

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/5511661>

# Self-Assembly of the $\beta$ 2-Microglobulin NHVTL SQ Peptide Using a Coarse-Grained Protein Model Reveals a $\beta$ -Barrel Species

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MAY 2008

Impact Factor: 3.3 · DOI: 10.1021/jp710592v · Source: PubMed

---

CITATIONS

41

---

READS

11

4 AUTHORS, INCLUDING:



Philippe Derreumaux

French National Centre for Scientific Research...

167 PUBLICATIONS 3,809 CITATIONS

SEE PROFILE

# Self-Assembly of the $\beta$ 2-Microglobulin NHVTL SQ Peptide Using a Coarse-Grained Protein Model Reveals a $\beta$ -Barrel Species

Wei Song and Guanghong Wei\*

Department of Physics, Fudan University, Shanghai 200433, China

Normand Mousseau

Département de Physique and Regroupement, Québécois sur les Matériaux de Pointe, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal, Québec, Canada

Philippe Derreumaux

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

Received: November 4, 2007; In Final Form: December 27, 2007

Although a wide variety of proteins can assemble into amyloid fibrils, the structure of the early oligomeric species on the aggregation pathways remains unknown at an atomic level of detail. In this paper we report, using molecular dynamics simulations with the OPEP coarse-grained force field, the free energy landscape of a tetramer and a heptamer of the  $\beta$ 2-microglobulin NHVTL SQ peptide. On the basis of a total of more than 17 ns trajectories started from various states, we find that both species are in equilibrium between amorphous and well-ordered aggregates with cross- $\beta$ -structure, a perpendicular bilayer  $\beta$ -sheet, and, for the heptamer, six- or seven-stranded closed and open  $\beta$ -barrels. Moreover, analysis of the heptamer trajectories shows that the perpendicular bilayer  $\beta$ -sheet is one possible precursor of the  $\beta$ -barrel, but that this barrel can also be formed from a twisted monolayer  $\beta$ -sheet with successive addition of chains. Comparison with previous aggregation simulations and the fact that nature constructs transmembrane  $\beta$ -sheet proteins with pores open the possibility that  $\beta$ -barrels with small inner diameters may represent a common intermediate during the early steps of aggregation.

## Introduction

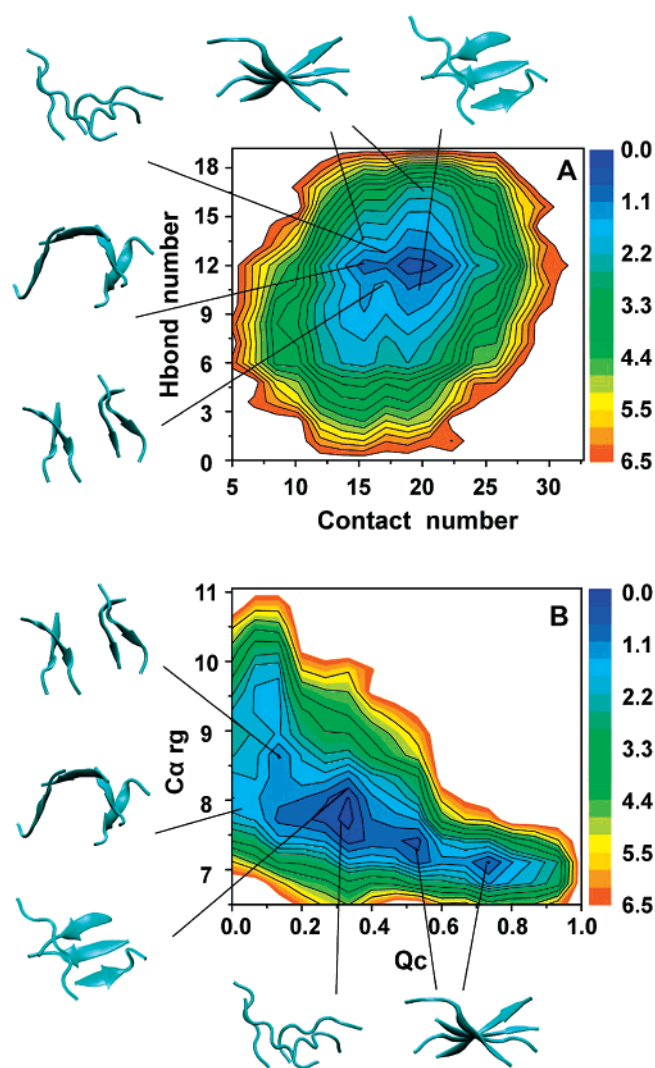
Formation of amyloid plaques is a hallmark of more than 20 neurodegenerative diseases such as Alzheimer's and Parkinson's diseases as well as dialysis-related amyloidosis. In spite of extensive experimental and theoretical studies, the early steps of aggregation are still mostly unknown at an atomic level of detail not only for the disease-related proteins but also for the short peptides identified in the last two decades. Attention has turned to this process in large part because the soluble aggregates are more cytotoxic than mature fibrils and these aggregates formed from various proteins and peptides are recognized by the same antibodies.<sup>1,2</sup> Understanding the first steps of peptide aggregation could therefore offer an answer to the molecular origin of the toxicity as well as some indications of what dynamical and structural properties are specific and generic for the various sequences.

Beyond these questions, understanding the structure and the formation of soluble aggregates is essential to explain the long lag phase observed before the appearance of fibrils and the nature of the critical nucleus.<sup>3–5</sup> Photoinduced cross-linking of unmodified proteins (PICUP) experiments on the amyloid  $\beta$ -protein ( $A\beta$ ) have shown, for example, that  $A\beta$ 40 exists in solution as a mixture of monomers, dimers, trimers, and tetramers while  $A\beta$ 42 is found mostly in pentameric and

heptameric forms,<sup>5</sup> suggesting that the critical nucleus is larger than four or seven for the  $A\beta$  peptide. Similarly, Fay et al. find also, from the variation of the lag phase as a function of concentration, a critical nucleus of six for the prion protein Ure2p.<sup>4</sup>

These numbers are not easily confirmed numerically. Nelson et al. proposed, from simple energetic considerations, that the minimal nucleus seed for fibril formation of the GNNQQNY peptide consists of only three to four.<sup>6</sup> Classical and quantum calculations by Tsemekhman et al.<sup>7</sup> on GNNQQNY indicate that the energy of adding a two-peptide layer to a preformed  $n$ -layer fibril (consisting of  $2n$  peptides) becomes constant when  $n \geq 3$ , suggesting a minimal nucleus of six. While these calculations provide information regarding the internal energy of ideal structures at 0 K, they fail to include the entropic effects that can be very important in the nucleation process. With a partial inclusion of entropy and an all-atom description, Hills and Brooks find the critical nucleus for fibril formation to be five for STVIYE.<sup>8</sup> The critical impact of entropy is also supported by the work of Pellarin and Caflisch,<sup>9</sup> who studied the aggregation of a peptide model adopting either amyloid-competent or amyloid-protected states. Even though each chain has only very limited motion, the authors show that, unless the sequence is a strong  $\beta$ -former, off-pathway intermediates will form and the critical nucleus can become much larger than four. Interestingly, this result is also obtained when more realistic empirical potentials are used. Nguyen et al.,<sup>10</sup> for example, find

\* To whom correspondence should be addressed. Phone: (86) 21 55665231. Fax: (86) 21 65104949. E-mail: ghwei@fudan.edu.cn.



