Sampling the self-assembly pathways of KFFE hexamers

Guanghong Wei†, Normand Mousseau†† and Philippe Derreumaux‡

†Département de Physique and RQMP, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7, and ‡Laboratoire de Biochimie Théorique, UPR 9080 CNRS, IBPC, Université Paris 7 Denis-Diderot, 13 rue Pierre et Marie Curie, 75005 Paris, France

Classification: Biological Sciences – Biophysics

Keywords: Activation-Relaxation Technique; peptide aggregation; Amyloid fibrils; simulations

May 14, 2004

†To whom correspondence should be addressed. Email: Normand.Mousseau@umontreal.ca
Abstract

The formation of amyloid fibrils is often encountered in Alzheimer’s disease, type II diabetes and transmissible spongiform encephalopathies. In the last few years, however, mounting evidence has suggested that the soluble oligomers of amyloid-forming peptides are also cytotoxic agents. Understanding the early pathway steps of amyloid self-assembly at atomic detail might therefore be crucial for the development of specific inhibitors to prevent amyloidosis in humans. Using the activation-relaxation technique and an approximate free energy model, we study in detail the aggregation of an hexamer of KFFE peptide. Our simulations show that a monomer remains disordered, but that six monomers placed randomly in an open box self-associate to adopt, with various orientations, three possible distant low-energy structures. Two of these structures show a double-layer $\beta$-sheet organization, in agreement with the structure of amyloid fibrils as observed by x-ray diffraction, while the third one consists of a barrel-like curved single-layer hexamer. Based on these results, we propose a bidirectional growth mode of amyloid fibril, involving alternate lateral and longitudinal growths.
Amyloid fibrils are often encountered in the late steps of Alzheimer’s disease, type II diabetes and transmissible spongiform encephalopathies (1, 2). This phenomenon is not limited to a few sequences, however, and recent data show that the ability to form these amyloid structures is a general property of any polypeptide chain under destabilising conditions (3); even peptides with as few as four to six residues can form well-defined fibrils showing all the characteristics of amyloid fibrils built from large proteins of 100 residues or more (4, 5, 6). Although a relation of causality between the amyloid fibrils and diseases still remains to be fully established, mounting evidence suggests that both insoluble fibrils and soluble oligomers are toxic in cell cultures (7). The structural characterization of these aggregation intermediates raises considerable challenges, however, as oligomers typically occur with a very low concentration and tend to be rather short-lived, making it very difficult to obtain precise structural information experimentally. For example, solid-state NMR has only been used to study the structural properties of stable fibrils (8) and in spite of much efforts, the structure of oligomers remains mostly unknown (9).

In view of these limitations, computer simulations have provided a very convenient complement to experimental approaches and the aggregation of peptides/proteins has been studied by using a range of protein models. While investigations based on lattice models have provided numerous insights into the thermodynamic and kinetic properties of protein aggregation (10, 11, 12), simplified off-lattice studies with a single-bead residue model and a Gō potential (13, 14) have helped clarify the kinetics and assembly processes of the β-sheet complex. Due to the simplicity of these models, however, it is not possible to extract detailed information on the structure and growth of oligomeric intermediates and more realistic representations must be used. Proceeding along these lines, various groups have studied the aggregation of dimers and trimers of amyloid-forming peptides using different atomic peptide models (15, 16, 17, 18). These simplified trajectories with only two or three strands help understand the first steps of aggregation, but they fall well below the size of the critical nucleus from which growth is rapid (19). Taking a different approach, Ma and Nussinov have studied the stabilities of various oligomers (20, 21) using molecular dynamics (MD) in explicit solvent, identifying stable structures of hexamers and octamers.

Complementing these previous studies, we examine here the self-assembly pathways of oligomers using an off-lattice peptide model (22, 23), which goes beyond the Ca representation. This peptide model, where each amino acid is represented by its N, H, Co, C, O and one bead for its side chain, allows us to study the explicit formation of hydrogen bonds (H-bonds), but also to simulate aggregation in systems with more than three peptides, the current system size. More precisely, we study the aggregation process of an hexamer of the KFFE peptide. This tetrapeptide was shown to form amyloid fibrils (4) and yet is small enough for simulations at an atomic level. The choice of an hexamer is based on results of a recent mathematical model of the nucleated polymerization mechanism which identifies this as the critical size for fibril formation (24).

