Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent

Yan Lu,[1] Philippe Derreumaux,[2] Zhi Guo,[1] Normand Mousseau,[3] and Guanghong Wei[1]*

[1] Department of Physics and National Key Surface Physics Laboratory, Fudan University, 220 Handan Road, Shanghai, 200433, China

[2] Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique et Université Paris 7,13 rue Pierre et Marie Curie, 75005 Paris, France

[3] Département de Physique and GÉPROM, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal (Québec), Canada

ABSTRACT

Aggregation of the full-length amyloid-β (Aβ) and β2-microglobulin (β2m) proteins is associated with Alzheimer's disease and dialysis-related amyloidosis, respectively. This assembly process is not restricted to full-length proteins, however, many short peptides also assemble into amyloid fibrils *in vitro*. Remarkably, the kinetics of amyloid-fibril formation of all these molecules is generally described by a nucleation-polymerization process characterized by a lag phase associated with the formation of a nucleus, after which fibril elongation occurs rapidly. In this study, we report using long molecular dynamics simulations with the OPEP coarse-grained force field, the thermodynamics and dynamics of the octamerization for two amyloid 7-residue peptides: the β2m83-89 NHVTLSQ and Aβ16-22 KLVFFAE fragments. Based on multiple trajectories run at 310 K, totaling 2.2 μs (β2m83-89) and 4.8 μs (Aβ16-22) and starting from random configurations and orientations of the chains, we find that the two peptides not only share common but also very different aggregation properties. Notably, an increase in the hydrophobic character of the peptide, as observed in Aβ16-22 with respect to β2m83-89 impacts the thermodynamics by reducing the population of bilayer β-sheet assemblies. Higher hydrophobicity is also found to slow down the dynamics of β-sheet formation by enhancing the averaged lifetime of all configuration types (CT) and by reducing the complexity of the CT transition probability matrix.

INTRODUCTION

Amyloid fibrillization is a generic feature of a large number of proteins with surprisingly little chain length similarity.[1] In spite of differences in amino acid compositions and experimental conditions, such as concentration and pH variations, these sequences all form final products displaying a common cross-β structure characterized by β-strands perpendicular and hydrogen bonds parallel to the fibril axis.[2] As it could be expected, these proteins seem to follow a similar two-step kinetics process for amyloid-fibril formation—characterized by a long lag phase associated with the formation of a nucleus followed by a rapid fibril elongation.[3] More remarkable is that these proteins also visit similar aggregates en route to fibrils including amorphous and annular aggregates and protofibrils, and that these soluble oligomers of various proteins and peptides display epitopes that can be recognized by the same antibodies.[4]

The high topological similarity of transient aggregates between different amyloid-forming proteins does not mean, however, that the nucleation process is totally independent of the amino acid length, especially in the very onset of aggregation. Photo-induced cross-linking of unmodified protein experiments on the amyloid β-protein (Aβ) have shown, for example, that in the early steps of aggregation, Aβ40 is in equilibrium between monomers, dimers, trimers, and tetramers whereas Aβ42 (resulting from the addition of the amino acids at the end of Aβ40) is in equilibrium between mostly monomers, pentamers, and heptamers.[5] Similarly, using both experiments and bioinformatics, it is understood that the location of both charged and hydrophobic amino acids[6] and the side chains details[7] are critical in accelerating or

slowing down the kinetics, and eventually preventing fibril formation. Even though recent Monte Carlo[8] and molecular dynamics simulations[9–12] have suggested that amyloid aggregation pathways are sequence-dependent, the full impact of hydrophobic variation on both the thermodynamics and dynamics of aggregation remains to be determined.

The aggregation of many amyloid-forming peptides has been studied experimentally and theoretically. The peptide sequences range from highly polar (GNNQQNY) to highly apolar (MVGGVV).[2] Here, we select two 7-residue peptides: the β2m83-89 fragment of sequence NHVTLSQ and the Aβ16-22 fragment KLVFFAE for two reasons.

Firstly, their amino acid compositions vary substantially: Aβ16-22 displays a continuous hydrophobic patch of five amino acids flanked by two charged amino acids, whereas β2m83-89, uncharged at neutral pH, is more hydrophilic and does not display any regular pattern of hydrophobic/hydrophilic amino acids. Although we know that Aβ16-22 forms in-register antiparallel β-sheets at pH 7 as determined by solid-state NMR experiments,[13] the β-sheet arrangement of β2m83-89 fibrils at pH 2 and 37 °C with 2.0 M NaCl remains to be determined.[14]

Secondly, both peptides have been studied in an attempt to characterize the critical nucleus size. Identification of the nucleus is very difficult by experiments since it is the highest free energy point along the aggregation pathways. It has been suggested from simulations and theoretical considerations using all-atom explicit[15] and implicit solvent simulations[16] that the nucleus size for Aβ16-22 certainly exceeds six chains and the nucleus for β2m83-89 might be larger than seven peptides.[17] Critical nuclei with sizes ranging from three to eight have been estimated theoretically for short peptides. For instance, the nucleus for GNNQQNY was found to vary from three to four chains[18] or four to five chains[19] (using all-atom REMD but no demonstration of convergence to equilibrium) to six chains[20] (using classical and quantum calculations). On the other hand, Hills and Brooks suggest a critical nucleus of five chains for STVIYE,[21] and Ma and Nussinov suggest that an octamer is the minimal oligomeric β-sheet size for the PrP fragment AGAAAAGA.[22] Finally, by using a tube model for peptides coupled to Monte Carlo simulations, Auer *et al.* find a critical nucleus of three chains for a 12-residue peptide.[12]

To understand the common dynamic and thermodynamic aggregation features shared by various peptides, it is essential to study several sequences of identical chain lengths using a single approach. Following previous studies leading to a maximal critical nucleus of eight for short peptides, we examine the aggregation of eight chains and report simulations of Aβ16-22 and β2m83-89 8-mers using the coarse-grained OPEP implicit solvent model coupled to molecular dynamics. It is to be noted that the size of such systems and the long timescale for oligomer growth or dissociation[23] preclude the use of all-atom MD simulations in explicit solvent.

MATERIALS AND METHODS

Molecular dynamics

Following our previous work on amyloid peptides, MD simulations are performed using a coarse-grained protein model.[24] All runs are carried out at 310 K using a 1 fs MD time step, Berendsen's thermostat with a coupling time of 0.5 ps enclosed in a sphere of diameter 60 Å with reflecting boundary conditions. The N- and C-termini of the peptides are unblocked and simulations are carried out at neutral pH since many systems including Aβ16-22 and the full-length β2m protein form fibrils over a wide range of pH conditions.