**Figure 1.** Free energy surfaces (kcal/mol) of the  $\beta$ 2m(83–89) 4-mer at 310 K as a function of (A) the number of interpeptide side chain contacts and number of interpeptide H-bonds and (B) the fraction of native interpeptide side chain contacts ( $Q_c$ ) and radius of gyration ( $r_g$ ) of the C $\alpha$  atoms.

that the locking time for an A $\beta$ (16–22) monomer keeps increasing even in the presence of a preformed pentamer. All-atom MD simulations on a prion-derived eight-residue peptide by Ma and Nussinov suggest that an octamer is the minimal  $\beta$ -sheet oligomeric size.<sup>11</sup> These apparently conflicting results can be explained, in part, by Hills and Brooks as discussed above and the discrete molecular dynamics simulations of Nguyen and Hall,<sup>12</sup> who showed, in agreement with numerous experimental studies, that the presence of a perfect preformed fibrillar state reduces the phase lag for fibril growth to zero. However, the probability of forming such a perfect aggregate is small, and other routes, with larger but defective aggregates, might be preferred in the absence of seeding.

To progress on that issue, it is necessary to obtain a precise picture of the free energy surface of small oligomers to extract the relative thermodynamical weight of the various structures. Given the long lag phase observed experimentally,<sup>4</sup> numerical studies are challenging and molecular dynamics simulations with explicit solvent remain too slow for extensive studies of large systems. All-atom MD or REMD simulations with explicit solvent, for example, can be run over an integrated time of 400–500 ns at best for dimers and tetramers of A $\beta$ (16–22)<sup>13,14</sup> and

$\beta$ 2m(83–89).<sup>15</sup> Moreover, as was shown recently,<sup>16</sup> convergence to equilibrium scales with the number of degrees of freedom and is difficult to obtain using replica exchange explicit solvent MD for these systems. To improve sampling, it is therefore useful to work with simplified models using various force fields and approaches.<sup>17–23</sup>

In this study, we use the recently developed coarse-grained protein molecular dynamics protocol, MD–OPEP, to determine the free energy surface of a tetramer and a heptamer of an amyloid-forming peptide. Comparing the various topologies obtained from the two oligomers, we can identify trends and common structures suggesting a critical nucleus or building blocks that might play an important role in fibrillization. MD–OPEP was used for calculating the free energy surface of the A $\beta$ (16–22) dimer<sup>24</sup>—which was found to be very similar to that obtained by Gnanakaran et al. using a full-atomic description and explicit solvent.<sup>14</sup> The coarse-grained OPEP force field has also been applied successfully to protein folding,<sup>25</sup> de novo peptide structure prediction,<sup>26</sup> and aggregation of model systems.<sup>27</sup>

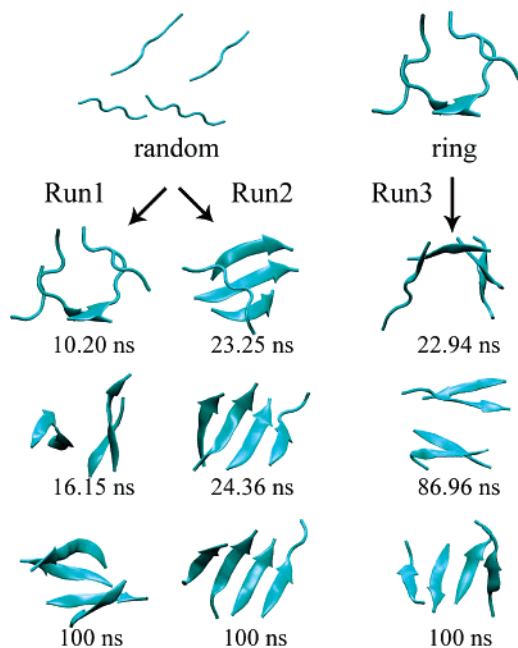
Many short peptides have been extensively studied. Here, we use the seven-residue  $\beta$ 2m(83–89) peptide for two main reasons. First, the fragment 83–89 of amino acid sequence NHVTLSQ forms fibrils in vitro.<sup>28</sup> Second, dimerization has been examined using implicit and explicit solvent all-atom MD simulations.<sup>15</sup> In what follows, we describe the protocol coupling MD simulations and the OPEP force field and how we analyze the trajectories and generate a free energy surface for four and seven  $\beta$ 2m(83–89) chains at 310 K. The main findings are then summarized. Notably, we find multiple free energy minima with cross- $\beta$  and amorphous signals that are similar between the tetramer and the heptamer, but we also find that these structures are in equilibrium with open and closed  $\beta$ -barrels for the heptamer.

## Materials and Methods

**Molecular Dynamics.** MD simulations at neutral pH were performed with our in-house package, Simulateur,<sup>21</sup> using the OPEP version 3.2 force field<sup>29</sup> at a constant temperature of 310 K in a sphere diameter of 60 Å with reflecting boundary conditions. The temperature was controlled by the Berendsen's bath with a coupling constant of 0.1 ps. For the four-chain system of the unblocked NHVTLSQ peptide, we perform twenty 100 ns MD simulations starting from a random state (15 runs) and a ringlike structure (five runs). For its part, the seven-chain system is subject to a total of 51 runs using various initial states for a simulation time varying between 100 and 300 ns each. The integration time step is 1 fs, and all bonded and nonbonded interactions are updated at every step.

**OPEP.** OPEP implements a coarse-grained representation with all backbone N, H, C $\alpha$ , C, and O atoms included and each side chain modeled by a single particle with appropriate van der Waals radius and geometry.<sup>30,31</sup> The OPEP energy function, which includes solvent effects implicitly, is expressed as a function of three types of interactions: local terms associated with the bond lengths, bond angles, improper torsions, and the backbone torsions, pairwise contact potentials between main chain and main chain, side chain and main chain, and side chain and side chain particles (considering all 20 amino acid types), and backbone two-body and four-body (cooperative) hydrogen-bonding interactions.<sup>30,32</sup>

The OPEP3.2 set of parameters has been shown to display capability very similar to the all-atom discrete optimized protein energy (DOPE) model.<sup>29</sup> This parameter set has also been found



**Figure 2.** Three OPEP-MD simulations of the tetramer at 310 K starting from a randomly chosen state (run 1 and run 2) or the ringlike state generated by run 1 (run 3). Representative snapshots are shown between 0 and 100 ns.

to describe accurately the free energy surface of the  $A\beta(21-30)$  monomer.<sup>33</sup> These results indicate that, although the competition with solvent is not treated explicitly and the complexity of the side chains is reduced, OPEP captures the physical forces stabilizing the native forms of proteins with various secondary structure contents. Since OPEP includes the solvent effects in an implicit manner, MD-OPEP simulations sample the phase space faster than all-atom MD with explicit solvent.

**Analysis.** The free energy surfaces are constructed using the relation  $-RT \log(H(x,y))$ , where  $H(x,y)$  is the histogram of two selected reaction coordinates,  $x$  and  $y$ . Umbrella sampling and thermodynamic integration techniques are not necessary because equilibrium is easily reached. Analysis of the 20 runs on the tetramer and the 51 runs on the heptamer shows multiple association-dissociation events. We also find an excellent (a good) overlap between the trajectories on the tetramer (on the heptamer). Given that the runs are independent, the fact that the trajectories visit, from time to time, the same part of the phase space indicates that the sampling of the phase space is ergodic and that the free energy surface we compute is at or near its equilibrium value.