Using the activation-relaxation technique (ART) (25, 26) combined with an optimized-potential for efficient peptide structure prediction (OPEP) (22, 23), we identify various self-assembly pathways leading to various hexameric metastable structures, including a two β-sheet structure displaying the structural characteristics of amyloid fibril. Our results demonstrate that small oligomers are in equilibrium between many different aggregation structures in addition to the fibril-like conformation, in agreement with recent structural-recognition measurements using antibody agents (27), and in-situ atomic force microscopy (28).
Methods

We have simulated the self-assembly of the KFFE hexamer using ART-OPEP. The N and C termini were neutralized using acetyl and amine groups, respectively. Twenty simulations were performed with open boundary conditions; each run generates 9000 events, starting from four different initial states and using different random number seeds. These four initial states include six monomers placed with random orientations in a sphere of diameter 40 Å (IN1, 10 runs), an antiparallel dimer and four monomers with random orientations (IN2, 4 runs), a double-layer three-stranded β-sheet structure with almost perpendicular orientations of the inter β-sheets (IN3, 4 runs), and a tetramer-dimer structure (IN4, 2 runs). Although the Metropolis temperature with ART cannot be directly associated with the real temperature, we ran at 310 K, the experimental incubation temperature of KFFE peptide fibrils (4). In this study, all the structures were produced using the the MolMol software (29) except for the structure in Fig. 5(d) produced using Rasmol (30), and the N-terminal end of each chain is located by a sphere.

Simulation method - ART

ART is a generic method for exploring the energy landscape of complex systems. An ART event consists of four basic steps: starting from a local minimum, a randomly chosen chain of the system is pushed and moves along a random direction until the lowest eigenvalue of the Hessian matrix becomes negative, indicating the vicinity of a saddle point; the system is then pushed along the eigenvector associated with the lowest eigenvalue until the total force approaches zero, and the system converges onto a first-order saddle point; the system is pushed slightly over the saddle point and is relaxed to a new local minimum, using standard minimization technique; the new local minimum is accepted or rejected using Metropolis criterion. As discussed in our previous works (31, 32), the trajectories sampled by ART are well controlled and the path from the initial to the final state by connecting all the accepted minima in sequence is a physical path; the folding trajectories generated this way have been found to be very similar to those obtained by MD and ensemble dynamics for a 14-residue α-helix (31) and the second β-hairpin of protein G (32). A detailed description of the algorithm and its implementation can be found in Refs. (25, 26, 31, 32).

Energy Model - OPEP

Our protein model is a coarse-grained off-lattice model where the main chain of each amino acid is represented explicitly and its side chain is represented by one bead (22, 23). The OPEP energy function, trained on the structures of monomeric peptides with various secondary structures, is expressed as a function of four types of interactions: harmonic potentials for maintaining the geometry of peptides (bond lengths and bond-angles for all particles, and improper dihedral angles of side chains with respect to the backbone); backbone two-body and four-body (cooperative) hydrogen bonding interactions (22, 33); pairwise potential between side chains represented by 12-6 potential if the interactions are hydrophobic and by a 6-potential, otherwise; and excluded-volume potentials between all particles. The impact of the force field on the ART trajectories and the energy barriers is discussed in Ref. (32).
Results

Structure characterization of monomer

We first characterize the structure of the monomeric peptide. We run four simulations for a total of 8000 accepted events. The secondary structure, computed using the $\phi$ and $\psi$ angles of the two inner residues following the criteria of Ref. (34), show that about 72% of all monomeric structures adopt a random coil conformation and 28% a $\beta$-strand structure; the population of $\alpha$-helical conformations is negligible (0.12%). These results are consistent with the CD analysis in solution at pH 6.0, with $\beta$-strands and random coil structures coexisting (no data shown) (4). The KFFE peptide in solution is therefore well described by OPEP.