OPEP force field

OPEP uses a coarse-grained representation with all backbone N, H, Cα, C, and O atoms included and each side chain is modeled by a single particle with appropriate van der Waals radius and geometry.[25,26] The OPEP force field, which includes solvent effects implicitly, is expressed as a function of local terms (associated with the bond lengths, bond angles, improper torsions, and the backbone torsions), pair-wise contact potentials between all particles (considering the hydrophobic and hydrophilic character of all 20 amino acid side chains), and backbone two- and four-body (cooperative) hydrogen bonding interactions.

Here, we use the OPEP version 3.2 parameter set. Though a coarse-grained force field averages out most of the fast-changing degrees of freedom,[27] this parameter set was found to discriminate native from non-native protein structures,[28] to describe accurately the equilibrium structures of the Aβ21-30 peptide in solution[29] and to reproduce the thermodynamics and three-dimensional structures of soluble peptides with α, β and mixed α/β compositions.

Analysis

In all the analysis, we discard the first 10 ns of each MD run for β2m83-89 and the first 50 ns for Aβ16-22 in order to remove the memory of initial states.

The free energy surfaces were constructed using the relation $-RT \times \log(H(x,y))$, where $H(x,y)$ is the histogram of two selected order parameters x and y. Here, umbrella sampling and thermodynamic integration techniques are not necessary to enhance sampling because the long MD simulations overlap and equilibrium ensemble is obtained for both sequences. For instance, among the 16 independent runs on β2m83-89 started from disordered states, four explore a twisted bilayer β-sheet

structure, with Cα root mean square deviations (rmsd) of 1.5–3.0 Å. Convergence issue is also discussed in "Results" section.

To follow the assembly process and characterize the dominant structures, Cα-rmsd from selected structure and the fraction of native contacts are commonly used. However, the former is not an order parameter of choice here because small deviations in the peptide orientations can lead to large rmsd even when chains are relabeled to minimize the total rmsd. In addition, in contrast to protein folding, a native state cannot be defined since the octamer is inherently unstable and found, at equilibrium, to populate many distinct topologies. In this work, we identify the aggregates by introducing a topological parameter, the connectivity length (CL), defined as the sum over the square root of the β-sheet size and of the number of disordered chains in each configuration. Here, the β-sheet size is the number of strands in a x-stranded β-sheet. For example, the β-sheet size of a four-stranded β-sheet is four. This specific formulation of CL ensures that all different assemblies are assigned a different CL. Two chains are considered to form a β-sheet if (1) their end-to-end distances are greater than 11 Å; (2) at least two consecutive residues in each chain visit the β-strand state; (3) they have more than three intermolecular hydrogen bonds (H-bonds). One H-bond is taken as formed if the distance between N and O is less than 3.5 Å and the angle of N—H...O is greater than 150 degrees. In our notation, a bilayer consisting of two four-stranded β-sheets is assigned a connectivity length CL of sqrt(4) + sqrt(4) = 4 and an assembly of eight disordered chains without β-sheet by a CL of sqrt(1) × 8 = 8. Thus, the higher the connectivity length, the more disordered the aggregate is.

We also analyze the assemblies by their β-strand contents using STRIDE program,[30] the percentage of in-register and out-of-register β-strands, and the ratio of parallel/antiparallel β-sheets: two β-strands are considered as forming an antiparallel (parallel) β-sheet if the scalar product of their normalized end-to-end unit vectors is less than −0.7 (greater than 0.7).

Table I

Details of All the MD Simulations Used for Studying the Aggregation of 8-mers of β2m(83–89) and Aβ(16–22)

System	MD runs	Initial states	Time (ns)
β2m(83–89)	1–13	Random A	100 × 13
	14–15	Random B	300 × 2
	16	Random C	300 × 1
Aβ(16–22)	1–6	Random A	300 × 6
	7–12	Random B	300 × 6
	13–16	Random C	300 × 4

RESULTS

Table I gives a summary for the set-up details of the 32 MD runs reaching a total simulation time of 2.2 μs for β2m83-89 and 4.8 μs for Aβ16-22. All β2m83-89 simulations are carried out for 100 ns, with runs 14–16 extended to 300 ns. By contrast, each of the 16 runs on Aβ16-22 covers a timescale of 300 ns to compensate for a slower kinetics of forming well-ordered structures. We use three distinct starting compact states, as shown in Figure 1, with randomly chosen orientations and conformations of the peptide chains (i.e. free of any β-sheet signal) and different seeds for the initial velocity distributions. For brevity, in what follows we also use β2m and Aβ for β2m83-89 and Aβ16-22.

Clustering based on the connectivity length

Table II reports the topological clustering of MD-generated conformations in 10–100 ns and 50–300 ns time intervals for β2m83-89 and Aβ16-22, respectively, based on their connectivity lengths. We focus here on the 10 most-populated configuration types (CT), since they represent 91–97% of the conformations explored by each sequence. We have checked that the configuration types observed in simulations represent equilibrium distributions by repeating the same analysis using another set of trajectories, namely 13 among a total of 16 trajectories by excluding runs 2, 6, and 16 for both Aβ and β2m. Using this second dataset, we find that the populations
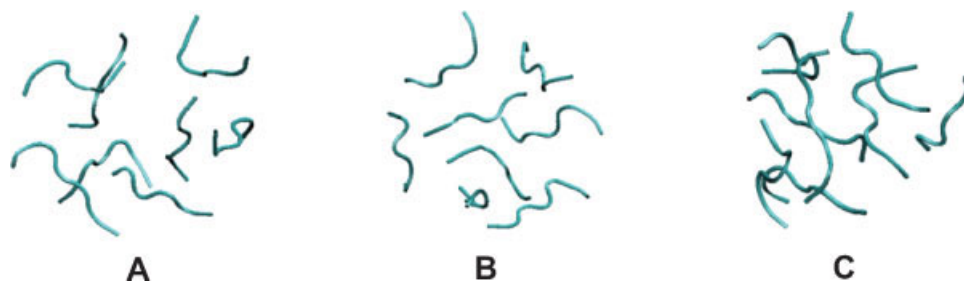

Figure 1

The three initial random states (A–C) used for the simulations of β2m83-89 and Aβ16-22 8-mers as defined in Table I. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table II
Percentage of Connectivity Length (CL) Of β2m(83–89) and Aβ(16–22) Aggregates Sampled in All the 16 MD Runs

