Multidimensional View of Amyloid Fibril Nucleation in Atomistic Detail

Fahimeh Baftizadeh,[†] Xevi Biarnes,[‡] Fabio Pietrucci,[¶] Fabio Affinito,[§] and Alessandro Laio*[,†]

[†]SISSA, Via Bonomea 265, Trieste, Italy
[‡]Institut Quimic di Sarria Universitat Ramon Llull, Barcelona, Spain
[¶]Centre Européen de Calcul Atomique et Moléculaire, EPFL, Lausanne, Switzerland
[§]CINECA, Bologna, Italy

Ⓢ *Supporting Information*

ABSTRACT: Starting from a disordered aggregate, we have simulated the formation of ordered amyloid-like beta structures in a system formed by 18 polyvaline chains in explicit solvent, employing molecular dynamics accelerated by bias-exchange metadynamics. We exploited 8 different collective variables to compute the free energy of hundreds of putative aggregate structures, with variable content of parallel and antiparallel β-sheets and different packing among the sheets. This allowed characterizing in detail a possible nucleation pathway for the formation of amyloid fibrils: first the system forms a relatively large ordered nucleus of antiparallel β-sheets, and then a few parallel sheets start appearing. The relevant nucleation process culminates at this point: when a sufficient number of parallel sheets is formed, the free energy starts to decrease toward a new minimum in which this structure is predominant. The complex nucleation pathway we found cannot be described within classical nucleation theory, namely employing a unique simple reaction coordinate like the total content of β-sheets.
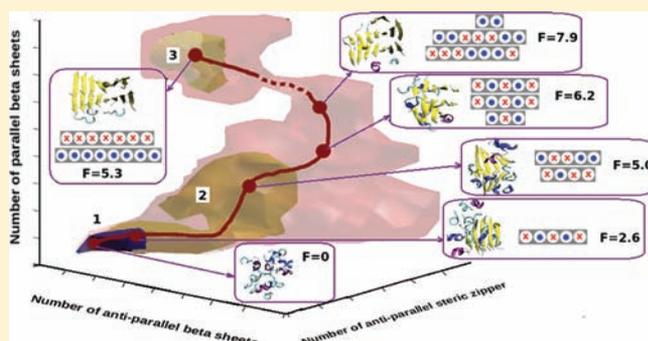
■ INTRODUCTION

Amyloid fibrils are large protein aggregates characterized by a high content of β-sheets. Their presence is a hallmark of several neurodegenerative disorders, including Alzheimer's, Huntington's, Parkinson's, and prion diseases.[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 β-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] For instance, crystals of small amyloid-like fragments, assembled from peptides 6−12 amino acids in length, have been successfully produced, allowing the construction of high-resolution three-dimensional models.[8,10] Also, in a recent paper by Liu et al.,[11] atomic structures of some oligomers of amyloidogenic sequences were obtained by incorporating them into macrocycles. Fibril formation is often described by a nucleation and growth mechanism, in which the proteins form intermediate oligomeric aggregates before they organize and grow into ordered fibrils with the characteristic cross-beta structure.[12] Intraresidue contacts have been measured by fluorescence to probe the process of self-assembly.[13−15] Quasi-elastic light scattering spectroscopy[16,17] has also been used for the same scope. Despite all these advances, 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.[18−20] Intermediate-resolution models with coarse-grained force fields or implicit solvent have been successfully investigated.[21−31] For instance, the aggregation of Aβ(16−22) hexamer has been studied in implicit solvent.[32] The aggregation of AcPHF6 segment with different numbers of chains has been simulated by a coarse-grained model.[33] On a more mesoscopic scale, the nucleation barrier of amyloid formation has been characterized using a tube model.[34] Using a more accurate explicit solvent description, the potential of mean force of peptide dimerization was calculated.[35] Relative stabilities of oligomers as well as mature filaments were also studied.[36] The binding free energies of several configurations of