Self-assembly pathways for the hexamers

All simulations lead to organized conformations with a total energy between $-55$ and $-60$ kcal/mol. Although each simulation generates a unique trajectory, it is possible to categorize them as belonging to one of three generic interconnected self-assembly pathways as discussed separately below. The first identified pathway leads to the formation of a double-layer three-stranded $\beta$-sheet with various in-plane orientations; this structure displays the characteristics of amyloid fibrils. The second pathway leads to a double-layer tetramer-dimer structure and the third one generates a monolayer $\beta$-sheet in a barrel-like shape. These three pathways are not independent: for example, we find that the double-layer three-stranded $\beta$-sheet can transform into a tetramer-dimer, and that both structures are possible intermediates towards the monolayer $\beta$-sheet.

Double layer three-stranded parallel-antiparallel-$\beta$-sheets

Of the ten runs starting from the random state IN1, two aggregate into a double-layer three-stranded parallel-antiparallel $\beta$-sheet. Although the detailed molecular structure of amyloid fibrils has not yet been determined experimentally — their non-crystalline and insoluble nature makes their characterization by x-ray diffraction or liquid-state NMR very difficult — X-ray fiber diffraction shows that amyloid fibrils contain a cross-$\beta$ structural motif. This motif is characterized by extended $\beta$-sheets in which the $\beta$-strands run perpendicular to, and the intermolecular H-bonds run parallel to, the fibril axis (35). These structural characteristics are present in both of our generated double-layer three-stranded parallel-antiparallel-$\beta$-sheet hexamers. Fig. 1 (a) and (b) show two different views of one hexamer (T1) which displays a perfect parallel match between layers even though each layer has a mixture of parallel and antiparallel $\beta$-strand. The other hexamer (T2) is less stable, with one antiparallel pair across the plane (figure not shown). This organization is less stable due to the misalignment of this pair. We describe below the assembly process for the parallel hexamer T1; results are qualitatively similar for T2. As shown in Fig. 1 (a) and (b), the distance between antiparallel $\beta$-strands is 4.5 Å, and the distance between two nearest nitrogen atoms located in the two parallel $\beta$-strands belonging to the two different layers is about 10-12 Å. These two distances are in agreement with the X-ray diffraction measured distances of 4.7 Å between $\beta$-strands and 10-11 Å between two $\beta$-sheets in fibrils (35, 36).

As can be seen from Fig. 2, there is considerable energy relaxation in the first 1000 events (see (a)), with most peptides (1, 3, 4, 5, and 6) becoming extended, with an end-to-end distance close to 12 Å the fully extended value (see (b)); however, no peptide shows a stable $\beta$-strand conformation (see (c)). Although
the end-to-end distance continues to fluctuate over the next 500 events, at event 1500, peptides 1, 4, 5, and 6 have moved in a β-strand conformation, with the remaining two adopting a random coil state. Panel (c) shows that the peptides only oscillate between these two states, with very little α-helical component. During these first 1500 events, there is considerably reorientation of the various peptides, as the first trimer, composed of 3, 5 and 6, stabilized into a parallel-antiparallel conformation. For the following 1500 events, this trimer adopts a more stable, fully extended state, while the second trimer (1, 2, 4), stabilizes, also in a parallel-antiparallel conformation, before locking into place, parallel to the first trimer. As is shown in panel (d), after these two trimers have settled into a two β-sheet structure, they rotate about each other by a full 360 degrees, adopting in passing a fully perpendicular orientation around event 2600 and antiparallel orientation around event 2900. During the whole self-assembly process, the configurational energy decreases steadily and converges to a plateau around -55 kcal/mol, avoiding being trapped in high-energy structures.

We summarize the assembly pathway for the fibril-like hexamer in Fig. 3: starting from a random conformation (a), an antiparallel dimer forms (b) and attracts a third monomer in a parallel orientation (c). At the same time, the other monomers come together and form an amorphous aggregate state (c). After inter-peptide interactions, a second trimer almost forms (d). The first formed antiparallel-parallel trimer helps stabilize the formation of the second trimer (e) which, after some rotation, forms a β-sheet perpendicular to the first trimer (e). This structure is not very stable and the two trimers rotate about each other and rearrange themselves into a parallel β-sheet structure (f).

