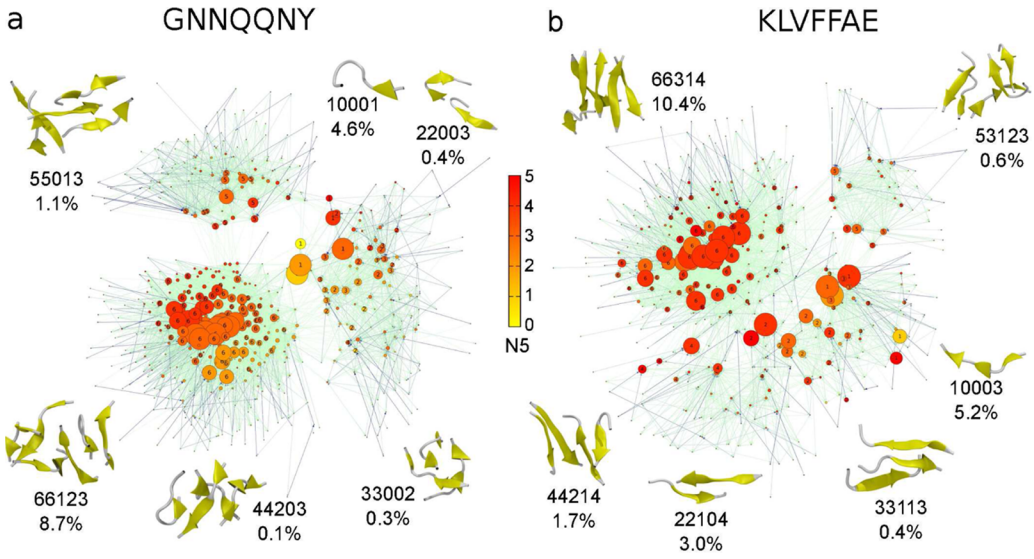


Figure 2. Kinetic transition networks for GNN (a) and KLV (b). For both peptides, the aggregation states *N*1*N*2*N*3*N*4*N*5 are shown as nodes connected by edges that represent transition probabilities. The size of each node corresponds to the population, and the color corresponds to the average number of β-strand amino acids, *N*5. The color of the edges reflects the value of the transition probability, from light green for small probabilities to dark blue for high probabilities. For each oligomer size, *N*1, we illustrate representative conformations corresponding to the aggregation state with the highest population, together with the population of that state in percentage.

dx.doi.org/10.1021/jp412648u | J. Phys. Chem. B 2014, 118, 1003−1011

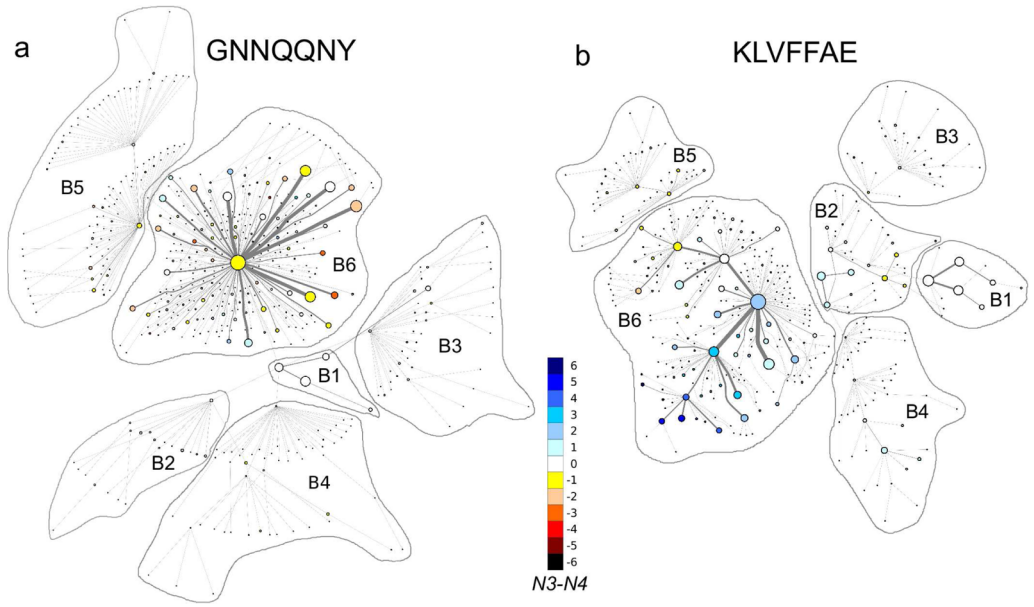


Figure 3. Mincut transition networks for GNN (a) and KLV (b). The coloring of the nodes reflects $N3-N4$, which indicates the excess number of antiparallel peptide pairs relative to parallel peptide pairs. Edges shown in gray represent the number of transitions returned by the mincut algorithm. Clusters of states with identical values of $N1$ have been outlined and identified as basin Bi, where i takes the value of $N1$.

higher than the number of parallel peptide pairs. In addition, the higher average number of β-strand amino acids indicates more ordered small oligomers for KLV (Figure 2b).