To follow the assembly process or characterize the dominant aggregates, we monitor the secondary structure composition using the STRIDE program,<sup>34</sup> the radius of gyration ( $r_g$ ), the chain-independent C $\alpha$  root-mean square deviation (rmsd) from a selected structure (i.e., the numbering of the chains is not considered), and the side chain-side chain contacts. Two side chains,  $k$  and  $l$ , of van der Waals  $R_k$  and  $R_l$ <sup>29</sup> are in contact if they deviate by less than  $R_k + R_l + 1$  Å.<sup>35</sup> Two chains are considered as forming an antiparallel (parallel)  $\beta$ -sheet if the scalar product of their normalized end-to-end unit vectors is less than  $-0.7$  (greater than  $0.7$ ) and at least two consecutive residues of each chain visit the  $\beta$ -strand state. One hydrogen bond (H-bond) is considered as formed if the distance between N and O is less than 3.5 Å and the angle N-H $\cdots$ O is greater than 150°. All representations of the aggregates are prepared using the VMD program.<sup>36</sup>

**TABLE 1: Details of the 51 MD Simulations Used for Studying the Aggregation of 7-mers of  $\beta 2m(83-89)$ <sup>a</sup>**

	initial state	time (ns)		initial state	time (ns)
runs 1–15	random A	100 $\times$ 15	runs 38–46	barrel A	100 $\times$ 7
runs 16–27	random B	100 $\times$ 12			290 $\times$ 2
runs 28–29	random C	100 $\times$ 2	runs 47–51	barrel B	290 $\times$ 4
runs 30–37	bilayer A	100 $\times$ 8			350 $\times$ 1

<sup>a</sup> For each run, we indicate the starting conformation (also shown in Figure 3) and the simulation time.

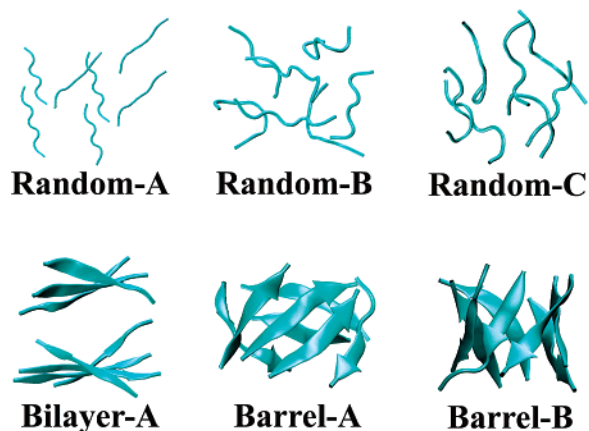
As a further step toward characterization of the aggregates, we first calculated the chain-independent rmsd between all conformations. The structure with the largest number of neighbors deviating by less than 4.0 and 3.0 Å rmsd (for the heptamer and tetramer, respectively) was taken, together with the neighbors, to form the first cluster and eliminated from the pool of structures. This process was repeated until there was no structure left in the pool. However, because the rmsd is very sensitive to the orientation of the chains and the registries of H-bonds, we further classify the clusters containing typical  $\beta$ -sheet structures into topologies by neglecting the orientation and alignment of the  $\beta$ -strands. Specifically, for the seven-chain system, we use the following topological parameters for defining the parallel and orthogonal bilayer  $\beta$ -sheets and the  $\beta$ -barrels. Two  $\beta$ -sheets are considered as forming a bilayer  $\beta$ -sheet if the two criteria given in the caption of Supporting Information Figure 1 are satisfied. A bilayer  $\beta$ -sheet is considered parallel (orthogonal) if the absolute value of the cosine angle between any two strands belonging to either layer is larger (smaller) than 0.7 (0.3). Finally, for  $\beta$ -barrels, we use the number ( $n$ ) of consecutive  $\beta$ -strands (six or seven are found here) and the shear number  $S$ <sup>37,38</sup> defined by  $S = [(2\pi R)^2 - (nb)^2]^{1/2}/a$ , where  $R \approx 6.5$  Å,  $a = 3.3$  Å, and  $b = 4.4$  Å.

## Results

**Tetramer.** Figure 1A shows the free energy surface of the tetramer generated from 2.0  $\mu$ s MD-OPEP at 310 K as a function of the total number of side chain-side chain (SC) contacts and the number of H-bonds, a set of reaction coordinates often used. Most structures visited have a very similar number of interpeptide SC contacts (between 15 and 22) and H-bonds (between 11 and 14) even though their 3D shape varies considerably. To show this, we plot in Figure 1B the free energy surface projected on the fraction of native SC contacts and the  $r_g$  measured on the C $\alpha$  atoms. Here, the native state refers to an ideal twisted parallel monolayer  $\beta$ -sheet. Using  $r_g$  and the native contacts, we see a better separation of the structures, but it is still not sufficient to really identify the dominant basins.

Using a topologically based clustering of all the conformations, we find five structural families. These include, from left to right in Figure 1A, two parallel  $\beta$ -sheets, ringlike conformations, random coil aggregates, a twisted monolayer  $\beta$ -sheet with parallel chains, and a three-stranded  $\beta$ -sheet stabilized by one extended chain located above the plane of the trimer. Amorphous aggregates constitute the dominant topology with a population of 66% and include random coil and ringlike conformations. For their part, the aggregates with high  $\beta$ -sheet contents include the parallel bilayer  $\beta$ -sheet and the monolayer  $\beta$ -sheet with populations of 7.7% and 26.5%, respectively. We emphasize that all these topologies display structural variability as measured by the orientation of the chains or the register of H-bonds. The ordered bilayer  $\beta$ -sheet with C $\alpha$  $\cdots$ C $\alpha$  distances





**Figure 3.** Initial structures of the seven  $\beta$ 2m(83–89) peptides used for simulations.

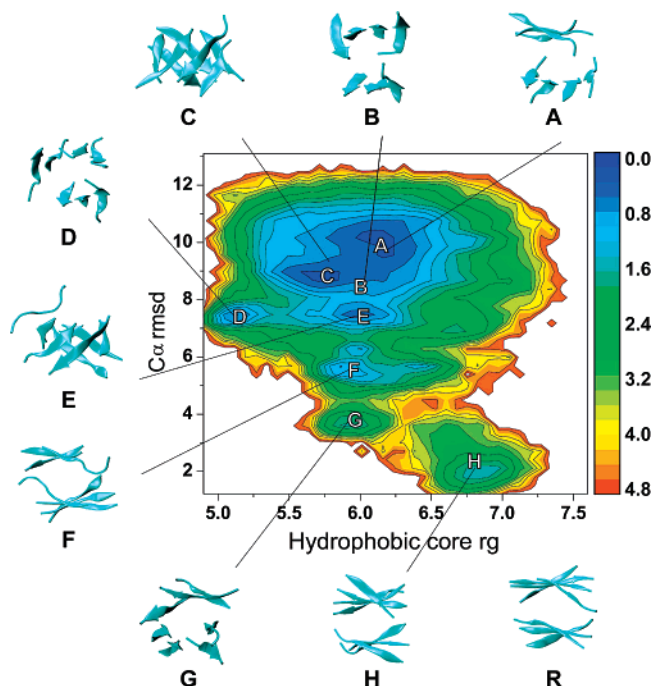
of 5.0 Å between the strands and 11.0 Å between the layers is consistent with the crystal structure of the peptide GNNQQNY.<sup>6</sup> For their part, the closed and open rings have outer diameters of about 13 Å. Interestingly, all these organizations were observed on multiple chains of KFFE,<sup>39</sup> A $\beta$ (16–22),<sup>21</sup> and A $\beta$ -(11–25) using various simulation methods, suggesting the universality and independence with respect to the exact amino acid composition.