Tetramer-dimer and monolayer β-sheets

Four out of the ten runs starting from IN1 lead to a tetramer-plus-dimer structure; two out of four runs starting from IN2 and one out of four starting from IN3 also aggregate into a similar structure. Fig. 4 shows the details of one representative trajectory from IN1. The conformational energy (Fig. 4(a)) drops rapidly and the number of H-bonds increases to 16 as the tetramer forms. After 900 accepted events, four out of the six monomers become extended (the end-to-end distance is about 12 Å) (see (b)), and the tetramer is already almost in its most stable state, in a well-aligned antiparallel β-sheet (see (d)). The two remaining peptides continue to move with respect to each other until event 2000, when they lock into an antiparallel dimer rotating into a parallel orientation with respect to the tetramer (Fig. 4(c)). The distance between two nearest nitrogen atoms between the two layers, is again 10 - 12 Å. The width of the β-sheets (defined as the length of β-strand) is around 12 Å which is consistent with the width 12 - 16 Å of 10 measured fibrils experimentally.

We can summarize qualitatively the assembly process of this structure by a sequence of structural steps. As in the case of the fibril-like hexamer, an antiparallel structure forms first. Two more peptides aggregate to this dimer, in sequence, with an antiparallel orientation, leading to a fully antiparallel tetramer. The two remaining peptides then assemble into a separate antiparallel dimer. During the whole process, the two peptides at the edge of the tetramer show considerably more flexibility than those in the center and will often break H-bonds.

The antiparallel four-stranded β-sheet (Fig. 4(d)) generated by ART during the folding of the KFFE hexamers is consistent with the proposed model of Tjernberg et al. Moreover, its assembly pathway is also in agreement with that proposed by Jang et al for a tetrametric β-sheet complex using MD simulations and a Go-energy model: four monomers → dimer and two monomers → trimer and monomers → tetramer.

Starting from IN4, a tetramer-dimer structure (Fig. 5(b)), both simulations, run over 9000 events, lead
to a third low-energy structure: a $\beta$-barrel structure (Fig. 5(c) and (d)). It is a curved monolayer hexamer, mostly antiparallel, of energy similar to the tetramer-dimer structure ($-60$ kcal/mol) and stabilized by H-bond interactions. Looking at the all-atom ball and stick representation of this structure (Fig. 5 (d)), we found six out of 12 Phe side-chains (orange ball) inside the barrel.

**Discussion**

**Inter-peptide organization**

Although we identify three low-energy metastable structures in our simulations, there is a considerable fluctuation in the details of the inter-peptide organization, both in terms of orientation and H-bond network. The existence of out-of-register alignments is to be expected for such short peptides. This observation has been reported during the aggregation of the trimeric heptapeptide GNNQQNN from yeast prion Sup35 by MD with implicit solvent (37), and the assembly of the dimeric heptapeptide KLVFFAE from A$\beta$ Alzheimer’s peptide by ART (18).

These fluctuations are also found in other amyloid-forming systems. Experiments and modeling of the human islet polypeptide suggest, for example, that the $\beta$-sheet of this peptide could be arranged in a parallel conformation or as a mixture of both parallel and antiparallel (38, 39). Similarly, an off-register staggered arrangement in A$\beta$34–42-fibrils was proposed recently (8). It was also shown by NMR solid-state analyses that the molecular conformations of fibrils depend strongly on the pH conditions. In particular, the human calcitonin fibril structure may be composed mainly of antiparallel $\beta$-sheets at pH 7.5 and 4.1, and a mixture of antiparallel and parallel $\beta$-sheets at pH 3.3 (40). Petkova et al. showed that A$\beta$11–25 fibrils adopt distinct antiparallel $\beta$-sheet registries as a function of pH (41). While the most favored arrangement might depend on the specificity of the environment, these results imply that all alignments are thermodynamically reachable.

