FAHIMEH BAFTIZADEH[1], XEVI BIARNES[1,3], FABIO PIETRUCCI[2], FABIO AFFINITO[4], ALESSANDRO LAIO[1]

[1] *SISSA, Trieste*
[2] *Centre Europeen de Calcul Atomique et Moleculaire, Lausanne, Switzerland*
[3] *Institut Quimic di Sarria Universitat Ramon Llull, Barcelona, Spain*
[4] *CINECA, Bologna*

A Multidimensional View of Amyloid Fibril Nucleation in Atomistic Detail

Simulation of Polypeptide Aggregation Using Bias-Exchange Metadynamics

1. Introduction

Amyloid fibrils are large protein aggregates characterized by a high content of beta-sheets. Their presence is a hallmark of several neurodegenerative disorders, including Alzheimer, Huntington, Parkinson and prion disease (1). The precise pathological role of fibrillar species is not completely understood. However, amyloid protofibrils as well as oligomers have been suggested to lead to neuronal cell death (2). Moreover, interruption of fibril formation prevented cell damage, suggesting that the early oligomers in the fibrillation process are probably toxic (3). Unlike other protein quaternary structures (4), amyloid fibrils involve a conformation, known as 'cross-beta structure', that is basically sequence independent (5-7). This structural motif consists of individual beta-sheets in a distinctive orientation perpendicular to the major axis of the fibril, with the side chains protruding from the sheets on each side in the characteristic *steric zipper* conformation (8,9).

Amyloid structures are large, insoluble in water and difficult to crystallize. Specific experimental strategies have been used to improve our understanding of the molecular mechanisms leading to their formation (1,10). However, the exact molecular assembly of a nascent fibril remains a matter of debate, and computer simulations have played an important role in addressing this question (11,12,13). Intermediate-resolution fibril models with coarse grained force-fields or implicit solvent have been successfully investigated (14) However, to the best of our knowledge, a detailed simulation of the fibril formation process with an all-atom force field and reasonable size of the system in explicit solvent is still missing even for short peptides.

We here exploited a powerful enhanced sampling technique, Bias-exchange metadynamics (15), to investigate the conformational free energy landscape of a set of 18 chains of Poly-Valine, each 8 residues long. The system is descried with an accurate all atom explicit solvent force field (16). We perform the simulation on a homo peptide, and not on a peptide of primary sequence that favors fibril formation, with the scope of optimally capturing sequence-independent features in the landscape. Using this approach we are able to compute the free energy of hundreds of putative aggregate structures, with various content of parallel and anti-parallel beta-sheets, and several different packing among the different sheets, characterizing in detail a possible nucleation pathway. Our results provide a strong indication that the first steps of the process leading to the formation of ordered aggregates in short peptides follow a highly non-trivial pathway that cannot be described by classical nucleation or a simple mesoscopic theory.

2. Results

2.1 Bias-exchange metadynamics (BE-META)

We performed a Bias-exchange metadynamics simulation of a system of 18 chains. This relatively large system size has been chosen in order to have a high chance of observing a critical nucleus: We used 8 replicas, all at the same temperature (350 *K*), each biased along a different CV (see Methods). The CVs describe the geometrical feature of amyloid structures, namely, lateral packing of polypeptide chains in a beta-sheet configuration and subsequently frontal packing of beta-sheets and in multiple layers. The simulation was run for a total time of 4240*ns*, 540*ns*, for each of the eight replicas, starting from a completely disordered aggregate. Due to the action of the bias, the system explores hundreds of structured configurations. 20 of the structures are represented in Figure 1 with their corresponding free energies (see below and Methods).

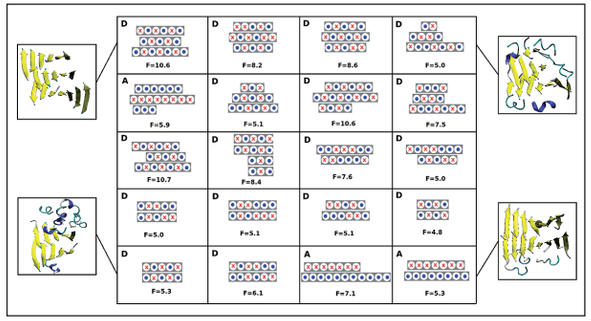

Fig. 1. – 20 representative structures obtained from BE-META simulation performed on the system of 18 chains of 8VAL in explicit solvent at 350K. The gray layers are beta-sheets, the circle with a central dot represents a beta strand pointing outside the paper plane, and the circle with a cross represents a beta strand pointing inside the paper plane. we also showed the cartoon representation for 4 of the structures. The free energy of each structure is also reported. Label *A* means that the structure is committed to basin 3 in Figure 3; label *D* means that the structure is committed to the disordered aggregate.

After 300*ns*, the history-dependent potentials acting on the replicas reaches a stationary state. After this time, trajectories were analyzed following the procedure (17). This allowed obtaining a converged estimate of the free energy of 500 structures. A representative configuration for some of these structures is provided in Figure 1, together with the estimated free energy.