Figure 2, which shows four representative snapshots from three 100 ns MD simulations, is a good indicator of the overlap between the different trajectories. Starting from a randomly chosen state, four chains can assemble into a four-stranded  $\beta$ -sheet structure by passing through a ringlike conformation which moves into a bilayer  $\beta$ -sheet (run 1) or through a three-stranded  $\beta$ -sheet + monomer (run 2). For its part, run 3 starting from a ringlike conformation shows that the system can explore a bilayer before a monolayer: there is a constant flux from one ordered structure to the others.

Interestingly, Ivanova et al. also studied a scrambled version of  $\beta$ 2m(83–89), QVLHTSN, which under the conditions used for the wild-type sequence, did not form fibrils.<sup>28</sup> To address this aspect, we performed 10 MD simulations of 100 ns on this variant starting from random states. We found that one simulation explores only amorphous aggregates within 100 ns, and averaged over 10 runs, the percentage of  $\beta$ -sheet drops from 43% in the wild-type sequence to 24% in the scrambled version. Notably also the percentage of bilayer  $\beta$ -sheets is 0.0% for the scrambled version vs 7.7%, indicating that an important impact of the scrambled sequence lies in the impossibility to populate two parallel, antiparallel, or orthogonal  $\beta$ -sheets. Although these results must be considered as qualitative only, they reinforce the capability of OPEP to recognize amyloid from nonamyloid sequences.<sup>40</sup>

**Heptamer.** Following the tetramer, we first describe the potential of mean force (PMF) and the structures of the heptameric species found. Then we present a detailed analysis of the pathways leading to  $\beta$ -barrels starting from randomly chosen states or allowing the dissociation of  $\beta$ -barrels.

**Dominant Free Energy Structures.** Table 1 gives a summary of the setup details of the 51 MD runs, reaching a total simulation time of 15  $\mu$ s. Runs 1–44 were carried out for 100 ns, and runs 45–51 were extended to 300 ns on average. The initial configurations, shown in Figure 3, are random with no  $\beta$ -sheet signal for runs 1–29 and consist of preformed  $\beta$ -barrels (barrel A and barrel B) or a bilayer  $\beta$ -sheet (bilayer A) for runs 30–51. The three random (A, B, C) starting points with



**Figure 4.** Free energy surface (kcal/mol) of the  $\beta$ 2m(83–89) 7-mers at 310 K as a function of the radius of gyration of the hydrophobic core and the C $\alpha$  rmsd with respect to a reference bilayer structure (state R). Representative ordered structures A–H are shown.

randomly chosen orientations and conformations of the chains were generated by a 30 ns MD simulation at 600 K starting from a bilayer  $\beta$ -sheet structure. The accumulation of statistics starting from these different initial states allows us to make sure that the system is equilibrated by verifying that the various trajectories overlap. Among the total of 29 simulations, starting from random states, we find that seven and six runs explore the bilayer  $\beta$ -sheet and the  $\beta$ -barrel, respectively. The overlap is not only topological but also structural. For example, the bilayer  $\beta$ -sheets observed in runs 17 and 22 deviate by 3.5 Å rmsd, and the  $\beta$ -barrels observed in runs 12 and 25 deviate by 2.9 Å rmsd.

Nevertheless, since trajectories starting from an ordered structure show a long equilibration time, we construct the free energy landscape (see Figure 4) from runs 1–29 only, starting from random conformations with different initial random seeds. We also reject the first 1 ns of all trajectories and compute the free energy surface and other averages using the remaining 99 ns. The reaction coordinates used to build the two-dimensional free energy landscape are the hydrophobic core radius of gyration ( $r_g$ ) and the C $\alpha$  rmsd calculated with respect to the parallel bilayer  $\beta$ -sheet structure R shown in Figure 4. This reference state, R, generated by our MD simulations, is composed of one antiparallel  $\beta$ -strand and six parallel  $\beta$ -strands. R was chosen for comparison, rather than its ideal counterpart with seven parallel chains, because it allows for a better separation of the various free energy basins. The hydrophobic core is computed using the valine and leucine residues of all chains.

The choice of reaction coordinates is not unique, and we tried many other sets that were found to be insensitive to the details of the chain organization. For example, projection on the hydrophilic core radius of gyration—involving all glutamine, histidine, threonine, serine, and glutamine residues—is found to depend mostly on the density of the aggregates and cannot discriminate between ordered and disordered structures. Simi-

larly, the fraction of C $\alpha$  contacts with respect to the state R is not well adapted for multichain systems characterized by thousands of minima and leads to a PMF almost flat along this coordinate. Indeed, all the structures generated by the first 29 simulations can be clustered into a total of 1900 and 1100 states using all and the five central residues of all chains, respectively, and a chain-independent C $\alpha$  rmsd cutoff of 4 Å, although many of them share equivalent topologies.

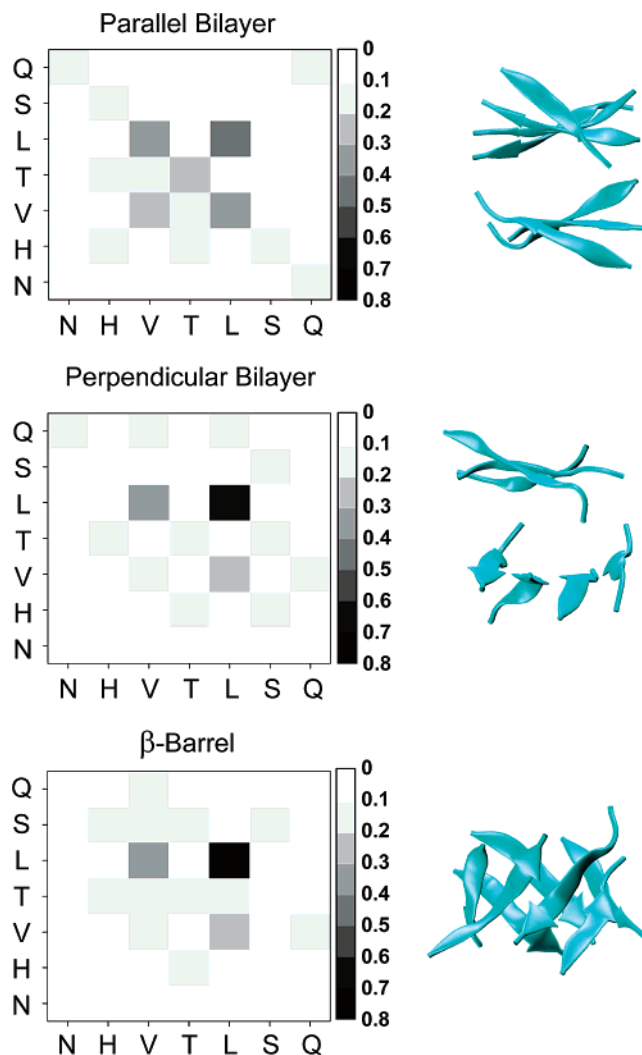
As for the tetramer, the heptamer PMF plot we obtain with the reaction coordinates discussed above reveals both disordered and ordered aggregates at 310 K. Amorphous aggregates are the most populated type (with a population of 64%). Ordered aggregates include three topologies: parallel bilayer  $\beta$ -sheets (states B, D, F, and H), orthogonal bilayer  $\beta$ -sheets (state A) and  $\beta$ -barrels (states C and E), with populations of 19.8%, 4.6%, and 12.3%, respectively.