The C2 clusters contain only states with $N1 = 5$, with a larger population for GNN. The representative aggregation states for this cluster are 55013 with a population of 1.1% for GNN and 53123 (0.6%) for KLV. Both states have more parallel pairs than antiparallel, with more pairs in total for KLV (antiparallel:parallel = 1:2) than for GNN (0:1) and thus more β-sheet-like structures. However, GNN has more chains connected by hydrogen bonds. The most populated cluster for both fragments is C3, containing states with $N1 = 6$. The most representative aggregation states for this cluster are 66123 (8.7%) for GNN and 66314 (10.4%) for KLV. GNN samples a higher average number of parallel pairs (1:2), while for KLV the states with more antiparallel pairs are more common (3:1). In terms of β-strand content, cluster C3 shows a gradual transition from low to high for GNN, while KLV has a more uniform distribution of higher β-strand content. State 66314 of KLV has four β-strand amino acids on average.

Mincut Transition Networks. The KTNs described in the previous section provide a detailed picture of the aggregation process. However, the large number of transitions and nodes makes it difficult to pinpoint the main pathways that dominate the assembly process. To identify the maximum flow in the transition network, we applied the minimum cut method (mincut) and derived MTNs.

The MTNs of the two peptides are shown in Figure 3, where we identify the states with the same value of $N1$ as basin Bi, with $i = 1, ..., 6$. For GNN, basin B6 has one central node corresponding to state 66123, which is connected radially to the other states. The highly populated states from B6 have an edge of high capacity in direct contact with the central node, and some of them have one to three edges directly connected to less populated nodes. The GNN assemblies from basin B6 have a higher propensity to form parallel pairs but also contain a significant number of antiparallel pairs. State 66123 is also the one that connects B6 to the other basins. An important observation is that the transition to/from B6 from/to B2, B3, and B4 occurs through B1, specifically through state 10002, which is not the highest populated state within B1. This observation indicates the central role of monomers in the assembly process for GNN.

The MTN of KLV is shown in Figure 3b. The topology of basin B6 is very different from that of GNN. Rather than having one central node surrounded radially by the other nodes, KLV displays several nodes that are at the center of local clusters. One can identify five such nodes connected to the central node, corresponding to state 66314. Three of four nodes are among the top 10 states with the highest population (Table 1). The difference in antiparallel to parallel pairs ($N3-N4$) shows a clear shift from negative to positive values for the five local clusters. This shift corresponds to a gradual transition from more parallel to more antiparallel pairs. Overall, the nodes in basin B6 have a higher propensity for antiparallel pairs. Interestingly, the nodes with more parallel pairs are connected to basin B5 through state 66124, while the majority of nodes

Table 1. The 10 Aggregation States with the Highest Populations for GNN and KLV[a]

GNNQQNY		KLVFFAE	
aggregation state	population (%)	aggregation state	population (%)
66123	8.6	66314	10.4
66023	6.2	10003	5.2
66124	5.6	66214	5.0
66013	5.2	66304	4.4
66113	4.9	66224	4.0
10001	4.6	10004	3.9
66223	3.7	10002	3.6
10002	3.3	66124	3.2
66213	2.9	66324	3.1
66033	2.1	22104	3.0

[a]For each peptide, the aggregation state $N1N2N3N4N5$ is shown in the left column with the percentage population in the right column.