polypeptide segments were estimated.[37] The structure and stability of oligomers of different sizes of the fragment Aβ(16–22) of the Alzheimer's β-amyloid peptide in several packing configuration have been also studied using atomic-detail molecular dynamics (MD) simulations.[38] 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 (BE-META),[39] to investigate the conformational free energy landscape of a set of 18 chains of polyvaline, each one 8 residues long. The system is described with an accurate all-atom explicit solvent force field.[40] The choice of a poly-amino acid is guided by the "generic hypothesis" of amyloid formation,[41] according to which the ability to assemble into ordered cross-beta structure is not a "strange" feature exhibited by a small group of peptides and proteins with special sequence or structural properties, but an inherent characteristic of polypeptide chains. The purpose of this work was to perform a simulation on a generic homopeptide and not on a peptide of primary sequence that favors fibril formation, with the goal of capturing sequence-independent features in the landscape. We simulate short peptides, because it has been reported by Sawaya et al.[9] that short segments of protein can assemble into structures which are similar to amyloids. In this perspective, valine has a steric hindrance that is close to the average of all the amino acids, making the structures that it forms prone to further analysis as a template for other amino acids. We choose the simulation box volume in such a way that the concentration of the system favors aggregated states. In this manner we focus the computational effort on the study of the ordering process within the aggregate. Using this approach we are able to compute the free energy of hundreds of putative aggregate structures, with various contents of parallel and antiparallel β-sheets and several different packings among the different sheets, characterizing in detail a possible nucleation pathway. The picture that emerges is quite surprising:

- Computing the free energy as a function of three selected reaction coordinates demonstrates that structures characterized by a specific intrasheet packing of parallel β-sheets form a large and well-defined free energy minimum, separated from the disordered aggregate by a barrier. This state is likely to be committed to a fully ordered amyloid-like aggregate.
- The natural reaction coordinate that one would use to describe the process, namely the total content of β-sheets, does not capture the presence of the amyloid-like minimum. Indeed, the free energy as a function of this variable is barrierless, leading to the (wrong) interpretation that all the structures, even with a high content of β-sheets, are kinetically committed to the disordered state.
- The minimum free energy path connecting the disordered aggregate with the amyloid-like free energy minimum involves first the formation of a nucleus of antiparallel β-sheets. Then, within this nucleus, a few parallel β-sheets start appearing, and when their number is sufficient the free energy starts to decrease toward the new minimum.

Even if the specific features of the free energy landscape are likely to depend on the primary sequence (polyvaline in our case), these 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 nontrivial pathway

that cannot be described by classical nucleation or a simple mesoscopic theory.

■ METHODS

Molecular Dynamics (MD). MD simulations were performed with the Amber99 force field[40] using GROMACS 4.0.7.[42] All the systems were solvated in TIP3P water[43] and equilibrated at 350 K by coupling the system to a Nosé−Hoover thermostat[44,45] and to a Parrinello−Rahman barostat,[46] both with a relaxation time of 1 ps. The particle-mesh Ewald method was used for long-range electrostatics with a short-range cutoff of 0.9 nm. A cutoff of 0.9 nm was used for the Lennard-Jones interactions. All bonds were constrained to their equilibrium length with the LINCS algorithm.[47] The time step for the MD simulation was 2 fs.

Bias-Exchange Metadynamics (BE-META).[39] We performed BE-META simulations using PLUMED.[48] The collective variables (CVs) specific to this system (see below) were coded by us in PLUMED. After 10 ns of equilibration, the BE-META simulation was started.