States B, D, F, and H all display the intrasheet C $\alpha$ –C $\alpha$  distance of 4.5 Å and the intersheet distance of 10–12 Å. They are, however, located in four different low free energy basins and therefore display distinct hydrophobic core  $r_g$  and C $\alpha$  rmsd values. State G consists of a curved, disordered five-stranded  $\beta$ -sheet with a ringlike shape interacting with a partially formed two-stranded  $\beta$ -sheet. Finally, the  $\beta$ -barrel can be characterized by a six (state E) or seven (state C) stranded curved  $\beta$ -sheet. According to the formula given in the section “Analysis”, the shear numbers  $S$  are 10 and 8 for state E ( $n = 6$ ) and state C ( $n = 7$ ), respectively.

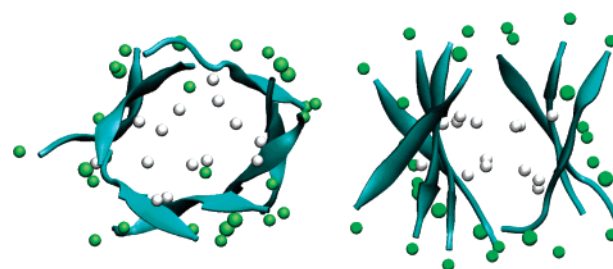
Diversity in the detailed atomic structures of each topology with high  $\beta$ -sheet content is also reflected by the observation of in-register and out-of-register alignments of the  $\beta$ -strands and mixed antiparallel/parallel  $\beta$ -strands with no strong preference for one orientation in all topologies (the population of parallel: antiparallel chains is 1.3:1.0), except in the bilayer state H, which displays only parallel  $\beta$ -strands. This slight preference for parallel arrangement is different from that observed in all-atom implicit and explicit solvent MD simulations of the dimer, where antiparallel  $\beta$ -sheets were found more stable than parallel  $\beta$ -sheets.<sup>15</sup> It is unclear whether this difference results from various pH conditions (pH 2 in all-atom MD vs pH 7 in this study) or the use of main-chain restraints to facilitate the formation of ordered dimers in all-atom MD<sup>15</sup> or is dependent on the size of the oligomeric species studied. Overall, our dynamic picture, with mixed orientated chains and different H-bond registries, is, however, consistent with previous all-atom MD simulations of amyloid peptides in settings varying from 3-mers to 9-mers.<sup>39,41–43</sup>

To identify which residue has a significant role in stabilizing the  $\beta$ -barrel and the parallel and perpendicular bilayer  $\beta$ -sheets, we compute the formation probability of each side chain–side chain interaction averaged over all structures belonging to the same cluster. The contact maps shown in Figure 5 reveal that the three topologies are stabilized by different patterns of side chain–side chain interactions, but they share three pairs of contacts: the leucine–leucine contact, present in 80% of the conformations of the perpendicular bilayer and  $\beta$ -barrel, and the double leucine–valine contact found in 30% of the conformations. This strong similarity between the contact maps of the perpendicular bilayer  $\beta$ -sheet and the  $\beta$ -barrel is a first indication that they may form from each other, as discussed further below.

Figure 6 shows the position of the side chains of the hydrophobic and hydrophilic residues in the twisted bilayer  $\beta$ -sheet and the  $\beta$ -barrel. Both topologies protect the valine and leucine amino acids from solvent exposure as is expected from



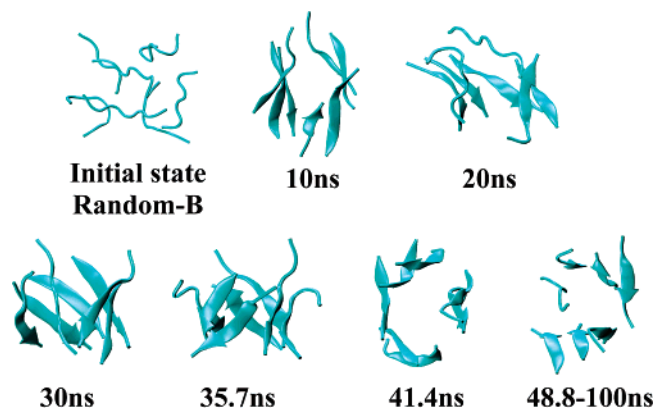
**Figure 5.** Contact map probability averaged over the seven  $\beta$ 2m(83–89) chains to form interpeptide side chain contacts in the parallel bilayer, perpendicular bilayer, and  $\beta$ -barrel structures.



**Figure 6.** Atomic descriptions of the  $\beta$ -barrel and bilayer structures found. The hydrophobic and hydrophilic side chains are in gray and green, respectively.

the contact map. This result, along with the number of H-bonds, indicates that both species optimize their backbone–backbone and side chain–side chain interactions.

Further analysis of the PMF indicates that all the dominant aggregates are separated by small free energy barriers that are on the order of 1  $kT$ . These barriers clearly depend on the reaction coordinates used,<sup>44</sup> but we have not found any significant changes by projecting on the first two principal components obtained from Cartesian principal component analysis. Moreover, these barriers are consistent with the kinetic information obtained from the MD trajectories and the fact that



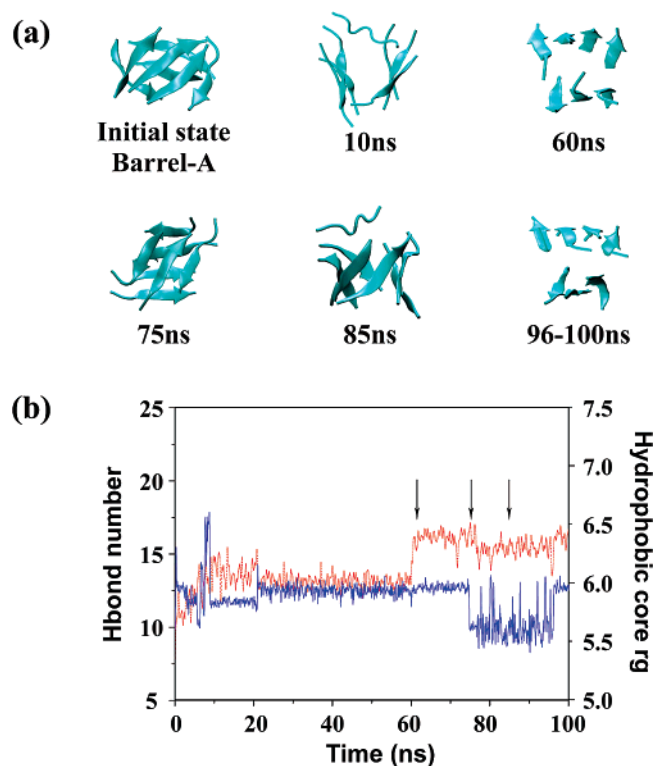
**Figure 7.** Representative snapshots generated in one trajectory showing a  $\beta$ -barrel formation starting from a random state.

the system is not trapped in any state that is overwhelmingly more stable and moves almost freely from one ordered state to another, with transition times often less than 10 or 20 ns.

**Formation and Dissociation of  $\beta$ -Barrel Aggregates.** Although the aim of this work is to identify the dominant species of equilibrated oligomers, we can combine the free energy surface and the trajectories to get insights into the pathways leading to  $\beta$ -barrel aggregates, a topology particularly interesting since it is well-organized and does not match the cross- $\beta$  structure.