dx.doi.org/10.1021/jp412648u | J. Phys. Chem. B 2014, 118, 1003−1011

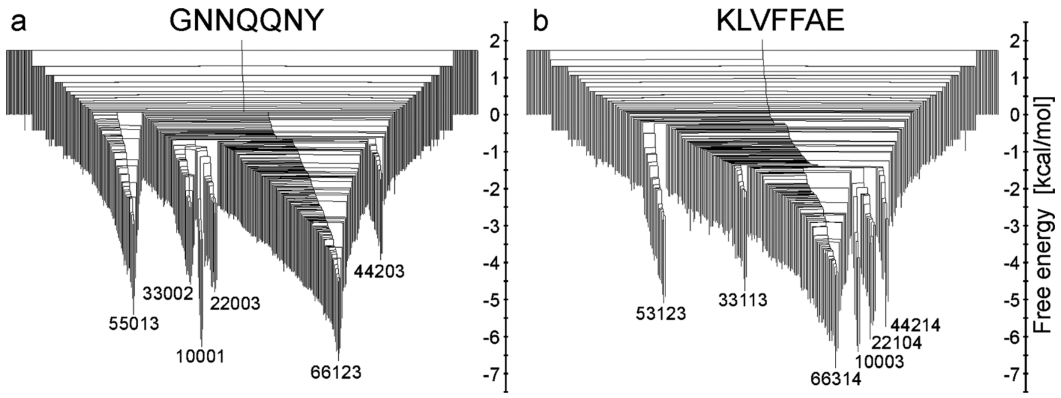


Figure 4. Mincut disconnectivity graphs for GNN (a) and KLV (b). The vertical axis is $F_i \sim -kT \ln Z_i$ and $F_{ij} \sim -kT \ln Z_{ij}$ for minima and barriers, respectively. Representative aggregation states for the lowest free energy minima are labeled, and corresponding structures are shown in Figure 2.

with more antiparallel pairs are closer and connected to basins B1−B4 through state 66314. Basins B1, B3, and B4 are connected to basin B6 through B2 (state 22103), suggesting that dimers play a central role in the pathway to hexamers.

One major difference between the above networks is the complex nature of basin B6, which suggests distinct paths for evolution of the collapsed hexamer in the two peptides. The distribution of parallel and antiparallel pairs is another characteristic of the two B6 basins. Numerous studies have reported a preference for parallel β-strands in GNN[23,29−31] and antiparallel β-strands in KLV,[32] in agreement with our findings. The central role of monomers in the assembly of small oligomers into hexamers is also clear from the MTN of GNN, in contrast to the importance of dimers for KLV. A two-stage "dock-lock" mechanism[32] for monomer addition to preformed structured oligomers was previously proposed for KLV fibril formation. Our results suggest that pathways involving dimers might also be relevant.

To explore the role of electrostatic interactions during oligomer formation, we identified salt bridges between amino acids K and E from different KLV peptides. We found one, two, and three salt bridges present in the system for 23.6, 4.3, and 0.2% of the simulation time, respectively. Given that we did not observe in-register antiparallel β-sheets, as found in the KLVFFAE fibrils,[45] and that the salt bridges are not as abundant as in a computational study of preformed KLVFFAE fibrils,[46] we suggest that salt bridges play a secondary role in the early aggregation of KLVFFAE; i.e., they are not as important as in fibril formation and stability.

While we did not observe actual fibrils stabilized by steric zippers in our simulations, we did find KLVFFAE hexamers resembling β-barrel structures (Figure 2b, state 66314). A similar hexamer arrangement has been recently observed in experiments performed by Laganowsky et al. and Liu et al.[47,48] when studying the aggregation of segments of the amyloid-forming protein αB crystallin. In these studies, they report cylindrical oligomers (cylindrins) formed by six out-of-register β-sheets as early oligomers off-pathway in terms of in-register fibril formation, but on-pathway for less stable out-of-register fibrils. The oligomers reported by Laganowsky et al. were found to be toxic, and the β-barrel conformation was suggested as a possible template for toxic amyloidogenic oligomers. The stability of the cylindrins was later studied computationally by Berhanu et al.[49] A recent computational study confirmed the existence of β-barrel oligomers containing eight KLVFFAE peptides,[50] similar to the hexamers observed in our study. In the case of KLVFFAE, the out-of-register arrangement of β-sheets has been observed experimentally and shown to lead to stable nanotubular structures, rather than the classical amyloid fiber with cross-β structure.[51] The fact that we do not observe fibrils, in contrast to Berhanu et al.,[46] is probably due to the limited number of peptides used in our study and also to the relatively short simulation time compared to the time necessary for fibril formation, which *in vitro* requires minutes or even hours. Moreover, if the hexamers with β-barrel-like structure observed in our study are indeed off-pathway for in-register fibrils, there might be a rather high energy barrier to formation of the in-register β-sheet fibrils, as indicated by Laganowsky et al.[47]