BE-META is a combination of replica exchange[49] and metadynamics[50] that allows reconstructing the free energy in a large number of CVs. In this method several copies (replicas) of the system are simulated, allowing them to periodically exchange conformation according to a replica exchange scheme. The bias acting on each replica is defined as

$$V_G(S(x), t) = w \sum_{\substack{t' = \tau_G, 2\tau_G, \cdots \\ t' < t}} \exp\left(-\frac{(S(x) - s(t'))^2}{2\delta s^2}\right)$$
(1)

where x represents the atomic coordinates of the system, s is the CV, and w and δs are the height and width of the Gaussian, respectively. After a certain time τ_{exch}, exchange between the different pairs of replicas is attempted using the Metropolis criterion. If the exchange move is accepted, the trajectory that was previously biased in the direction of the first variable continues its evolution biased by the second, and vice versa. In this manner, a large number of different variables can simultaneously be biased, and, ideally, the dimensionality of the space explored by metadynamics can be made so large that the residual barriers orthogonal to the reaction coordinates can be crossed in the available simulation time.

Collective Variables (CVs). The following set of CVs has been selected as putative reaction coordinates to explore the conformational regions of 18 chains of 8-valine (hereafter termed VAL8).

CV-A: coordination number, defined as

$$S_A = \sum_{i,j} C_{ij}$$

with

$$C_{ij} = \frac{1 - \left(r_{ij}/r_0\right)^n}{1 - \left(r_{ij}/r_0\right)^m}$$
(2)

where r_{ij} is the distance between atoms or groups i and j. We used this CV to count the number of contacts between side chains with $n = 6$, $m = 12$, and $r_0 = 0.6$ nm. The sum in eq 2 runs over the C_β atoms.

CV-B, antibetarmsd: This variable counts how many fragments of 3+3 residues in the protein chain form a β-sheet secondary structure, by computing their distance with respect to an ideal antiparallel β-sheet conformation:[51]

$$S_B = \sum_\alpha g[rdist(\{R_i\}_{i\in\Omega_\alpha}, \{R^0\})]$$

with

$$g(rdist) = \frac{1 - (rdist/rdist_0)^n}{1 - (rdist/rdist_0)^m}$$
(3)

where $\{R_i\}_{i\in\Omega_\alpha}$ are the atomic coordinates of a set Ω_α of six residues of the protein (including only N, C_α, C, O, and C_β atoms), while $\{R^0\}$ is the corresponding atomic position of the ideal β conformation. $rdist$ is

the root-mean-square deviation of the distance matrix calculated for $\{R_i\}_{i \in \Omega_\alpha}$ with respect to the distance matrix calculated for $\{R^0\}$. $g(rdist)$ is a function switching smoothly between 0 and 1. We used $n = 8$, $m = 12$, and $rdist_0 = 0.1$.

CV-C, parabetarmsd: This variable has the same definition as CV-B, but the reference structure is an ideal parallel β-sheet conformation;[51] therefore, it counts the number of parallel β-sheet blocks in the structure.

By changing the coordinates of the reference structure in eq 3, CVs of this form have been used for counting the number of parallel and antiparallel β-strands, and the packing of different layers on top of each other (see Figure 1; for details see SI).
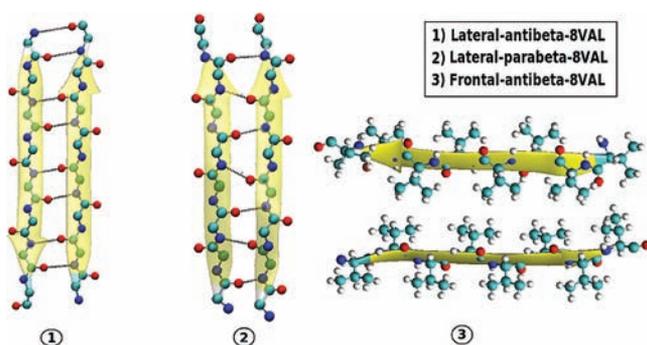

Figure 1. Templates 1, 2, and 3 that were used for CV-D(E), CV-F, and CV-G(H).

CV-D, number of antiparallel β-sheets: This variable counts how many pairs of VAL8 chains form a fully antiparallel β-strand. This is estimated by computing their $rdist$ with respect to the structure template 1 in Figure 1, with $n = 3$, $m = 5$, and $rdist_0 = 0.1$.