**Oligomer growth process**

The early steps of the oligomeric assembly consist in the association of monomers, resulting in the formation of different arrangement of hexamers. In our simulations, starting from randomly chosen conformations, the peptides undergo a dramatic increase in $\beta$-strand content, consistent with the hypothesis proposed by Serpell and Smith (42), leading to two distinct double-layers with trimer-trimer or tetramer-dimer $\beta$-sheets. Some simulations show that the trimer-trimer structure can transform into a tetramer-dimer and, from there, into a curved monolayer six-stranded $\beta$-sheet. The connection between the three different lowest-energy hexamers is shown in Fig. 5. Interestingly, all these conformations are structurally relevant intermediates in the aggregation of fibrils, in the sense that they represent building blocks that are fully compatible with the structural properties of fibrils.

The three different oligomeric structures generated in our simulations are consistent with a large body of data. The double-layer three-stranded $\beta$-sheets and the tetramer-dimer intermediate have the structural characteristics of fibril. In a recent work of Zanuy et al (21), the stabilities of all-atom double-layer and triple-layer $\beta$-sheets were studied by MD simulations in explicit solvent. The mono-layer hexamer was not considered in this theoretical study, but it has been observed by light scattering and atomic force microscopy for a related peptide (43). Fung et al (43) have investigated the concentration effects on the aggregation of the EAK16-II peptide. Three types of nanostructures were observed: the filaments, the globular aggregates...
and the fibrils. They also found that the filaments are monolayers of β-sheets, but the globular aggregates are made of at least 2 layers of β-sheets.

Although all our simulations require first the formation of an antiparallel dimer intermediate, they suggest that there does not exist a single route to amyloid fibrils: these can form by the addition of single monomer or small oligomers to preformed oligomers of various size, in contradiction with the unique addition-reaction mechanism proposed in Refs. (37, 44). Even for a hexamer, dimers and trimers are seen to attach to existing structures in a single step.

Based on our simulations, we can identify an assembly mechanism. Two different models for the assembly of human amylin (hA) fibril have been proposed based on morphological studies using time-lapse atomic force microscopy (45, 46). The first model is that the hA fibril forms by lateral growth of oligomers followed by longitudinal growth into mature fibrils (45). The second scenario postulates that profibrils elongate (longitudinal growth) and subsequently associate laterally to form thicker fibrils (46). Also using atomic force microscopy, Khurana et al. (47) studied the fibril assembly of three proteins: α-synuclein, insulin, and the B1 domain of protein G. They observed that the fibrils formed by the lateral association of preformed protofilaments. From our simulations of the self-assembly of KFFE hexamers, we find that the growth of the fibril-like hexamer is by a bidirectional growth mode, with an alternate longitudinal growth and lateral growth. We propose that the growth of fibrils made of short peptides follow a bidirectional growth mode as shown by a cartoon in Fig. 6, which complements the two previously proposed fibril growth modes.

Conclusion

Understanding the process of aggregation is the first step for the rational design of therapeutics preventing neurodegenerative diseases. In this study, we have simulated the assembly of six monomers of the KFFE peptide into ordered oligomers using ART-OPEP. To our knowledge, this is the first aggregation study that takes into account more than three monomers using a more realistic chain representation than one- or two-beads and without using a Gō-energy model. Because of the simplicity of the side chains and solvent effects, the assembly process generated here might not reproduce the exact aggregation pathway as by experiments. Nevertheless, we can extract some qualitative features about its self-assembly pathway and the structures of oligomers. Our main results are: three different low-energy hexameric structures are in equilibrium, and both parallel and antiparallel β-strand arrangements are obtained. Two of these oligomers are double-layer β-sheets with structural characteristics of amyloid fibrils observed by x-ray diffraction, and one is a monolayer β-sheet. These results indicate that oligomers can adopt many distinct aggregation intermediates, in agreement with recent experimental observations using different techniques, whose populations will be determined by the solvent condition and the crowded environment. The assembly of ordered hexamers follows three different processes: 1) six monomers → dimer and four monomers → trimer and three monomers → two trimers → bilayer hexamer; 2) six monomers → dimer and four monomers → trimer and three monomers → tetramer and two monomers → bilayer hexamer; 3) bilayer trimer trimer or bilayer tetramer dimer → single layer hexamer. Finally, from our simulations, we propose a bidirectional growth mode of amyloid fibril, alternate lateral growth and longitudinal growth.