Representative snapshots for run 1 are shown in Figure 7. Starting from a random organization of seven chains, the peptides assemble into two  $\beta$ -sheets with four and three twisted and misaligned  $\beta$ -strands at 10 ns that convert into a twisted five-stranded sheet interacting with two random chains at 20 ns. Subsequently, the two free chains diffuse around the preformed  $\beta$ -sheet structure, and at 30 ns, one free chain is associated with the  $\beta$ -sheet, allowing the formation of a twisted six-stranded  $\beta$ -sheet. From there, the last free chain and the preformed  $\beta$ -sheet explore the conformational space until the formation of a closed  $\beta$ -barrel ( $t = 35.7$  ns). The  $\beta$ -barrel in this run is not stable and continues to evolve into two  $\beta$ -sheets of five and two strands ( $t = 41.4$  ns) and then into a bilayer  $\beta$ -sheet of six strands interacting with a disordered chain ( $t = 100$  ns). Although some variations are observed in the other three runs forming a  $\beta$ -barrel starting from the random states A, B, and C, they follow closely the same pathway involving the successive addition of one chain to a twisted  $\beta$ -sheet.

The transition time from the closed  $\beta$ -barrel to bilayer  $\beta$ -sheet aggregates varies significantly. Figures 8 and 9 show the analysis of MD runs 38 and 47 starting from two closed  $\beta$ -barrels, A and B, deviating by less than 3 Å rmsd from each other. In the barrel A trajectory (Figure 8), the closed barrel remains stable for less than 5 ns and a transition is observed between a parallel-like bilayer  $\beta$ -sheet ( $t = 60$  ns) and an open six-stranded  $\beta$ -barrel ( $t = 85$  ns) through a perpendicular bilayer intermediate ( $t = 75$  ns). As seen in Figure 8b, this topological change between 60 and 85 ns takes place with little variation in the number of H-bonds and hydrophobic core radius of gyration. In contrast to the barrel A trajectory, the close barrel B is preserved for 65 ns in run 47 (Figure 9a) and then evolves into two  $\beta$ -sheets with four-stranded and two-stranded chains surrounded by a mobile chain ( $t = 118$  ns). This transition now occurs through an important decrease of H-bonds and a slight decrease of the hydrophobic core radius of gyration (Figure 9b).



**Figure 8.** Analysis of MD run 38 starting from  $\beta$ -barrel A: (a) representative snapshots, (b) evolution of the number of H-bonds (in red) and hydrophobic core radius of gyration (in blue) as a function of time. The three arrows indicate three time points of 60, 75, and 85 ns.

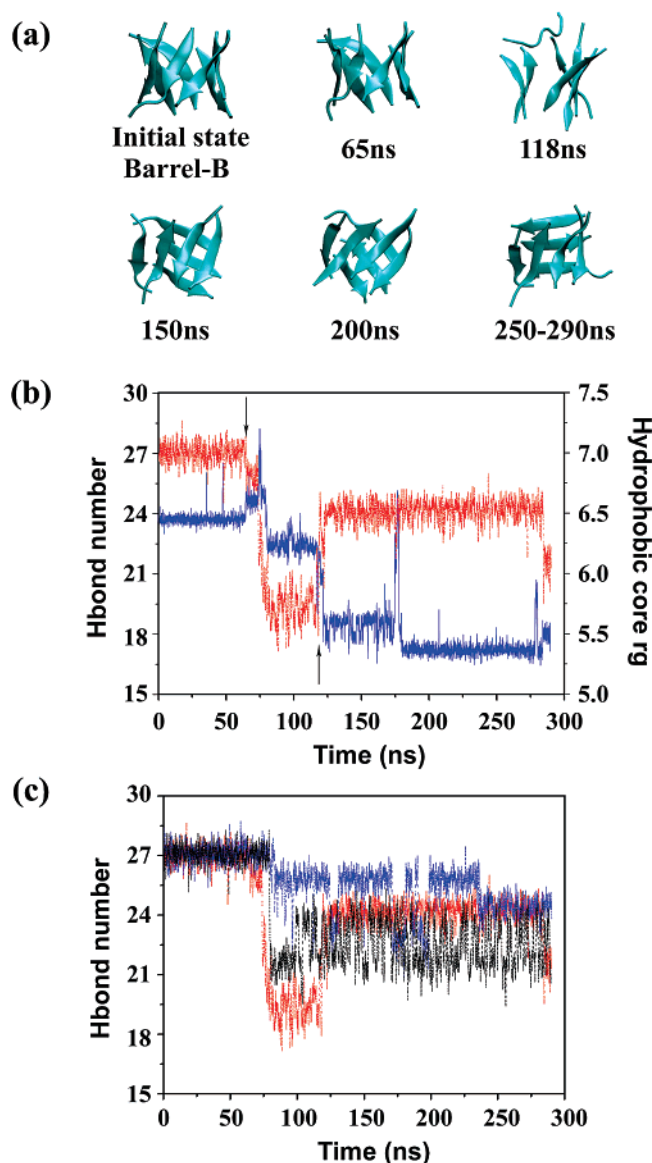
The destabilization of the closed seven-stranded  $\beta$ -barrel is further examined in Figure 9c, where the time evolution of the number of H-bonds are plotted for runs 47, 48, and 49 starting from the same barrel B, but using different random seeds. We see that the closed barrel B is destabilized after 65–84 ns in the three runs and then explores distinct topologies according to the simulations. In run 48 (blue), the peptides essentially fluctuate between open  $\beta$ -barrels (number of H-bonds 25) and a bilayer  $\beta$ -sheet within 80–125 ns, while in run 49 (black) the closed  $\beta$ -barrel is transformed into parallel and perpendicular bilayer  $\beta$ -sheets in equilibrium within 84–130 ns. This diversity of time scales and dissociation routes is consistent with the multiple routes seen in the present free energy surface and previous on-lattice and off-lattice amyloid-forming peptide simulations.<sup>22,23,35,45</sup>

## Discussion and Conclusion

The free energy surfaces of the  $\beta$ 2m(83–89) tetramer and heptamer, calculated using long, multiple MD simulations with the OPEP coarse-grained protein model, offer a clear picture of the first steps of aggregation. With 20 and 51 overlapping trajectories, respectively, each of 100 ns or more, starting from many different conformations, we obtain well-converged results sampling a large number of distinct structures. The seven peptides populate 1900 distinct clusters using a chain-independent C $\alpha$  rmsd cutoff of 4.0 Å. It is possible, however, to obtain a more meaningful classification by turning to a topological analysis which, although it neglects the differences between parallel and antiparallel arrangements of the chains, allows us to follow more closely the important elements of the kinetic and thermodynamical evolution.