2.2 A classical nucleation picture

In the attempt of explaining the formation of an ordered aggregate according to classical nucleation theory, we first used the results obtained in BE-META to compute the free energy profile as a function of a single order parameter, namely, the number of beta strands. To do this, we first computed the average number of beta strands in each cluster

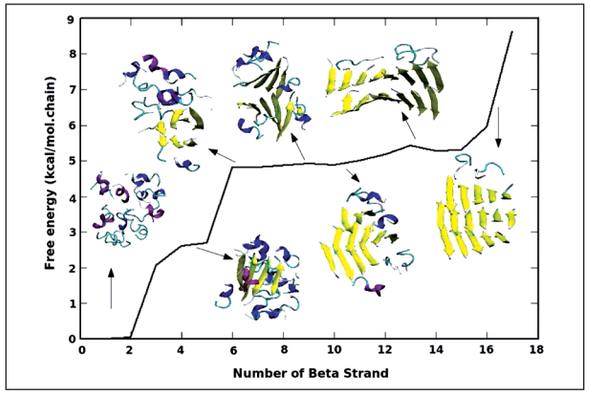

Fig. 2. – Free energy as a function of the number of beta strands.

(n_α). The free energy as a function of the number of strands n is then given by:

$$F(n) = -K_B T \log\left(\sum_\alpha \chi_n(n_\alpha) \exp\left(-\frac{F_\alpha}{K_B T}\right)\right) \quad (1)$$

where F_α is the free energy of cluster α and $\chi_n(n_\alpha)$ equal to 1 if $n_\alpha \in \left[n-\frac{1}{2}, n+\frac{1}{2}\right]$ and 0 otherwise. The result is shown in Figure 2. No barrier in the free energy as a function of the number of beta strands is present. The disordered aggregate is the minimum in the free energy. Moving towards structures with a higher content of beta-sheet the free energy rises to around 5 *kcal/mol/chain* for $n = 6$. Afterwards, in spite of the fact that the amount of order increases, the free energy grows slowly, and no barrier is present. From this graph one would conclude that the critical nucleus is even larger than $n = 18$, or that ordered structures are intrinsically metastable and committed to the disordered aggregate.

2.3 A 3-dimensional picture of the nucleation process

Since the free energy profile as a function of a single coordinate measuring the number of beta-sheets is barrierless, we started considering the free energy landscape in more dimensions. A completely different picture arises if one considers the free energy landscape as a function of three variables, quantifying the anti-parallel and parallel packing of beta strands inside a layer and the steric zipper packing of the beta strands in front of each other. A volumetric representation of free energy

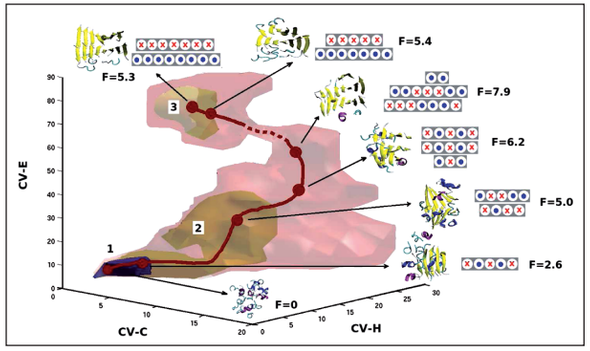

Fig. 3. – Free energy as a function of 3 collective variables: CV-C is the number of anti-parallel beta-sheets, CV-E is the number of parallel beta-sheets and CV-H is the number of anti-parallel steric zippers. The red isosurface represent the region of CV-space explored during the simulation. The blue isosurface corresponds to a free energy of 3.3 kcal/mol/chain, and contains the global minimum, a disordered aggregate. The orange isosurface, at 5.8 kcal/mol/chain, highlights a plateau region (region 2) connected to region 1 and another basin (region 3), which is separated from regions 1 and 2 by a barrier. The red line follows the lowest free-energy path connecting the disordered aggregate (region 1) with region 3. We also show 7 representative structures along the path, with their corresponding free energies.

with respect to these 3 CVs is presented in Figure 3. The global minimum of the system still corresponds to the disordered aggregate. The plateau observed in the one-dimensional profile is also present, at a free energy of approximately 5.8 *kcal/mol/chain* (light orange region labeled with 2 in Figure 3). It mostly consists of structures with a high content of anti-parallel beta-sheets, that seem to form more easily from the disordered aggregate. At variance with what observed in the one dimensional free energy landscape, in three dimensions a well defined free energy minimum is present (labeled 3 in Figure 3). The minimum free energy in this region is 5.8 *kcal/mol/chain*, like in the plateau, but this region is separated from the rest by a barrier of at least 2 *kcal/mol/chain*. Only a lower bound for this value can be provided, as the free energy of the structures close to the transition state is not converged within the 4000*ns* BE-META simulation. Region 3 is characterized by highly ordered structures which are very rich in parallel beta-sheets.