GW and NM are supported in part by the Fonds québécois de la recherche sur la nature et les technologies and the Natural Sciences and Engineering Research Council of Canada. NM also acknowledges funding from the Canada Research Chair Program. Most of the calculations were done on the computers of the Réseau
québécois de calcul de haute performance (RQCHP). NM is a Cottrell Scholar of the Research Corporation.

References

Figure captions:

Fig. 1: Two different views of the two double-layer three-stranded $\beta$-sheet hexamers. The two layers are perfectly parallel: (a) view perpendicular to the fibril axis; (b) view parallel to the fibril axis. Blue line represents H-bonds formed between the carbonyl oxygen and the amide hydrogen. The N-terminal end of each chain is located by a sphere.

Fig. 2: Characterization of the self-assembly pathway as a function of accepted event number leading to the formation of fibril-like hexamer. Evolution of (a) the conformational energy; (b) end-to-end Ca distance for the six monomers; (c) secondary structure of each residue in the six monomers — the secondary structure is assigned according to the values of dihedral angles $\phi$ and $\psi$ in Ref. (34), blue for random coil, red for $\alpha$-helix, and green for $\beta$-strand, $F_{ji}$ is the $j$th ($j = 2, 3$, the two inner residues in each peptide) residue Phe in peptide $i$ ($i = 1-6$); and (d) the inter-sheet orientation for the three final pairs. Orientation is calculated using the scalar product between the end-to-end unit vectors of each chain.

Fig. 3: Assembly pathway ((a) $\rightarrow$ (f)) leading to a fibril-like hexamer with a parallel-antiparallel $\beta$-sheet/parallel layer structure: (a) six monomers with random orientations; (b) a dimer with four monomers; (c) A trimer with three monomers; (d) A trimer with another trimer which is not well-aligned; (e) two trimers with perpendicular inter-strand arrangement; (f) two trimers with parallel inter-strand arrangement (view parallel to fibril axis and perpendicular to $\beta$-strands). The colors of the six peptides, from monomer 1 to 6, respectively are yellow, green, red, blue, cyan, and magenta. Blue line represents H-bonds formed between the carbonyl oxygen and amide hydrogen.

Fig. 4: Characterization of the self-assembly pathway as a function of accepted event number leading to the formation of a tetramer plus a dimer. Evolution of (a) the conformational energy (solid line) and number of H-bonds (dashed line); (b) end-to-end distance for the six monomers; (c) intra-sheet orientation for selected pairs in the top (red: 5-6) and the bottom layer (blue: 2-4, magenta: 4-1, black: 1-3), inter-sheet orientation between a monomer in the tetramer and one in the dimer (cyan: 5-2). (d) A close-up view of the tetramer, in a well-aligned four-stranded antiparallel configuration. From left to right, the four monomers are 2, 4, 1, 3, respectively. Blue line represents H-bonds formed between the carbonyl oxygen and amide hydrogen.

Fig. 5: The three lowest-energy structures generated in our simulations. (a) the double-layer three-stranded $\beta$-sheet, (b) the double-layer tetramer-dimer structure, and (c) the $\beta$-barrel hexamer with the main-chain atoms shown. (d) the $\beta$-barrel hexamer in a ball and stick representation and the side-chains of Phe residue in orange. The arrows indicate pathways generated in our simulations.

Fig. 6: Schematic diagram showing a bidirectional grow mode of fibril. The growth can take place by adding monomer, dimers or larger oligomers. It also alternates between lateral and longitudinal growth. A newly-generated $\beta$-strand is in grey, an old one is in black.
Figure 1  Authors: Wei G. et al.
Figure 2  Authors: Wei G. et al.
Figure 3
Authors: Wei G. et al.
Figure 4  Authors: Wei G. et al.
Figure 5  
Authors: Wei G. et al.
Figure 6  Authors: Wei G. et al.