CV-E: In order to control the sensibility toward the β-strand template structure, we also included this variable with the same definition of CV-D, but with $rdist_0 = 0.2$. The first set of parameters (in CV-D) allows counting the number of β-strands only when they are nearly perfect as compared to the template. The second set of parameters (in CV-E) is less strict with respect to the exact shape of the β-strand. The metadynamics bias potential acting on the latter drives the system toward β-like structures when the system is fully disordered.

CV-F, number of parallel β-sheets: This variable counts how many pairs of VAL8 chains form a fully parallel β-strand. This is estimated by computing their $rdist$ with respect to the structure template 2 in Figure 2, with $n = 3$, $m = 5$, and $rdist_0 = 0.2$.

CV-G, number of antiparallel steric zippers: This variable counts how many antiparallel steric zippers exist in a structure by estimating the $rdist$ of each pair of VAL8 chains with respect to the structure template 3 in Figure 1, with $n = 3$, $m = 5$, and $rdist_0 = 0.1$.

CV-H: Same definition as CV-G, but with $rdist_0 = 0.2$.

An α helix could be also considered as a relevant CV for this simulation; however, our tests showed that formation of an α helix is a "fast" process in our simulation time. Therefore, it is not necessary to dedicate a CV to bias this observable. The parameters adopted in BE-META simulations were as follow: Gaussian height, 2 kJ/mol; Gaussian width, 3 for CV-A, 0.2 for CV-B and CV-C, and 0.1 for CV-D to CV-H; deposition time τ_G of the Gaussians, 5 ps; exchange of bias attempted every 50 ps (τ_{exch}). The procedure for choosing an optimal set of parameters is explained in ref 52. BE-META simulations were performed biasing 8 replicas, one for each CV for a total simulation time of 4320 ns, 540 ns per replica. The boundaries were treated following the procedure explained in ref 53. The initial configuration for all replicas was a completely disordered aggregate.

Cluster Analysis and Thermodynamic Model. In BE-META the convergence of the bias potential $V_G(s,t)$ in eq 1 is monitored like in standard metadynamics:[54] after a transient time t_{eq}, $V_G(s,t)$ reaches a

stationary state in which it grows evenly, fluctuating around an average profile. The free energy as a function of s is estimated as the time average of $V_G(s,t)$:

$$-F(s) \approx \overline{V_G}(s) = \frac{1}{t_{sim} - t_{eq}} \int_{t_{eq}}^{t_{sim}} dt V_G(s, t) \tag{4}$$

where t_{sim} is the total simulation time. Convergence is evaluated independently over the profile reconstructed by each replica. After convergence is achieved, one can reconstruct the free energy as a function of several CVs by using the approach of ref 55. In short, the CV space is subdivided so that all the frames of the BE-META trajectories are grouped in sets (clusters) whose members are close to each other in CV space. Then, the free energy F_α of each cluster α is estimated by a weighted-histogram analysis approach (WHAM).[55,56] The free energy of cluster α is given by

$$F_\alpha = -k_B T \log \frac{\sum_i n_\alpha^i}{\sum_j e^{(f^j - V_\alpha^j)/k_B T}} \tag{5}$$

where n_α^i is the number of times cluster α is observed in the trajectory of replica i and V_α^i is the bias potential acting on cluster α in replica i. V_α^i is estimated as the time average of the history-dependent potential acting on replica i evaluated in s_α, the center of cluster α:

$$V_\alpha^i = \overline{V_G^i}(s_\alpha) = \frac{1}{t_{sim} - t_{eq}} \int_{t_{eq}}^{t_{sim}} dt V_G^i(s_\alpha, t)$$

The cluster-based thermodynamic model was constructed employing the METAGUI program.[57] An important issue in this procedure is how many and which CVs to use for the clusterization. Of course, it is not necessary to use all the CVs that have been explicitly biased, as some of these CVs might prove to be less relevant for the process or strongly correlated with other variables. In order to find the optimal set of CVs for clustering, we have done several tests with different combinations of CVs. For each setup we monitored how many different free energy minima are obtained and the level of convergence in the free energy of the clusters. At the end the best setup gives two distinct free energy minima, connected to each other in the space of CVs through a set of well-populated clusters. A satisfactory description of the thermodynamics of VAL8 aggregation is achieved employing the three variables CV-C, CV-E, and CV-H (for details see SI).