Disconnectivity Graphs. Disconnectivity graphs were originally introduced to describe the dynamics of a small peptide[15] and were subsequently employed to gain insight into the emergence of observable properties from the underlying potential energy landscape in a wide variety of systems.[16] In the present work, we have constructed DGs for each peptide based on the method described by Krivov et al.[26] (Figure 4). In Figure 4a, one can identify several free energy minima for GNN with the lowest value at 310 K for state 66123, followed in order of increasing FE by 10001, 55013, 22003, 33002, and 44203. For KLV, the state with the lowest FE at 310 K is 66314, followed by states 10003, 22104, 44214, 53123, and 33113. The graphs indicate a FE barrier for the GNN trimer formation from monomers, or to state 33002 from state 10001, of about 5.4 kcal/mol. This value is slightly higher than the FE barriers (around 3 kcal/mol) between disordered aggregates and in-register mixed parallel/antiparallel aggregates reported by Gsponer et al.[30] In the case of dimer formation, or transitions to state 22003 from 10001, the FE barrier is around 5.1 kcal/mol. Barriers of 2 and 3 kcal/mol for GNN dimer formation have been reported previously by Reddy et al.[52] and Strodel et al.[14] The higher FE barrier observed in our simulations might be due to the fact that the monomers are initially spatially separated in explicit solvent. Overall, GNN has higher barriers for transitions to the most stable aggregation state than KLV in all cases, except for states with $N1 = 2$ and 4. This is an interesting result in the context of the earlier observation that monomers and dimers are central to the assembly process in GNN and KLV, respectively. In the case of GNN, the higher FE barrier for aggregation might be due to the polar nature of the peptide, while for KLV the fact that the FE barrier is higher for dimers and tetramers might be related to the reduced hydrophobic surface compared to monomers.

dx.doi.org/10.1021/jp412648u | J. Phys. Chem. B 2014, 118, 1003−1011

First Passage Time Distributions. To provide further insight into the underlying dynamics, we have calculated first passage time distributions (FPTDs) for the states with the highest populations, i.e., 66123 for GNN and 66314 for KLV (Figure 5). The distributions differ in two important aspects:

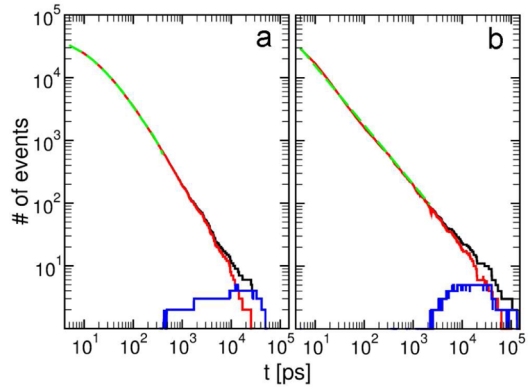


Figure 5. First passage time distributions for the most populated aggregation states 66123 and 66314 of GNN (a) and KLV (b). The black, red, and blue curves correspond to FPTDs from all times, from any time at which a state with $N1 = 6$ is present and from any time at which a state with $N1 < 6$ occurred. The green curves are nonlinear fitting functions for the FPTD for the time intervals with $N1 = 6$ that do not overlap with the blue curve. For part a, the fitting function is $y = a_1e^{-t/b1} + a_2e^{-t/b2} + a_3e^{-t/b3}$ with parameters $a_1 = 26620$, $b_1 = 8.9$ ps, $a_2 = 15878$, $b_2 = 36.5$ ps, $a_3 = 3854$, and $b_3 = 216$ ps. For part b, the fitting function is $y = ct^{-d}$ with parameters $c = 137294$ and $d = 0.94$ ps.

for time intervals below 1 ns, GNN displays significant curvature, while the FPTD of KLV is practically linear, and for longer times, the distribution decays faster for GNN, reaching one event per time interval around 60 ns, compared to 200 ns for KLV. To understand these differences, we calculated FPTDs starting from states with $N1 = 6$ and with $N1 < 6$. The two curves are shown in red ($N1 = 6$) and blue ($N1 < 6$) in Figure 5. For both peptides, the FPTD is dominated by transitions between states with $N1 = 6$, equivalent to the rearrangement of the hexamer. The FPTs starting from states with $N1 < 6$ occur only at lag times of around 400 ps to 50 ns for GNN and 2 to 200 ns for KLV, and remain below 4 events per time interval. These time ranges indicate faster assembly of hexamers for GNN than for KLV, confirmed by the larger number of hexameric states observed in the CGTN. A previous study reported a mean first passage time to form an ordered KLV trimer from monomers of 244 ns with extremes between 7 and 600 ns at 300 K for monomer concentrations of 40−70 mM.[32] We estimate a FPT from KLV monomers to state 66314 below about 200 ns. The two values are not directly comparable due to different concentrations (10 mM in the present work), different temperatures (310 K here), and the fact that state 66314 contains six peptides and is not as ordered as the trimer studied in ref 32. A trimer in a large bath of oligomers might exhibit a shorter FPT and might not reach a more ordered state due to interactions with other oligomers.