3. Discussion.

The aggregation process of 18 chains of poly-valine has been investigated by molecular dynamics using an all-atom force field in which the solvent is modeled explicitly. To enhance the sampling of the configuration space, the BE-META technique has been employed. The dynamics of the system has been driven by a set of eight reaction coordinates. These have been chosen in an attempt of capturing the most important degrees of freedom associated with aggregation, especially the parallel and anti-parallel arrangements of strands in a beta-sheet layer (lateral packing), and the arrangement of beta strands in layers facing each other (frontal packing). Each of the eight reaction coordinates has been biased on a different replica of the system, until convergence in the reconstructed free energy projections has been achieved. Using this methodology, we were able to simulate for the first time the formation of ordered beta structures from a disordered aggregate. We obtained several independent structures, and for each of these structures we computed the free energy. These data can be used for optimizing a coarse grained potential, or as a template for constructing possible structures of aggregates of peptides with a different primary sequence.

The free energy as a function of three coordinates, shows that while structures rich in anti-parallel beta-sheets belong to the same free energy minimum of the disordered aggregate, structures with an even smaller fraction of beta-sheets, but rich in parallel strands form a well defined minimum, separated from the disordered aggregate by a relatively high barrier. The comparison of Figures 2 and 3 shows that for this system projecting the free energy on a single "natural" reaction coordinate does not capture the qualitative features of the process (18). The minimum free energy path is shown in Figure 3 as a red line. Clearly, at least in the space of the three variables chosen for the representation, the path is highly non trivial: first the system forms anti-parallel beta-sheets, that are favored over parallel ones when the structure is highly disordered. Then, within a relatively large ordered nucleus formed mostly by anti-parallel beta-sheets, a few parallel beta-sheets start appearing. The relevant nucleation process seems to happen at this point: when a sufficient number of parallel sheets is formed, the free energy finally starts to decrease towards a new minimum, in which parallel sheets are predominant. By monitoring the structures close to the barrier we found that they contain a similar number of strands as the structures of the ordered aggregate. The only qualitative difference is in the number of parallel beta-sheets, that is much larger in the structures close to the free energy minimum. The structures close to the barrier have only a few parallel beta-sheets, surrounded by anti-parallel beta-sheets.

4. Methods

We performed Bias-exchange metadynamics(BE-META) (15) using PLUMED (19). The collective variables (CVs) specific to this system were coded by us in PLUMED. After 10*ns* of equilibration, the BEMETA simulation was started. BEMETA is a combination of replica exchange (20) and metadynamics (21) that allows reconstructing the free energy as a function of a large number of CVs (15).

Acknowledgement: We acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support

References

(1) F. Chiti and C. Dobson, *The Annual Review of Biochemistry,* **75**, 333 (2006).
(2) D. Hartley et al., *The Journal of Neuroscience,* **19**, 8876 (1999).
(3) M. Bucciantini et al., *Nature,* **416**, 507 (2002).
(4) C. Branden and J. Tooze, *Introduction to protein struture,* (Garland Publishing, 2nd edition, ADDRESS, 1998).
(5) M. Sunde et al., *Journal of Molecular Biology,* **273**, 729 (1997).
(6) J. Sipe and A. Cohen, *Journal of Structural Biology,* **130**, 88 (2000).
(7) M. Fandrich and C. Dobson, *The EMBO Journal,* **21**, 5682 (2002).
(8) R. Nelson et al., *Nature,* **435**, 773 (2005).
(9) M. Sawaya et al., *Nature,* **447**, 453 (2007).
(10) J. Rochet and P. Lansbury, *Current Opinion in Structural Biology,* **10**, 60 (2000).
(11) P. Nguyen et al., *Proceedings of the National Academy of Sciences of the United States of America,* **104**, 111 (2007).
(12) R. Pellarin and A. Caflisch, *Journal of Molecular Biology* **360**, 882 (2006).
(13) B. Ma and R. Nussinov, *Current Opinion in Chemical Biology,* **10**, 445 (2006).
(14) S. Auer, C. Dobson, and M. Vendruscolo, *HFSP Journal,* **1**, 137 (2007).
(15) S. Piana and A. Laio, *J. Phys. Chem.* B, **111**, 4553 (2007).
(16) W.D. Cornell et al., *Journal of the American Chemical Society* **117**, 5179 (1995).
(17) F. Marinelli, F. Pietrucci, A. Laio, and S. Piana, *PLoS Comput. Biol.,* **5**, e100045 (2009).
(18) P. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, *Annual Review of Physical Chemistry,* **53**, 291 (2002).
(19) M. Bonomi et al., *Comp. Phys. Comm.,* **180**, 1961 (2009).
(20) B.A. Berg and T. Neuhaus, *Phys. Rev. Lett.,* **68**, 9 (1992).
(21) A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA,* **99**, 12562 (2002).