Index	CL	Configuration type	Percentage	Sd (%)
β2m(83–89)				
1	4.732	4 + 3 + 1	18.5 (19.1)	1.1
2	6.146	3 + 2 + 1 + 1 + 1	13.2 (14.5)	0.9
3	5.414	4 + 2 + 1 + 1	12.7 (12.1)	0.7
4	4.000	4 + 4	11.4 (8.7)	2.1
5	5.464	3 + 3 + 1 + 1	9.5 (10.7)	0.7
6	6.000	4 + 1 + 1 + 1 + 1	8.2 (7.3)	0.6
7	6.732	3 + 1 + 1 + 1 + 1 + 1	7.8 (8.3)	0.7
8	6.828	2 + 2 + 1 + 1 + 1 + 1	4.0 (4.6)	0.4
9	7.414	2 + 1 + 1 + 1 + 1 + 1 + 1	3.1 (3.5)	0.5
10	5.236	5 + 1 + 1 + 1	2.6 (2.0)	0.3
Aβ(16–22)				
1	6.146	3 + 2 + 1 + 1 + 1	33.7 (31.1)	4.1
2	6.828	2 + 2 + 1 + 1 + 1 + 1	24.7 (23.7)	2.7
3	7.414	2 + 1 + 1 + 1 + 1 + 1 + 1	9.3 (6.3)	1.5
4	6.732	3 + 1 + 1 + 1 + 1 + 1	8.8 (7.2)	1.3
5	5.464	3 + 3 + 1 + 1	5.3 (8.5)	1.7
6	5.414	4 + 2 + 1 + 1	3.7 (5.5)	1.1
7	8.000	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	3.5 (1.4)	2.2
8	6.243	2 + 2 + 2 + 1 + 1	3.4 (4.8)	1.1
9	5.560	3 + 2 + 2 + 1	3.3 (5.1)	1.8
10	6.000	4 + 1 + 1 + 1 + 1	1.8 (2.5)	0.8

For β2m(83–89), the first percentage value uses 10–100 ns and the second in parenthesis uses 55–100 of all MD runs. For Aβ(16–22), the first percentage value uses 50–300 ns and the second in parenthesis uses 175–300 ns of all MD runs. Similarly, the standard deviation (sd) is calculated on 10–100 ns for β2m(83–89) and 50–300 ns for Aβ(16–22).

of the dominant topologies do not vary by more than 4%. The convergence was further checked by calculating the percentage of the 10 most-populated CT using the second half of the MD runs (i.e. 55–100 ns for β2m83-89 and 175–300 ns for Aβ16-22). The calculated CT percentage is given in the parenthesis in the fourth column of Table II. Comparing these results with those obtained in the full equilibrated runs, we find that the population difference of the dominant states between the two different time intervals is within 3.2% for both systems, indicating a reasonable convergence of our MD simulations. The standard deviation (sd) of the percentage over the 16 different runs for each configuration type is also given in the last column of Table II. These fluctuations are relatively small for all 10 dominant configuration types, showing that all runs converge to fairly similar distributions of various aggregates.

Comparing the connectivity lengths, we see that there is a significant overlap in the 10 most-visited configuration types between the two peptides. Not taking into account the occurrence probability, only configuration types 4 + 4, 4 + 3 + 1 and 5 + 1 + 1 + 1 with populations of 11.4%, 18.5% and 2.6% in β2m are absent in the top 10 of Aβ, being replaced by eight disordered chains (CL = 8, with probability 3.5%), three two-stranded β-sheets with two disordered chains (2 + 2 + 2 + 1 + 1, CL = 6.243, probability 3.4%), and a three-stranded β-sheet with two two-stranded β-sheets and one disordered

chain (3 + 2 + 2 + 1, CL = 5.56, probability 3.3%). These 4 + 4, 4 + 3 + 1, and 5 + 1 + 1 + 1 configuration types are, however, explored in Aβ simulations with probabilities of 0.1, 1.1, and 0.01%, respectively. The five most-populated aggregates in β2m (65%) and Aβ (82%) show that higher order configuration types dominate in β2m with mixtures of three- and four-stranded β-sheets (4 + 3 + 1, 18.5%). By contrast, Aβ aggregates are better described by a two- or three-stranded β-sheet generally interacting with a two-stranded β-sheet (2 + 2 + 1 + 1 + 1 + 1 and 3 + 2 + 1 + 1 + 1).

By analyzing the probability of the dominant β-sheet sizes calculated using all conformations shown in Figure 2, we find that the most-populated β-sheets for β2m are three-stranded (14%), followed by two- and four-stranded (12.5% each), whereas for Aβ, the probability decreases rapidly with the β-sheet size: two-stranded (21%), three-stranded (10%), and four-stranded (only 2%). The β-sheet propensity difference between Aβ and β2m is further illustrated by the probabilities of six- and seven-stranded β-sheets: low (less than 2%) but nonnegligible in β2m versus 0% in Aβ.

2D-free energy surfaces

Figures 3 and 4 show the free energy surfaces of β2m83-89 and Aβ16-22. Although it would have been preferable to use the same reaction coordinates for both sequences to build the two-dimensional (2D)-free energy landscape, we did not manage to find a set of coordinates that would allow, in both cases, a sufficient separation for the various topologies. For instance, if the bilayer β-sheet is explored in all simulations, it is a mixed anti-parallel/parallel twisted state with out-of-register strands in β2m (structure F in Fig. 3), whereas it is a flat antiparallel state with in-register strands in Aβ (structure G′ in Fig. 4) deviating by 8.8 Å rmsd from structure F.

As seen in Figure 3, there are six localized low free-energy basins and one delocalized region of higher free energy for β2m83-89 using the number of intermolecular H-bonds and the chain-independent and orientation-independent (i.e. neglecting the antiparallel and parallel character of the chains) Cα-rmsd calculated with respect to the bilayer β-sheet structure F—the most ordered structure—as order parameters. Note that the basins labeled from A to F belong to configuration types ranked 9th, 7th, 3rd, 5th, 1st, and 4th, respectively in Table II.