RESULTS

Plain Molecular Dynamics. We started our investigation by performing plain MD simulations with a number of chains of VAL8 ranging from 2 to 18, arranged in a single, double, or triple layer (see Figure 2). The chains of VAL8 were capped with ACE and NME peptide at two ends. All the simulations are performed in explicit solvent and at a temperature of 350 K (see Methods). Most of the systems are simulated for 100 ns. In Figure 2, the starting and final configurations of some of these simulations are presented. Single layer β-sheets of 2 chains and 4 chains lost their conformation in a few nanoseconds. Consistent with this result, in a bias-exchange simulation including 2 chains starting from disordered structure, we found that the antiparallel and parallel dimers are unstable, with a free energy of formation of approximately 10 kcal/mol. Double layers of 8 chains are stable after 100 ns. Triple layers of 9 chains twist slightly around the axis perpendicular to the β-sheets but still maintain their β content. We also performed a simulation for 18 chains of VAL8 in triple layers, which were also very stable. These calculations clarify that both lateral (in plane) and frontal (plane in front of each other) stackings of strands are essential to determine the stability. Specially, frontal packing termed *steric zipper*,[8,9] in which the side chains of the opposing β-sheets tightly

System	Starting Configuration		Final Configuration	Simulation time
2 chains - VAL8				100 ns
4 chains - VAL8				10 ns
4 chains - VAL8				100 ns
8 chains - VAL8				100 ns
8 chains - VAL8				100 ns
9 chains - VAL8				100 ns
9 chains - VAL8				100 ns
18 chains - VAL8				20 ns
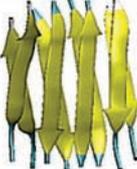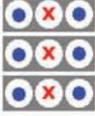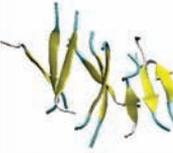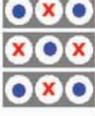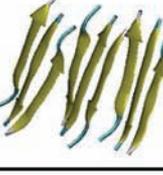
Figure 2. Structural evolution during the MD simulation of several configurations of VAL8 system with different number of chains (ranging between 2 and 18). The first column reports the number of chains. The second column is a schematic representation of the starting configuration, in which the gray layers are β-sheets, the circle with a central dot represents a β-strand pointing outside the paper plane, and the circle with a cross represents a β-strand pointing inside the paper plane. The third and fourth columns are cartoon representation of the structures before and after the MD simulation. The last column is the simulation time.

compenetrate each other, leads to very stable conformations. Therefore, in the next simulation, we decided to bias the presence of this conformational motif with a dedicated CV (see Methods).