Power law behavior for the FPTD is associated with a barrierless process, while exponential behavior indicates a rate limiting barrier.[18] In Figure 5, we selected time ranges corresponding to FPTs starting from $N1 = 6$ and excluding overlapping states with FPTs starting from $N1 < 6$. For these time ranges, 0−400 ps for GNN and 0−1600 ps for KLV, we fitted different combinations of exponential ($ae^{-t/b}$) and power law (ct^{-d}) functions. While for GNN the best fit was obtained using a combination of at least three exponentials, a single power law function was sufficient for KLV. Although it is difficult to attribute the three exponential functions for GNN to a particular state or group of states, it is clear that the dynamical processes involved in the rearrangement of the GNN and KLV hexamers are different. To obtain further insight into this difference, we analyzed the states that contribute most to the FPTD, starting from $N1 = 6$.

Since the FPTs starting from $N1 = 6$ dominate the distribution, we identified all the contributing states and ranked them according to their populations. For each of the top five states, we recalculated the FPTD originating from that state and ending in 66123 for GNN and 66314 for KLV. These distributions are included in Figure 6 together with the total

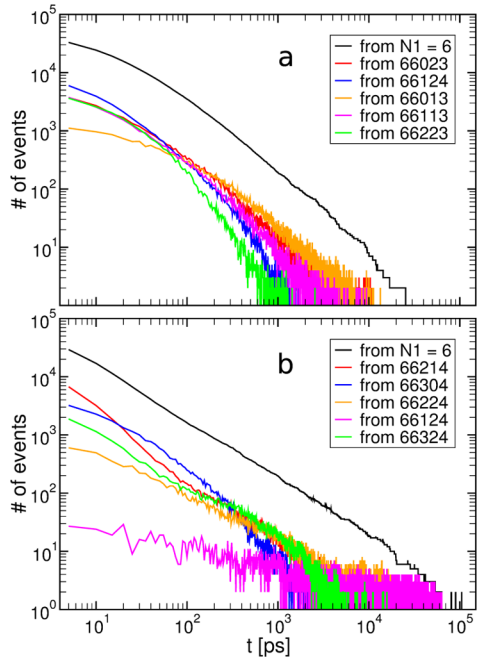


Figure 6. First passage time distributions for the most populated aggregation states 66123 and 66314 of GNN (a) and KLV (b). The black curve corresponds to FPTDs from any time at which a state with $N1 = 6$ is present. The colored curves correspond to FPTDs started from specific states. The specific states are the top five with the highest population that contribute most to the distribution from the black curves and are arranged in order of decreasing population from top to bottom.

FPTDs starting from states with $N1 = 6$. For GNN, the FPTD from state 66124 dominates very short time intervals below 100 ps. Since the target state is 66123, this result suggests that the decrease of $N5$ from four to three or backbone reorganization is the rate limiting process for the hexamer rearrangement at short time intervals. Beyond 100 ps, the FPTD from states 66013, 66023, and 66113 makes the largest contribution, indicating the creation/breakage of antiparallel ($N3$) and parallel ($N4$) pairs as the rate limiting process. For KLV, the five states with the highest populations are of type 66XX4, where $X = 0, ..., 6$. Since the target state is 66314, this observation suggests that the formation and breakage of antiparallel ($N3$) and parallel ($N4$) pairs is the rate limiting process in the evolution of the KLV hexamer.

dx.doi.org/10.1021/jp412648u | J. Phys. Chem. B 2014, 118, 1003−1011

CONCLUSIONS

In this work, we have highlighted differences between the aggregation pathways of two peptides with alternative amino acid sequences. While the two fragments might reflect the aggregation behavior of the full proteins, the relation between sequence and assembly mechanism is also of general interest. The early aggregation process of the two protein fragments from a large number of monomers has been studied before *in silico*, mostly using coarse-grained methods.[23,24,28,29] More detailed all-atom studies have focused on the assembly of two or three monomers,[14,21,30,32,52,53] the addition of monomers to a preformed oligomer,[32,54,55] or the stability of preordered fibrilar structures.[31,56,57] Here we investigated the aggregation of six monomers using all-atom MD simulations in explicit solvent.