Basins A and B are amorphous in character and display either a two-stranded (A) or a three-stranded β-sheet (B) interacting with six or five disordered peptide chains, respectively. Basins C and D with an increasing number of H-bonds are associated with configurations types 4 + 2 + 1 + 1 (C) and 3 + 3 + 1 + 1 (D). In basins E and F, bilayer β-sheets of configuration type 4 + 3 + 1 (i.e. with one disordered chain) and 4 + 4 are observed, while the basin G of higher free energy, but of
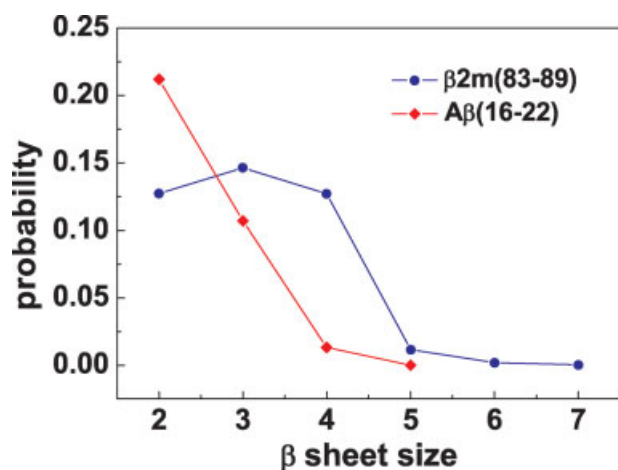
Figure 2

Normalized probability of the different β-sheet sizes observed in all the 16 MD runs of (**A**) β2m83-89 and (**B**) Aβ16-22 8-mers. [Color figure can be viewed in the online issue, which is available at www. interscience.wiley.com.]

large surface area contains many various configurations—notably the second most-populated configuration type 3 + 2 + 1 + 1 + 1 as shown in Figure 3. In line with the clustering based on the connectivity lengths, β2m 8-mers at 310 K are mainly characterized by amorphous aggregates with ordered two-, three-, and four-stranded β-sheets, and to a much lesser extent by β-sheet
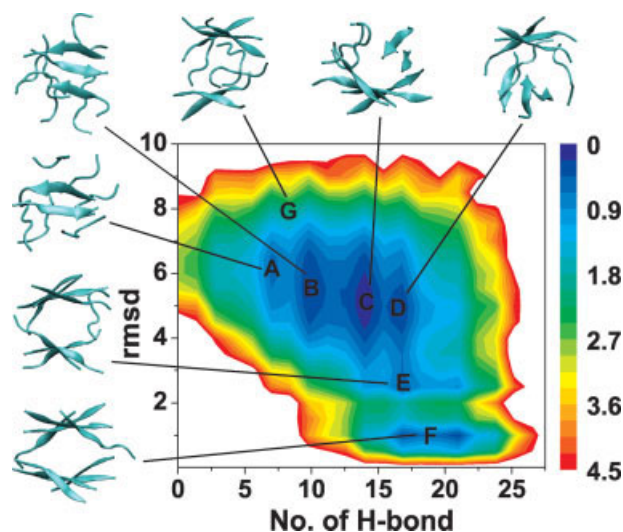
bilayers, with a population of 12%. These well-ordered bilayers include all conformations deviating by less than 3.0 Å rmsd from the structures F (CT of 4 + 4) and E (CT of 4 + 3 + 1); these results are insensitive to the precise cutoff value, increasing the threshold to 4.0 Å changes the population by less than 4%.

Figure 4 reports the same analysis for Aβ16-22 8-mers at 310 K using chain- and orientation-independent Cα rmsd calculated with respect to both the parallel bilayer β-sheet structure G′ and the perpendicular bilayer β-sheet structure E′, the two most ordered structures for this sequence; in particular, our MD generated fibril-like structure G′ is in good agreement with the Ma-Nussinov Aβ16-22 model (model D: antiparallel 1 in their Fig. 1),[31] with a Cα rmsd of 2.7 Å between the two structures. Basins A′ and B′ contain the amorphous configuration types ranked 3rd and 4th in Table II, characterized by a two-stranded β-sheet with six disordered chains (A′) and a three-stranded β-sheet with five disordered chains (B′). Basins D′, E′, and G′ correspond to antiparallel (G′) and perpendicular (D′, E′) bilayer β-sheets with antiparallel β-strands. These three localized basins are, however, marginally populated (1.5%). Finally, we show the position of two structures F′ and C′ associated with the very populated configuration types 3 + 2 + 1 + 1 + 1 and 2 + 2 + 1 + 1 + 1 + 1, ranked 1st and 2nd in Table II, respectively.

We stress that it is very difficult to find order parameters able to separate efficiently all configuration types
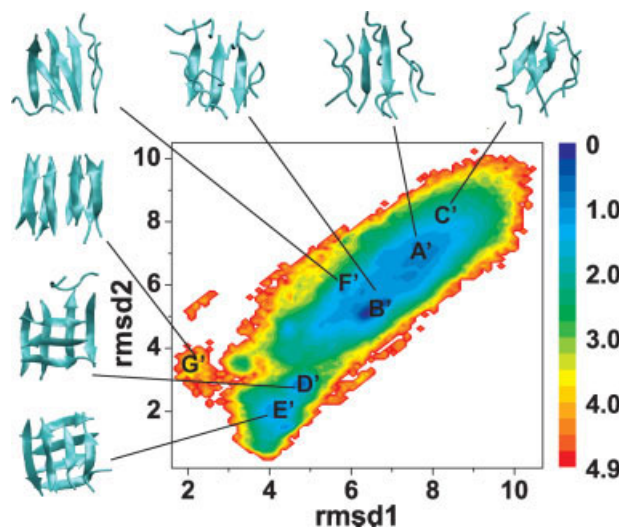

Figure 3

Free energy surface (in kcal/mol) of β2m83-89 8-mers at 310 K as a function of the number of intermolecular H-bonds and the Cα-rmsd (in Å) with respect to a reference bilayer structure (structure F). Representative ordered and disordered structures **A–H** are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Figure 4