These simulations allow us to conclude that ordered aggregates formed by less than 6 chains are for sure unstable. Aggregates formed by 8 chains or more are instead stable on the time scale of 100 ns. However, the results do not provide
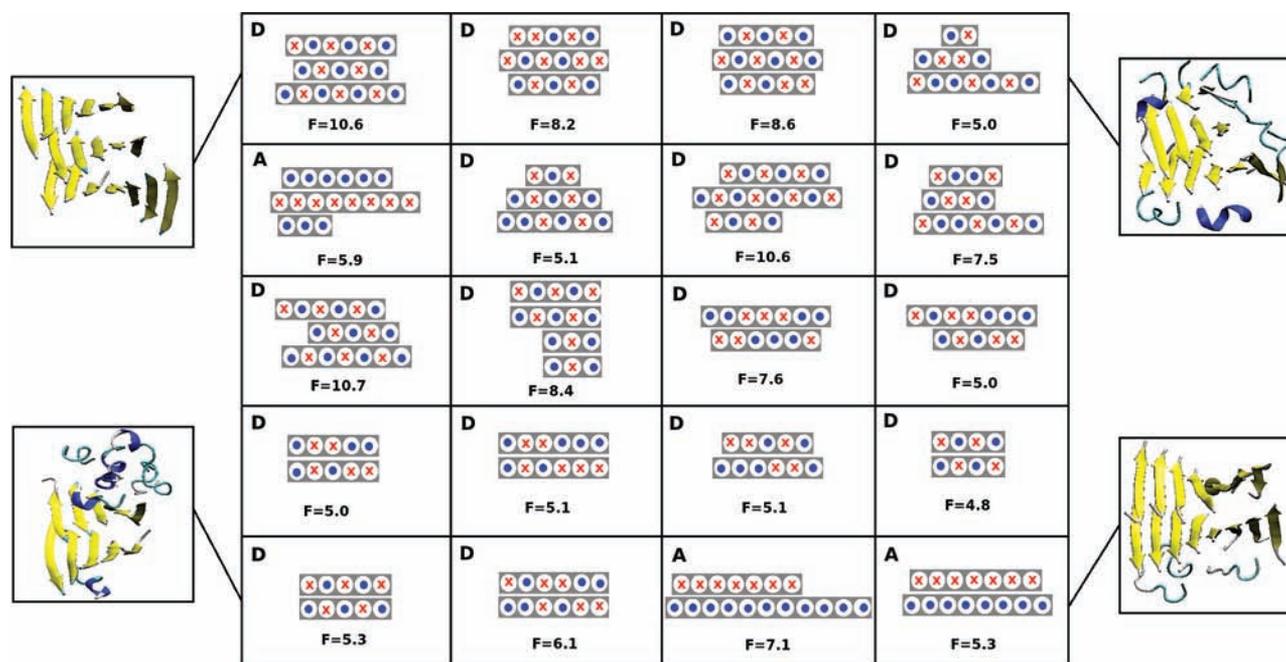
Figure 3. Twenty representative structures obtained from BE-META simulation performed on the system of 18 chains of VAL8 in explicit solvent at 350 K. The structures are represented like in Figure 2. We also show cartoon representations for 4 of the structures. The free energy of each structure is also reported. Label A means that the structure is committed to region 3 in Figure 5; label D means that the structure is committed to the disordered aggregate (region 1 in Figure 5).

decisive evidence neither on the size of the critical nucleus (the "stable" structures in figure could break apart on a longer time scale) nor on the relative stability of the possible structures. In particular, from MD simulations of this duration it is not possible to decide if the system prefers to form parallel or antiparallel β-sheets, and how the different sheets preferentially pack together. The set of CVs that are biased in the following BE-META simulation is chosen exactly in order to provide quantitative evidence on these issues.

Bias-Exchange Metadynamics. We then performed a BE-META simulation on a system of 18 chains (the chains were capped with ACE and NME peptide at two ends). This relatively large system size has been chosen in order to have a high chance of observing a critical nucleus (from the plain MD results it is clear that already aggregates of 8−9 chains are stable on the 100 ns time scale). We used 8 replicas, all at the same temperature (350 K), each biased along a different CV (see Methods). The high temperature was chosen to make exploration faster. The exchange rate between replicas was 50 ps (for further details on the optimal choice of BE-META parameters, see ref 52). The CVs capture the geometrical features of amyloid structures, namely, lateral packing of polypeptide chains in a β-sheet configuration and subsequently frontal packing of β-sheets and in multiple layers. This set of CVs was used in an attempt to include all degrees of freedom which may be important for distinguishing the relevant structures. The simulation was run for a total time of 4340 ns, 540 ns for each of the 8 replicas, starting from a completely disordered aggregate. Due to the action of the bias, the system explores hundreds of configurations different in the number of layers (single, double, triple, and quadruple) and also in the way β-strands pack, laterally (antiparallel β-sheet and parallel β-sheet) or frontally (parallel or antiparallel β-strands in different layers face each other). Twenty of the structures are represented in Figure 3 with their corresponding free energy (see below and Methods).