We have considered both coarse-grained and more detailed kinetic transition networks, which identify monomers as the most important species in the aggregation of GNN. The kinetic transition networks indicate a higher β-strand content for all KLV assemblies compared to GNN, due to the hydrophobic residues, while in the case of the GNN hexamer a gradual transition from low to high β-strand content is observed. Mincut transition networks reveal a radial distribution of aggregation states of GNN for hexamer rearrangements dominated by parallel peptide pairs and a modular distribution in the case of KLV, with a gradual increase in the relative number of antiparallel to parallel peptide pairs. Disconnectivity graphs reveal higher FE barriers for hexamer formation from monomers, trimers, and pentamers of GNN compared to KLV and from dimers and tetramers of KLV compared to GNN. This result is probably due to the hydrophilic/hydrophobic nature of the two peptides and emphasizes the importance of the stability of monomers and dimers for the aggregation of GNN and KLV, respectively. The first passage time distributions indicate an assembly process dominated by the rearrangement of the hexamer for both peptides, with faster aggregation for GNN, perhaps due to the presence of the tyrosine residue.

We find that monomers play a key role in the aggregation of GNN, consistent with previous studies indicating that the main pathway for the prion protein Sup35 propagation is monomer addition.[58] Moreover, the reduced β-strand content in the intermediate oligomers of GNN observed in our study agrees with a previous observation that oligomers of Sup35 lack β-strand structure before being incorporated into fibrils.[59] For KLV, the observation that the dimers play an important role in the aggregation process is consistent with the contribution of dimers and hexamers to aggregation of Aβ protein (1−42).[6,7] Furthermore, the high β-strand content of the KLV oligomers in our study is consistent with a higher β-strand propensity in Aβ protein oligomers.[7] The antiparallel nature of the fragment KLV, which contrasts with the in-register parallel fibrils formed by the full-length Aβ protein,[60] might play an important role in unstructured toxic oligomers that most probably lack the in-register β-sheet structure.[61,62]

In conclusion, our study highlights key differences in the assembly process of two aggregation-prone peptides with contrasting amino acid sequences, as well as connections to the aggregation of the corresponding full-length proteins. Furthermore, the present study provides proof of principle that the method introduced here allows a concise description of the thermodynamics and kinetics of protein aggregation. In future work, it will be applied to systems including full-length Aβ, where the results obtained for KLV will be of great value for understanding the aggregation. Finally, our approach reveals for the first time in a one-to-one comparsion of KLV and GNN their different aggregation pathways, highlighting the importance of sequence.

AUTHOR INFORMATION

Corresponding Author

*E-mail: b.strodel@fz-juelich.de. Phone: +49 2461 613670. Fax: +49 2461 619497.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

B.B. and B.S. gratefully acknowledge the Jülich Supercomputing Centre for the computer time made available through grants JICS61 and JICS62. D.J.W. gratefully acknowledges financial support from the EPSRC and the ERC. We also thank Prof. Gerhard Stock for fruitful discussions.

REFERENCES

(1) Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297*, 353−356.

(2) Kirkitadze, M. D.; Bitan, G.; Teplow, D. B. Paradigm Shifts in Alzheimer's Disease and Other Neurodegenerative Disorders: The Emerging Role of Oligomeric Assemblies. *J. Neurosci. Res.* **2002**, *69*, 567−577.

(3) Silveira, J. R.; Raymond, G. J.; Hughson, A. G.; Race, R. E.; Sim, V. L.; Hayes, S. F.; Caughey, B. The Most Infectious Prion Protein Particles. *Nature* **2005**, *437*, 257−261.

(4) Haass, C.; Selkoe, D. J. Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid β-Peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101−112.

(5) Klein, W. L.; Stine, W. B.; Teplow, D. B. Small Assemblies of Unmodified Amyloid β-Protein are the Proximate Neurotoxin in Alzheimer's Disease. *Neurobiol. Aging* **2004**, *25*, 569−580.

(6) Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. Amyloid β-Protein (Aβ) Assembly: Aβ40 and Aβ42 Oligomerize through Distinct Pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330−335.

(7) Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B. Elucidation of Amyloid β-Protein Oligomerization Mechanisms: Discrete Molecular Dynamics Study. *J. Am. Chem. Soc.* **2010**, *132*, 4266−4280.

(8) Barz, B.; Urbanc, B. Dimer Formation Enhances Structural Differences between Amyloid β-Protein (1−40) and (1−42): An Explicit-Solvent Molecular Dynamics Study. *PLoS One* **2012**, *7*, e34345.

(9) Melquiond, A.; Dong, X.; Mousseau, N.; Derreumaux, P. Role of the Region 23−28 in Aβ Fibril Formation: Insights from Simulations of the Monomers and Dimers of Alzheimer's Peptides Aβ40 and Aβ42. *Curr. Alzheimer. Res.* **2008**, *5*, 244−250.