Free energy surface (in kcal/mol) of Aβ16-22 8-mers at 310 K as a function of rmsd1: Cα-rmsd with respect to the reference parallel bilayer structure (structures G′) and rmsd2: Cα-rmsd with respect to the reference perpendicular bilayer structure (structure E′). Here, the Cα-rmsd is in Å. Representative ordered and disordered structures A′–F′ are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
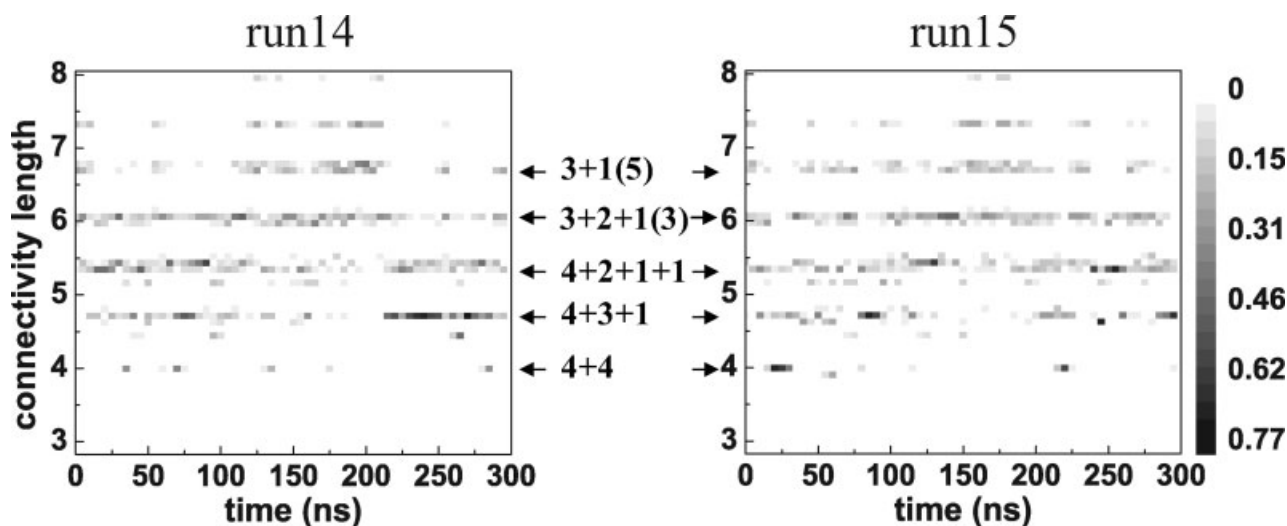
Figure 5

Time evolution of the normalized probabilities of the configuration types (CT) and connectivity lengths (CL) sampled in two simulations of β2m83-89 at 310 K. The white–gray–black bars give the normalized probability within a 5-ns time window. For simplicity, x disordered peptides in a given configuration type are denoted by 1(x).

sampled in the simulations. This aspect has already been discussed in a recent study using the hydrophobic core radius of gyration.[17] Here, we tried free energy representations based on principal component analysis,[32] with unsatisfactory results (data not shown). This indicates that some regions of our 2D-free energy surface can be unambiguously associated with specific configuration types (e.g. D′, E′, and G′ for Aβ), but other regions such as A′ and B′ in Aβ contain multiple configuration types. Finally, our orientation-independent Cα-rmsd parameter does not render the diversity of the detailed atomic structures, since it neglects the chains' antiparallel and parallel character. Averaged over all conformations, we find that Aβ16-22 has a stronger preference for in-register alignment of the β-strands than β2m83-89: 93% in Aβ versus 28% in β2m, and the percentage of fully antiparallel and mixed parallel/antiparallel β-sheets is 90% and 3% in Aβ versus 60% and 38% in β2m.

Dynamics of β-sheet association/dissociation

Because aggregation is a dynamical process, it is useful to characterize the time evolution of the octamers studied here. Figures 5 and 6 show the time series of the connectivity length for two trajectories of β2m83-89 and Aβ16-22. The normalized probability here is the number of configuration types satisfying a specific value of connectivity length CL divided by the total number of configurations saved within a consecutive 5-ns window. Results remain almost the same when using a consecutive 2-ns time window.

We first consider two β2m 300-ns trajectories, runs 14 and 15, starting from a disordered aggregate (CL = 8).

As shown in Figure 5, the dynamics varies between the two runs, but similar generic properties are easily observed. First, we see many transitions between low-ordered configuration types (e.g. 3 + 1 + 1 + 1 + 1 + 1) and higher ordered ensembles (e.g. 4 + 2 + 1 + 1 and 4 + 3 + 1) within the first 20 ns. A bilayer β-sheet structure (CL = 4) is first formed at 32 ns in run 14 and at 21 ns in run 15. This state is not stable, however, and subsequently dissociates/associates four times in run 14 versus twice in run 15 throughout the simulations. Interestingly, β2m is also able to re-explore the fully disordered states four times in run 14 and twice in run 15 between 100 and 220 ns. Overall, these two runs indicate, and this is supported by the 14 other runs, that β2m83-89 is almost in dynamic equilibrium between two extreme configuration types: fully disordered monomers (CL = 8) and bilayer β-sheets (CL = 4).

Figure 6 reports the same analysis for runs 1 and 16 of Aβ starting from two distinct disordered aggregates. These runs as well as the other 14 show that the dynamics is much slower in Aβ than in β2m. Within the first 25 ns, Aβ only explores very low ordered configuration types: 2 + 1 + 1 + 1 + 1 + 1 + 1 and 2 + 2 + 1 + 1 + 1 + 1. Slow dynamics is also indicated by the reduced number of transitions between the fully disordered and the higher ordered assemblies: CL of 8 is re-explored around 125 and 280 ns in run 1, while it is re-sampled in run 16 around 50 ns. This slow evolution prevents the Aβ octamer from extensively visiting all states even in 300-ns runs, although there seems to be strong preference for two-layered β-sheets.

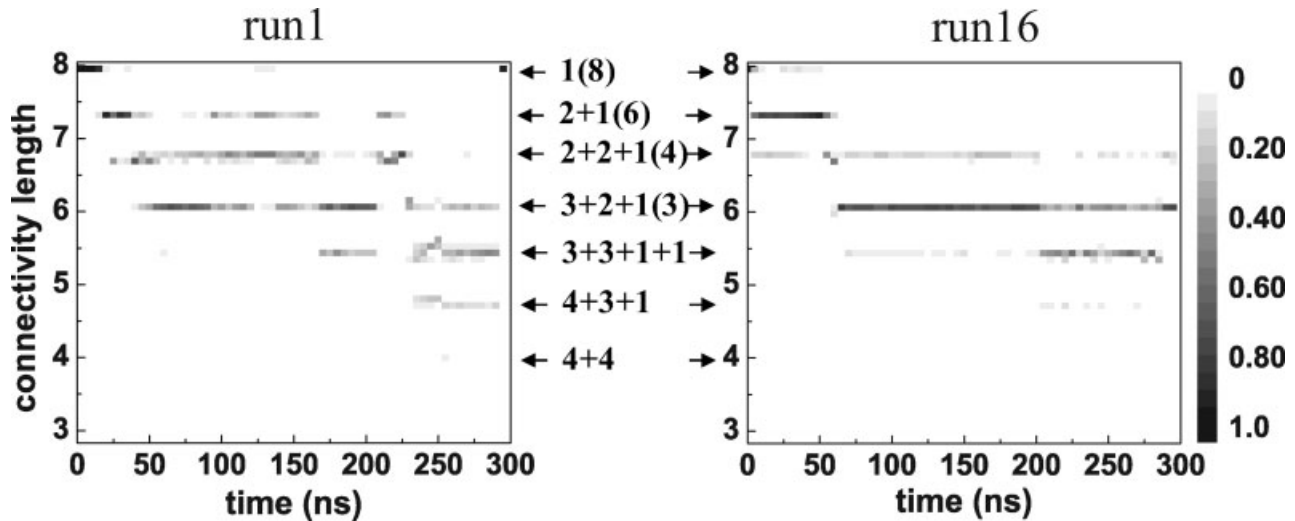Figure 7 shows the transition probability matrix between the 10 dominant MD-generated configuration