Although the present pH condition (pH 7) varies from the experiments carried out by Ivanova et al.<sup>28</sup> at pH 2 (affecting





**Figure 9.** Analysis of MD runs starting from  $\beta$ -barrel B: (a) representative snapshots, (b) time evolution of the number of H-bonds (in red) and hydrophobic core radius of gyration (in blue) from run 47, (c) H-bond numbers from runs 47 (red), 48 (blue), and 49 (black). The arrow points to the value of the H-bond number at  $t = 65$  ns when the  $\beta$ -barrel starts to break.

the protonation state of the histidine side chain), many peptides from full-length  $\beta$ 2m including fragments  $\beta$ 2m(21–31),<sup>46</sup>  $\beta$ 2m(59–71),<sup>47</sup> and  $\beta$ 2m(20–41)<sup>48</sup> form amyloid fibrils in vitro under a wide range of pH values and notably at neutral pH. Moreover, fibrillization of full-length  $\beta$ 2m protein was also observed in vitro at 310 K and pH 7.<sup>49</sup> All together these experimental results suggest that our MD-based analysis should not vary much at pH 2, although we cannot exclude the possibility that the registry of H-bonds may be affected by the pH conditions, as was observed for the  $A\beta$ (11–25) fragment.<sup>50</sup>

Our results show that the four and the seven peptides visit structures that are fully compatible and are in rapid equilibrium between cross- $\beta$ , amorphous, and, for the heptamer,  $\beta$ -barrel aggregates. The structure of the  $\beta$ -barrel is not unique, and its topology includes closed six- and seven-stranded  $\beta$ -barrels and open barrels. Clearly, these results do not support a critical nucleus for this system, between four and seven chains, at least as it is often defined in numerical work. Indeed, we do not find

any evidence of a specific oligomeric topology that is particularly stable or prevalent, but rather a large ensemble of disordered and ordered structures in dynamic equilibrium. Only a simulation with a much larger number of chains would allow us, in the spirit of the classical nucleation theory and following the work of Nguyen and Hall,<sup>12</sup> to fully write down the nucleation process.

The cross- $\beta$  structures and the amorphous aggregates found here display morphological characteristics that have been observed for many peptides and proteins en route to an amyloid fibril by electron microscopy and atomic force microscopy.<sup>51,52</sup> On the other hand,  $\beta$ -barrels made of six or seven short peptide chains have never been reported experimentally. Only annular or ringlike species with inner diameters larger than 15–20 Å have been observed for many proteins. It was estimated for instance that 40–60 full-length  $A\beta$  peptides or eight PrP(90–232) prion proteins are required to form these annular structures.<sup>52,53</sup> This result raises the question as to whether the  $\beta$ -barrel described here can serve as a metastable building block for further polymerization, by allowing the formation of larger barrels through successive addition of monomers or collision and coalescence of several small barrels. Simulations of higher order species, and 16-mers, are currently under investigation to address this issue. However, the present simulations make it clear that the perpendicular bilayer  $\beta$ -sheet is one possible precursor of the  $\beta$ -barrel, and this barrel can also be formed from a twisted monolayer  $\beta$ -sheet with successive addition of chains.

The observation of  $\beta$ -barrel topologies from numerical simulations is not totally new. This topology has previously been reported during the self-assembly of 6-mers and 7-mers of KFFE using the activation–relaxation technique,<sup>39,54</sup> 24-mers of 16-residue polyglutamine chains using discontinuous molecular dynamics,<sup>55</sup> and six  $A\beta$ (16–22) peptide chains using all-atom Monte Carlo simulations.<sup>20</sup> The present  $\beta$ 2m barrel differs from the KFFE barrel in that it can be fully closed and the hydrophobic core is well protected from solvent. It also differs from the polyglutamine peptide annular structure, which is not regular. On the other hand, it is very similar to the six-stranded  $\beta$ -barrel formed by the  $A\beta$ (16–22) peptide, which has the same chain length as the  $\beta$ 2m(83–89) peptide. It is noted that the six-chain  $\beta$ 2m(83–89)  $\beta$ -barrels are observed in six of the 29 MD simulations from random organizations, while the six-chain  $A\beta$ (16–22)  $\beta$ -barrel is found in only one of the 30 Monte Carlo simulations attempted.<sup>20</sup> This indicates that the population and formation time of  $\beta$ -barrels are likely to vary with amino acid composition, chain length, and solvent conditions. However, the observation of the  $\beta$ -barrel topology using four distinct sequences and simulation protocols and the fact that nature constructs transmembrane proteins such as OmpW of  $\beta$ -barrel character with a solvent-protected hydrophobic core and a hydrophobic gate closing the channel<sup>56</sup> suggest that the  $\beta$ -barrel may represent a common intermediate during the early steps of amyloid formation.

Whether such  $\beta$ -barrel intermediates could play a role in disrupting or interacting with cell membranes and would be cytotoxic remains to be determined. In particular, it is unknown whether this  $\beta$ -barrel in solution with most hydrophobic residues protected from solvent can interconvert into a  $\beta$ -barrel in the membrane with this time most hydrophilic residues protected from the membrane environment.

**Acknowledgment.** This work is funded by NSFC (Grant No. 10674029), PCSIRT, and the Young Foundation of Fudan University. N.M. and P.D. are grateful to Fudan University for



a Senior Visiting Scholar Grant. N.M. acknowledges partial support from NSERC, the Canada Research Chair Foundation, and FQRNT. P.D. acknowledges support from CNRS and the Université de Paris 7 Denis Diderot. Simulations were performed at the Shanghai Supercomputing Center and the National High Performance Computing Center of Fudan University.