(10) Sekijima, M.; Motono, C.; Yamasaki, S.; Kaneko, K.; Akiyama, Y. Molecular Dynamics Simulation of Dimeric and Monomeric Forms of Human Prion Protein: Insight into Dynamics and Properties. *Biophys. J.* **2003**, *85*, 1176−1185.

(11) Balbirnie, M.; Grothe, R.; Eisenberg, D. S. An Amyloid-Forming Peptide from the Yeast Prion Sup35 Reveals a Dehydrated β-Sheet Structure for Amyloid. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2375−2380.

(12) Yvinec, R.; D'Orsogna, M. R.; Chou, T. First Passage Times in Homogeneous Nucleation and Self-Assembly. *J. Chem. Phys.* **2012**, *137*, 244107.

(13) Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3−10.
(14) Strodel, B.; Whittleston, C. S.; Wales, D. J. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005−16014.
(15) Becker, O. M.; Karplus, M. The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics. *J. Chem. Phys.* **1997**, *106*, 1495−1517.
(16) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal Energy Landscapes. *Nature* **1998**, *394*, 758−760.
(17) Krivov, S. V.; Karplus, M. Hidden Complexity of Free Energy Surfaces for Peptide (Protein) Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766−14770.
(18) Rao, F.; Karplus, M. Protein Dynamics Investigated by Inherent Structure Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 9152−9157.
(19) Baba, A.; Komatsuzaki, T. Construction of Effective Free Energy Landscape from Single-Molecule Time Series. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19297−19302.
(20) Gfeller, D.; Rios, P. D. L.; Caflisch, A.; Rao, F. Complex Network Analysis of Free-Energy Landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817−1822.
(21) Riccardi, L.; Nguyen, P. H.; Stock, G. Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1471−1479.
(22) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the Cross-β Spine of Amyloid-like Fibrils. *Nature* **2005**, *435*, 773−778.
(23) Nasica-Labouze, J.; Mousseau, N. Kinetics of Amyloid Aggregation: A Study of the GNNQQNY Prion Sequence. *PLoS Comput. Biol.* **2012**, *8*, e1002782.
(24) Nasica-Labouze, J.; Meli, M.; Derreumaux, P.; Colombo, G.; Mousseau, N. A Multiscale Approach to Characterize the Early Aggregation Steps of the Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35. *PLoS Comput. Biol.* **2011**, *7*, e1002051.
(25) Ford, L. R.; Fulkerson, D. R. Maximal Flow through a Network. *Can. J. Math.* **1956**, *8*, 399−404.
(26) Krivov, S. V.; Karplus, M. Free Energy Disconnectivity Graphs: Application to Peptide Models. *J. Chem. Phys.* **2002**, *117*, 10894−10903.
(27) Gomory, R. E.; Hu, T. C. Multi-Terminal Network Flows. *J. Soc. Ind. Appl. Math.* **1961**, *9*, 551−570.
(28) Cheon, M.; Chang, I.; Hall, C. K. Spontaneous Formation of Twisted Aβ(16−22) Fibrils in Large-Scale Molecular-Dynamics Simulations. *Biophys. J.* **2011**, *101*, 2493−2501.
(29) Osborne, K. L.; Bachmann, M.; Strodel, B. Thermodynamic Analysis of Structural Transitions During GNNQQNY Aggregation. *Proteins* **2013**, *81*, 1141−1155.
(30) Gsponer, J.; Haberthr, U.; Caflisch, A. The Role of Side-Chain Interactions in the Early Steps of Aggregation: Molecular Dynamics Simulations of an Amyloid-Forming Peptide from the Yeast Prion Sup35. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5154−5159.
(31) Vitagliano, L.; Esposito, L.; Pedone, C.; De Simone, A. Stability of Single Sheet GNNQQNY Aggregates Analyzed by Replica Exchange Molecular Dynamics: Antiparallel Versus Parallel Association. *Biochem. Biophys. Res. Commun.* **2008**, *377*, 1036−1041.
(32) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer Adds to Preformed Structured Oligomers of Aβ-Peptides by a Two-Stage Dock-Lock Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111−116.
(33) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; The Jerusalem Symposia on Quantum Chemistry and Biochemistry 14; Springer: Houten, The Netherlands, 1981; pp 331−342.
(34) 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.
(35) Scott, W. R. P.; Hnenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(36) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20*, 786−798.