Figure 6

Time evolution of the normalized probabilities of the configuration types (CT) and connectivity lengths (CL) sampled in two simulations of Aβ16-22 at 310 K. The white–gray–black bars give the normalized probability within a 5-ns time window. For simplicity, x disordered peptides in a given configuration type are denoted by 1(x).

types (note the different scale for both sequences). For this, we compute the number of transitions from one configuration type to another in all the 16 MD runs. The calculation distinguishes between i --> j and j --> i transitions. The *y*-axis is the final state and the *x*-axis is the initial state. Although the forward and backward transitions are very similar, they are actually different. We clearly see that this asymmetric transition matrix (almost perfectly symmetric by visualization) is much simpler in Aβ than in β2m. This simplicity and the approximate
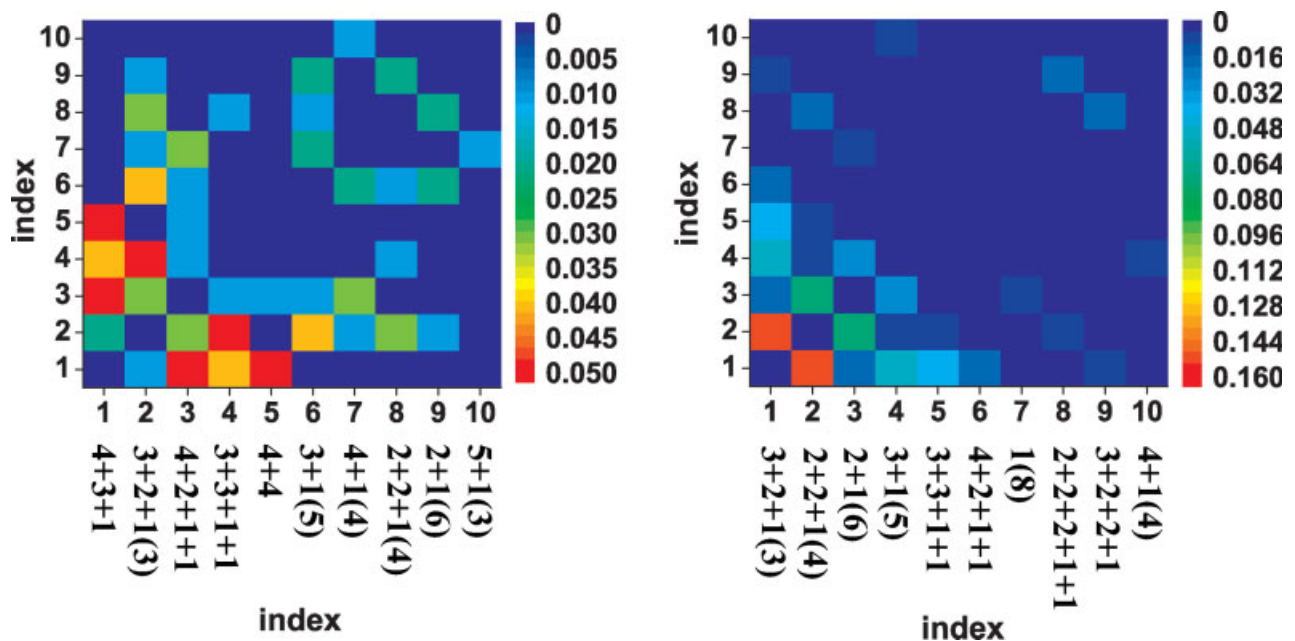

Figure 7

Transition probability matrices between the 10 dominant configurations types. These are obtained by averaging over all 16 simulations of β2m83-89 (left) and Aβ16-22 (right). Here, the configuration types are indexed, and we give below each index the configuration type. Note, as reported in Table II, the description of the 10 CT's varies from β2m83-89 to Aβ16-22, and x disordered peptides in a given configuration type are denoted by 1(x).

symmetry of the transition matrix reflect the existence of reversible transitions between two different states and indicate the convergence of our MD simulations. We see two configurational transitions dominate in Aβ: 3 + 2 + 1 + 1 + 1 <--> 2 + 2 + 1 + 1 + 1 + 1, each with a probability of 14% and associated with the reversible evolution of one two-stranded β-sheet into a three-stranded β-sheet in contact with another dimeric β-sheet. The matrix is rather sparse: we observe 28 transitions with a probability higher than 0.5% among a total of 90 possible transitions, and 14 forbidden transitions.

For β2m, we see six equally populated transitions (each of 5%) involving high-ordered configurations: 4 + 3 + 1 <--> 4 + 2 + 1 + 1, 4 + 3 + 1 <--> 4 + 4 and 3 + 2 + 1 + 1 + 1 <--> 3 + 3 + 1 + 1. In contrast to Aβ, we observe 44 transitions using a 0.5% probability cutoff and only four forbidden transitions. It is beyond the scope of this study to characterize in details the transition rate between each configuration type, the formation/dissociation process of each configuration type, and the lifetime distribution of each configuration type for both sequences. Nonetheless, we find that the averaged lifetimes for all the configuration types of Aβ and β2m are 0.6 and 0.2 ns, respectively. The longest lifetime for β-sheet aggregates are 1.9 and 0.7 ns in Aβ and β2m and they are associated with 3 + 1 + 1 + 1 + 1 + 1 configuration type and the bilayer β-sheet, respectively.