**Supporting Information Available:** Schematic diagram showing the variables used for defining a bilayer  $\beta$ -sheet during the simulations of seven chains. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Kaye, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **2003**, *300*, 486–9.
- (2) Kaye, R.; Glabe, C. G. Conformation-dependent anti-amyloid oligomer antibodies. *Methods Enzymol.* **2006**, *413*, 326–44.
- (3) Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. Amyloid  $\beta$ -protein ( $A\beta$ ) assembly:  $A\beta$  40 and  $A\beta$  42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330–335.
- (4) Fay, N.; Inoue, Y.; Bousset, L.; Taguchi, H.; Melki, R. Assembly of the yeast prion Ure2p into protein fibrils. Thermodynamic and kinetic characterization. *J. Biol. Chem.* **2003**, *278*, 30199–205.
- (5) Ban, T.; Goto, Y. Direct observation of amyloid growth monitored by total internal reflection fluorescence microscopy. *Methods Enzymol.* **2006**, *413*, 91–102.
- (6) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* **2005**, *435*, 773–778.
- (7) Tsemekhan, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* **2007**, *16*, 761–764.
- (8) Hills, R.; Brooks, C. Hydrophobic cooperativity as a mechanism for amyloid nucleation. *J. Mol. Biol.* **2007**, *368*, 894–901.
- (9) Pellarin, R.; Cafilisch, A. Interpreting the aggregation kinetics of amyloid peptides. *J. Mol. Biol.* **2006**, *360*, 882–892.
- (10) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer adds to preformed structured oligomers of  $A\beta$ -peptides by a two-stage dock-lock mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111–116.
- (11) Ma, B.; Nussinov, R. Molecular dynamics simulations of alanine rich  $\beta$ -sheet oligomers: Insight into amyloid formation. *Protein Sci.* **2002**, *11*, 2335–2350.
- (12) Nguyen, H. D.; Hall, C. K. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180–16185.
- (13) Takeda, T.; Klimov, D. K. Dissociation of  $A\beta$ 16–22 Amyloid Fibrils Probed by Molecular Dynamics. *J. Mol. Biol.* **2007**, *368*, 1202–1213.
- (14) Gnanakaran, S.; Nussinov, R.; Garcia, A. E. Atomic-level description of amyloid  $\beta$ -dimer formation. *J. Am. Chem. Soc.* **2006**, *128*, 2158–2158.
- (15) Lei, H.; Wu, C.; Wang, Z.; Duan, Y. Molecular dynamics simulations and free energy analyses on the dimer formation of an amyloidogenic heptapeptide from human  $\beta$ 2-microglobulin: implication for the protofibril structure. *J. Mol. Biol.* **2006**, *356*, 1049–1063.
- (16) Sgourakis, N.; Yan, Y.; McCallum, S.; Wang, C.; Garcia, A. The Alzheimer's peptides  $A\beta$ 40 and 42 adopt distinct conformations in water: a combined MD / NMR study. *J. Mol. Biol.* **2007**, *368*, 1448–1457.
- (17) Bernstein, S. L.; Wyttenbach, T.; Baumketner, A.; Shea, J. E.; Bitan, G.; Teplow, D. B.; Bowers, M. T. Amyloid  $\beta$ -protein: monomer structure and early aggregation states of  $A\beta$ 42 and its Pro19 alloform. *J. Am. Chem. Soc.* **2005**, *127*, 2075–2084.
- (18) Cellmer, T.; Bratko, D.; Prausnitz, J.; Blanch, H. Protein aggregation in silico. *Trends Biotechnol.* **2007**, *25*, 254–261.
- (19) Cecchini, M.; Rao, F.; Seebler, M.; Cafilisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **2004**, *121*, 10748–56.
- (20) Irbäck, A.; Mitternacht, S. Spontaneous  $\beta$ -barrel formation: An all-atom Monte Carlo study of  $A\beta$ (16–22) oligomerization. *Proteins*, in press.
- (21) Derreumaux, P.; Mousseau, N. Coarse-grained protein molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 025101.
- (22) Harrison, P.; Chan, H.; Prusiner, S.; Cohen, F. Conformational propagation with prion-like characteristics in a simple model of protein folding. *Protein Sci.* **2001**, *10*, 819–35.
- (23) Dima, R.; Thirumalai, D. Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics. *Protein Sci.* **2002**, *11*, 1036–49.
- (24) Wei, G.; Mousseau, N.; Derreumaux, P. Computational simulations of the early steps of protein aggregation. *Prion* **2007**, *1*, 3–8.
- (25) Wei, G.; Mousseau, N.; Derreumaux, P. Complex folding pathways in a simple  $\beta$ -hairpin. *Proteins* **2004**, *56*, 464–74.
- (26) Floquet, N.; Pasco, S.; Ramont, L.; Derreumaux, P.; Laronze, J. Y.; Nuzillard, J. M.; Maquart, F. X.; Alix, A. J.; Monboisse, J. C. 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*, 2091–100.
- (27) Mousseau, N.; Derreumaux, P. Exploring the early steps of amyloid peptide aggregation by computers. *Acc. Chem. Res.* **2005**, *38*, 885–891.
- (28) Ivanova, M. I.; Sawaya, M. R.; 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. U.S.A.* **2004**, *103*, 10584–10589.
- (29) Maupetit, J.; Tuffery, P.; Derreumaux, P. A coarse-grained protein force field for folding and structure prediction. *Proteins* **2007**, *69*, 394–408.
- (30) Derreumaux, P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. *J. Chem. Phys.* **1999**, *111*, 2301–2310.
- (31) Derreumaux, P. Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations. *Phys. Rev. Lett.* **2000**, *85*, 206–209.
- (32) Wei, G.; Mousseau, N.; Derreumaux, P. Exploring the energy landscape of proteins: A characterization of the activation-relaxation technique. *J. Chem. Phys.* **2002**, *117*, 11379–11387.
- (33) 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.
- (34) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566–79.
- (35) Santini, S.; Wei, G. H.; Mousseau, N.; Derreumaux, P. Pathway complexity of Alzheimer's  $\beta$ -amyloid A  $\beta$ (16–22) peptide assembly. *Structure* **2004**, *12*, 1245–1255.
- (36) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (37) Murzin, A. G.; Lesk, A. M.; Chothia, C. Principles determining the structure of  $\beta$ -sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **1994**, *236*, 1369–1381.
- (38) Liu, W.-M. Shear numbers of protein  $\beta$ -barrels: definition refinements and statistics. *J. Mol. Biol.* **1998**, *275*, 541–545.
- (39) Melquiond, A.; Mousseau, N.; Derreumaux, P. Structures of soluble amyloid oligomers from computer simulations. *Proteins* **2006**, *65*, 180–191.
- (40) Melquiond, A.; Boucher, G.; Mousseau, N.; Derreumaux, P. Following the aggregation of amyloid-forming peptides by computer simulations. *J. Chem. Phys.* **2005**, *122* (17), 174904.
- (41) Gsponer, J.; Haberthur, U.; Cafilisch, 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*, 5154–9.
- (42) Zanuy, D.; Porat, Y.; Gazit, E.; Nussinov, R. Peptide sequence and amyloid formation: molecular simulations and experimental study of a human islet amyloid polypeptide fragment and its analogs. *Structure* **2004**, *12*, 439–55.
- (43) Lopez de la Paz, M.; de Mori, G. M.; Serrano, L.; Colombo, G. Sequence dependence of amyloid fibril formation: insights from molecular dynamics simulations. *J. Mol. Biol.* **2005**, *349*, 583–96.
- (44) Nguyen, P.; Stock, G.; Mittag, E.; Hu, C.; Li, M. Free energy landscape and folding mechanism of a  $\beta$ -hairpin in explicit water: a replica exchange molecular dynamics study. *Proteins* **2005**, *61*, 795–808.
- (45) 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.
- (46) Hasegawa, K.; Ohhashi, Y.; Yamaguchi, I.; Takahashi, N.; Tsutsumi, S.; Goto, Y.; Gejyo, F.; Naiki, H. Amyloidogenic synthetic peptides of b2 microglobulin—a role of the disulfide bond. *Biochem. Biophys. Res. Commun.* **2003**, *304*, 101–106.
- (47) Jones, S.; Manning, J.; Kad, N. M.; Radford, S. E. Amyloid-forming Peptides from b2-Microglobulin—Insights into the Mechanism of Fibril Formation in Vitro. *J. Mol. Biol.* **2003**, *325*, 249–257.
- (48) Iwata, K.; Fujiwara, T.; Matsuki, Y.; Akutsu, H.; Takahashi, S.; Naiki, H.; Goto, Y. 3D structure of amyloid protofilaments of b2-microglobulin fragment probed by solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18119–24.

- (49) Thomas, R. J.; Martin, J. P.; Steve, W. H.; Sheena, E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **2006**, *13*, 195–201.
- (50) Petkova, A. T.; Buntkowsky, G.; Dyda, F.; Leapman, R. D.; Yau, W. M.; Tycko, R. Solid state NMR reveals a pH-dependent antiparallel  $\beta$ -sheet registry in fibrils formed by a  $\beta$ -amyloid peptide. *J. Mol. Biol.* **2004**, *335*, 247–60.
- (51) Zhu, M.; Han, S.; Zhou, F.; Carter, S. A.; Fink, A. L. Annular oligomeric amyloid intermediates observed by in situ atomic force microscopy. *J. Biol. Chem.* **2004**, *279*, 24452–9.
- (52) Lashuel, H. A.; Lansbury, P. T., Jr. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q. Rev. Biophys.* **2006**, *39*, 167–201.

- (53) Lashuel, H.; Hartley, D.; Petre, B.; Walz, T.; Lansbury, P. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature* **2002**, *418*, 291.
- (54) Wei, G. H.; Mousseau, N.; Derreumaux, P. Sampling the self-assembly pathways of KFFE hexamers. *Biophys. J.* **2004**, *87*, 3648–3656.
- (55) Marchut, A. J.; Hall, C. K. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys. J.* **2006**, *90*, 4574–4584.
- (56) Hong, H.; Patel, D. R.; Tamm, L. K.; van den Berg, B. The outer membrane protein OmpW forms an eight-stranded  $\beta$ -barrel with a hydrophobic channel. *J. Biol. Chem.* **2006**, *281*, 7568–77.