(37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(38) 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.
(39) 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*, 10089−10092.
(40) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(41) Klimov, D. K.; Thirumalai, D. Dissecting the Assembly of Aβ_{16-22} Amyloid Peptides into Antiparallel β Sheets. *Structure* **2003**, *11*, 295−307.
(42) Brandes, U.; Wagner, D. Visone - Analysis and Visualization of Social Networks. *Graph Drawing Software*; Springer-Verlag Berlin: Heidelberg, Germany, 2003; pp 321−340.
(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(44) Baftizadeh, F.; Pietrucci, F.; Biarns, X.; Laio, A. Nucleation Process of a Fibril Precursor in the C-Terminal Segment of Amyloid-β. *Phys. Rev. Lett.* **2013**, *110*, 168103.
(45) Colletier, J.-P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. Molecular Basis for Amyloid-β Polymorphism. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16938−16943.
(46) Berhanu, W. M.; Hansmann, U. H. E. Side-Chain Hydrophobicity and the Stability of Aβ_{16-22} Aggregates. *Protein Sci.* **2012**, *21*, 1837−1848.
(47) Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P. K.; et al. Atomic View of a Toxic Amyloid Small Oligomer. *Science* **2012**, *335*, 1228−1231.
(48) Liu, C.; Zhao, M.; Jiang, L.; Cheng, P.-N.; Park, J.; Sawaya, M. R.; Pensalfini, A.; Gou, D.; Berk, A. J.; Glabe, C. G.; et al. Out-of-Register β-Sheets Suggest a Pathway to Toxic Amyloid Aggregates. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20913−20918.
(49) Berhanu, W. M.; Hansmann, U. H. E. The Stability of Cylindrin β-Barrel Amyloid Oligomer Models—A Molecular Dynamics Study. *Proteins* **2013**, *81*, 1542−1555.
(50) Xie, L.; Luo, Y.; Wei, G. A β(16−22) Peptides Can Assemble into Ordered β-Barrels and Bilayer β-Sheets, while Substitution of Phenylalanine 19 by Tryptophan Increases the Population of Disordered Aggregates. *J. Phys. Chem. B* **2013**, *117*, 10149−10160.
(51) Mehta, A. K.; Lu, K.; Childers, W. S.; Liang, Y.; Dublin, S. N.; Dong, J.; Snyder, J. P.; Pingali, S. V.; Thiyagarajan, P.; Lynn, D. G. Facial Symmetry in Protein Self-Assembly. *J. Am. Chem. Soc.* **2008**, *130*, 9829−9835.
(52) Reddy, A. S.; Chopra, M.; de Pablo, J. J. GNNQQNY-Investigation of Early Steps during Amyloid Formation. *Biophys. J.* **2010**, *98*, 1038−1045.
(53) Hwang, W.; Zhang, S.; Kamm, R. D.; Karplus, M. Kinetic Control of Dimer Structure Formation in Amyloid Fibrillogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12916−12921.
(54) Meli, M.; Morra, G.; Colombo, G. Investigating the Mechanism of Peptide Aggregation: Insights from Mixed Monte Carlo-Molecular Dynamics Simulations. *Biophys. J.* **2008**, *94*, 4414−4426.
(55) Zhang, Z.; Chen, H.; Bai, H.; Lai, L. Molecular Dynamics Simulations on the Oligomer-Formation Process of the GNNQQNY Peptide from Yeast Prion Protein Sup35. *Biophys. J.* **2007**, *93*, 1484−1492.
(56) Berryman, J. T.; Radford, S. E.; Harris, S. A. Thermodynamic Description of Polymorphism in Q- and N-Rich Peptide Aggregates Revealed by Atomistic Simulation. *Biophys. J.* **2009**, *97*, 1−11.

(57) Park, J.; Kahng, B.; Hwang, W. Thermodynamic Selection of Steric Zipper Patterns in the Amyloid Cross-β Spine. *PLoS Comput. Biol.* **2009**, *5*.
(58) Collins, S. R.; Douglass, A.; Vale, R. D.; Weissman, J. S. Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. *PLoS Biol.* **2004**, *2*, e321.
(59) Narayanan, S.; Walter, S.; Reif, B. Yeast Prion-Protein, Sup35, Fibril Formation Proceeds by Addition and Substraction of Oligomers. *ChemBioChem* **2006**, *7*, 757−765.
(60) Petkova, A. T.; Leapman, R. D.; Guo, Z.; Yau, W.-M.; Mattson, M. P.; Tycko, R. Self-Propagating, Molecular-Level Polymorphism in Alzheimer's β-Amyloid Fibrils. *Science* **2005**, *307*, 262−265.
(61) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. Amyloid-β Protein Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology of Alzheimer's Disease. *Nat. Chem.* **2009**, *1*, 326−331.
(62) Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. Structural Conversion of Neurotoxic Amyloid-β_{1-42} Oligomers to Fibrils. *Nat. Struct. Mol. Biol.* **2010**, *17*, 561−567.