DISCUSSION AND CONCLUSIONS

While amyloid formation is observed for a wide range of sequences, external conditions, such as pH, temperature and ionic strength, and internal sequence variations are known to affect the growth kinetics and thermodynamics. The relation between these variations and amyloid growth is not clear, however. For example, it has been shown that amino acid substitutions cause variations in fibril formation[6] and stability[33] only at specific positions and both tolerant and restrictive positions to mutation can be found within an amyloid sequence. There is also both theoretical and experimental evidence that the *in vitro* aggregation rates depend on a delicate balance between many factors, including hydrophobic interactions, electrostatic interactions and propensity for β-strand and α-helix.[7,34] This active research field has led to the emergence of several techniques such as FOLDX,[35] PASTA[36] for predicting the amyloid-prone regions from the sequences.

Although fully polar sequences such NNQQ form fibrils and molecular charges contribute to the aggregation process, the hydrophobic effects play an essential role in the early steps of aggregation.[37,38] Moreover, hydrophobic cooperativity has been proposed as a mechanism for amyloid nucleation.[21] Hydrophobic effects on

aggregation are often studied computationally by using a single peptide model and varying the strength of the hydrophobic interaction relative to that of the hydrogen-bonding interaction.[39,40] Recently, their effects were also analyzed by studying the aggregation of two different chain length peptides using a Monte Carlo approach.[8]

In this study, we have explored the thermodynamics and dynamics of the octamerization for two amyloid-forming peptides of identical chain length but different hydrophobicity. The use of a coarse-grained implicit solvent model certainly simplifies the free energy surface with respect to an all-atom explicit solvent model, but we are more interested in generic properties rather than quantitative effects. Moreover, the use of a simplified representation makes it possible to explore equilibrium thermodynamical ensembles on a time scale that is not reachable with standard all-atom representations, providing us well converged properties. Our findings can be summarized as follows.

Firstly, an increase in hydrophobicity changes the free energy surface of 8-mers, by stabilizing and destabilizing specific topologies and changing the ratio of in-register/out-of-register β-strands and parallel/antiparallel β-sheets. These findings are consistent with high temperature MD simulations aimed at understanding the stability of protofibrillar models upon amino acid mutations[41] and landscape topology graph analysis that an increment of hydrophobicity increases the population of metastable monomeric states.[42] Of particular importance for peptide oligomerization is that, at equilibrium, the population of amyloid-competent states consisting of parallel, antiparallel, and perpendicular β-sheets, and the population of β-sheets of large size consisting of five, six, and seven β-strands change drastically from Aβ16-22 to β2m83-89 8-mers. While the same topologies are accessible to both sequences, except for topologies consisting of six- or seven-stranded β-sheet, it is their respective weight that varies.

Secondly, we find that increasing hydrophobicity impacts the dynamics of well-ordered aggregate formation: Aβ16-22 is a slow aggregating peptide, whereas β2m83-89 is a very fast aggregating peptide. Though there is not a unique aggregation mechanism,[43–45] we find that an increase in the hydrophobic character of the peptides enhances the averaged lifetime of amorphous or partially ordered aggregates and thus slows down the kinetics of forming well-ordered aggregates, and at the same time reduces the complexity of the transition probability matrix between all configuration types. In particular, the presence of large hydrophobic patches introduces an energetic frustration that traps oligomers in a large number of disordered states. Our dynamical picture is consistent with the difference in nucleation processes observed in Langevin dynamics simulations of 125 peptide chains with low and high β-prone sequences.[46,47] Since the difference in dynamics was directly connected

to the energy difference between the amyloid-competent (i.e. β-strand) and amyloid-protected states of the monomer, we hypothesize that the former state is more populated in β2m83-89 than in Aβ16-22. Our simulations are also in agreement with Monte Carlo simulations of multiple chains of Aβ16-22 and the hydrophilic Aβ25-35 peptides, although these peptides have different chain lengths.[8] It was shown that Aβ16-22 peptides assemble into amorphous oligomers that subsequently undergo a process of conformational conversion leading to highly ordered β-sheet species, whereas Aβ25-35 peptides, of higher hydrophilic character with respect to Aβ16-22, form very rapidly ordered aggregates with very high β-sheet content.

Finally, using a rough estimate provided by Rao et al.[48] for the sampling speed-up factor—about 100—associated with the use of an implicit solvent over an explicit solvent, we predict that the longest lifetime of all β-sheet sizes should be in the hundreds of nanoseconds range in explicit solvent. We also note that the population of bilayer β-sheets does not increase from 7- to 8-mers of β2m83-89 and we did not find evidence that any octameric state is sufficiently stable to behave as a nucleus in both Aβ16-22 and β2m83-89. Taken together these results indicate that the critical nucleation size from which rapid fibril growth occurs is very likely larger than eight chains for both peptides.

In summary, our long molecular dynamics simulations show that an increase in the hydrophobic character of the peptide, especially as a hydrophobic patch, as observed in Aβ16-22 with respect to β2m83-89, introduces an element of frustration that increases the proportion of visited amorphous structures and decreases the population of well-ordered β-sheet assemblies, effectively introducing energy traps that slow down the dynamics of β-sheet formation by enhancing the averaged lifetime of all the aggregates. This finding is fully consistent with and provides a microscopic confirmation of the observation from positional scanning mutagenesis that most sequences leading to a large amount of fibrils polymerize slowly.[6]

ACKNOWLEDGMENTS

REFERENCES

1. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 2006;75:333–366.
2. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT, Madsen AO, Riekel C, Eisenberg D. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature 2007;447:453–457.
3. Grant MA, Lazo ND, Lomakin A, Condron MM, Arai H, Yamin G, Rigby AC, Teplow DB. Familial Alzhemier's disease mutations alter the stability of the amyloid beta-protein monomer folding nucleus. Proc Natl Acad Sci USA 2007;104:16522–16527.
4. Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, Glabe CG. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science 2003;300: 486–489.
5. Bitan G, Teplow DB. Rapid photochemical cross-linking—a new tool for studies of metastable, amyloidogenic protein assemblies. Acc Chem Res 2004;37:357–364.
6. López de la Paz M, Serrano L. Sequence determinants of amyloid fibril formation. Proc Natl Acad Sci USA 2004;101:87–92.
7. Dubay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. J Mol Biol 2004;341:1317–1326.
8. Cheon M, Chang I, Mohanty S, Luheshi LM, Dobson CM, Vendruscolo M, Favrin G. Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. PLoS Comput Biol 2007;3:1727–1738.
9. Huet A, Derreumaux P. Impact of the mutation A21G (Flemish variant) on Alzheimer's beta-amyloid dimers by molecular dynamics simulations. Biophys J 2006;91:3829–3840.
10. Zheng J, Jang H, Nussinov R. Beta2-microglobulin amyloid fragment organization and morphology and its comparison to Abeta suggests that amyloid aggregation pathways are sequence specific. Biochemistry 2008;47:2497–2509.
11. Liang C, Derreumaux P, Mousseau N, Wei G. The beta-strand-loop-beta-strand conformation is marginally populated in beta2-microglobulin (20–41) peptide in solution as revealed by replica exchange molecular dynamics simulations. Biophys J 2008;95:510–517.
12. Auer S, Dobson CM, Vendruscolo M. Characterization of the nucleation barriers for protein aggregation and amyloid formation. HFSP J 2007;1:137–146.
13. Petkova AT, Buntkowsky G, Dyda F, Leapman RD, Yau WM, Tycko R. Solid state NMR reveals a pH-dependent anti-parallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. J Mol Biol 2004;335:247–260.
14. Ivanova MI, Sawaya MR, Gingery M, Attinger A, Eisenberg D. An amyloid-forming segment of beta 2-microglobulin suggests a molecular model for the fibril. Proc Natl Acad Sci USA 2004;101:10584–10589.
15. Nguyen PH, Li MS, Stock G, Straub JE, Thirumalai D. Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. Proc Natl Acad Sci USA 2007;104: 111–116.
16. Irback A, Mitternacht S. Spontaneous beta-barrel formation: an all-atom Monte Carlo study of Abeta(16–22) oligomerization. Proteins 2008;71:207–214.
17. Song W, Wei GH, Mousseau N, Derreumaux P. Self-assembly of the beta 2-microglobulin NHVTLSQ peptide using a coarse-grained protein model reveals beta-barrel species. J Phys Chem B 2008;112:4410–4418.
18. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D. Structure of the cross-beta spine of amyloid-like fibrils. Nature 2005;435:773–778.
19. De Simone A, Esposito L, Pedone C, Vitagliano L. Insights into stability and toxicity of amyloid-like oligomers by replica exchange molecular dynamics analyses. Biophys J 2008;95:1965–1973.
20. Tsemekhman K, Goldschmidt L, Eisenberg D, Baker D. Cooperative hydrogen bonding in amyloid formation. Protein Sci 2007;16:761–764.
21. Hills RD, Brooks CL. Hydrophobic cooperativity as a mechanism for amyloid nucleation. J Mol Biol 2007;368:894–901.

22. Ma B, Nussinov R. Molecular dynamics simulations of alanine rich beta-sheet oligomers: insight into amyloid formation. Protein Sci 2002;11:2335–2350.

23. Takeda T, Klimov DK. Dissociation of Abeta(16–22) amyloid fibrils probed by molecular dynamics. J Mol Biol 2007;368:1202–1213.

24. Derreumaux P, Mousseau N. Coarse-grained protein molecular dynamics simulations. J Chem Phys 2007;126:025101.

25. Derreumaux P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J Chem Phys 1999;111:2301–2310.

26. Floquet N, Pasco S, Ramont L, Derreumaux P, Laronze JY, Nuzillard JM, Maquart FX, Alix AJP, Monboisse JC. The antitumor properties of the alpha 3(IV)-(185–203) peptide from the NC1 domain of type IV collagen (tumstatin) are conformation-dependent. J Biol Chem 2004;279:2091–2100.

27. Derreumaux P, Wilson KJ, Vergoten G, Peticolas WL. Conformational studies of neuroactive ligands. I. Force field and vibrational spectra of crystalline acetylcholine. J Phys Chem 1989;93:1338–1350.

28. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. Proteins 2007;69:394–408.

29. Chen W, Mousseau N, Derreumaux P. The conformations of the amyloid-beta (21–30) fragment can be described by three families in solution. J Chem Phys 2006;125:084911.

30. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins 1995;23:566–579.

31. Ma BY, Nussinov R. Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (Abeta(16–22) Abeta(16–35) and Abeta(10–35)): sequence effects. Proc Natl Acad Sci USA 2002;99:14126–14131.

32. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11:205–217.

33. López de la Paz M, de Mori GMS, Serrano L, Colombo G. Sequence dependence of amyloid fibril formation: insights from molecular dynamics simulations. J Mol Biol 2005;349:583–596.

34. Caflisch A. Computational models for the prediction of polypeptide aggregation propensity. Curr Opin Chem Biol 2006;10:437–444.

35. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res 2005;33:W382–W388.

36. Trovato A, Seno F, Tosatto SC. The PASTA server for protein aggregation prediction. Protein Eng Des Sel 2007;20:521–523.

37. Mousseau N, Derreumaux P. Exploring the early steps of amyloid peptide aggregation by computers. Acc Chem Res 2005;38:885–891.

38. Melquiond A, Gelly J-C, Mousseau N, Derreumaux P. Probing amyloid fibril formation of the NFGAIL peptide by computer simulations. J Chem Phys 2007;126:065101–065107.

39. Marchut AJ, Hall CK. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. Biophys J 2006;90:4574–4584.

40. Yun SJ, Urbanc B, Cruz L, Bitan G, Teplow DB, Stanley HE. Role of electrostatic interactions in amyloid beta-protein (Abeta) oligomer formation: a discrete molecular dynamics study. Biophys J 2007;92:4064–4077.

41. Zheng J, Ma BY, Tsai CJ, Nussinov R. Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. Biophys J 2006;91:824–833.

42. Auer S, Miller MA, Krivov SV, Dobson CM, Karplus M, Vendruscolo M. Importance of metastable states in the free energy landscapes of polypeptide chains. Phys Rev Lett 2007;99:178104.

43. Harrison PM, Chan HS, Prusiner SB, Cohen FE. Conformational propagation with prion-like characteristics in a simple model of protein folding. Protein Sci 2001;10:819–835.

44. Santini S, Wei GH, Mousseau N, Derreumaux P. Pathway complexity of Alzheimer's beta-amyloid Abeta(16–22) peptide assembly. Structure 2004;12:1245–1255.

45. Santini S, Mousseau N, Derreumaux P. In silico assembly of Alzheimer's Abeta(16–22) peptide into beta-sheets. J Am Chem Soc 2004;126:11509–11516.

46. Pellarin R, Caflisch A. Interpreting the aggregation kinetics of amyloid peptides. J Mol Biol 2006;360:882–892.

47. Pellarin R, Guarnera E, Caflisch A. Pathways and intermediates of amyloid fibril formation. J Mol Biol 2007;374:917–924.

48. Rao F, Settanni G, Guarnera E, Caflisch A. Estimation of protein folding probability from equilibrium simulations. J Chem Phys 2005;122:184901.