After 300 ns, the history-dependent potentials acting on the replicas reach a stationary state: in the relevant region, they grow evenly, fluctuating around a stable estimator (see Figure 1 in SI). After this time, trajectories were analyzed following the procedure in ref 55, namely by dividing the CV space in a hypercubic grid. All the structures whose CV values fall in the same hypercube define a cluster. The equilibrium free energy of each cluster is then computed by removing the effect of the bias potentials from the populations. This allowed us to obtain a converged estimate of the free energy of almost 500 structures (see Methods). A representative configuration for some of these structures is provided in Figure 3, together with the estimated free energy.

Classical Nucleation Picture. In the attempt to explain the formation of an ordered aggregate according to classical nucleation theory, we first used the results obtained from BE-META simulations to compute the free energy profile as a function of a single order parameter, namely the number of β-strands. To do this, we first computed the average number of β-strands in each cluster (n_α). The free energy as a function of the number of strands n is then given by

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

(6)

where F_α is the free energy of cluster α and $\chi_n(n_\alpha)$ is equal to 1 if $n_\alpha \in [n - 1/2, n + 1/2]$ and 0 otherwise. The result is shown in Figure 4. No barrier in the free energy as a function of the number of β-strands is present. The disordered aggregate is the minimum in the free energy. Moving toward structures with a higher β-sheet content, 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. Finally, for $n > 16$ the free energy sharply increases. From this graph one would conclude that the critical
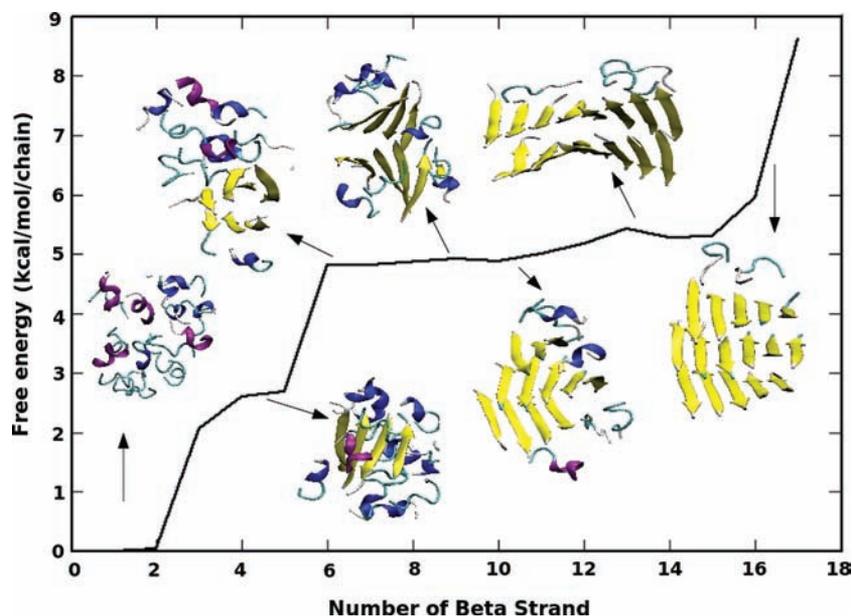
**Figure 4.** Free energy as a function of the number of β-strands.

Figure 5. Free energy as a function of three CVs: CV-C is the number of antiparallel β-sheets, CV-E is the number of parallel β-sheets, and CV-H is the number of antiparallel steric zippers. The red isosurface represents the region of CV space explored during the simulation. The blue isosurface corresponds to a free energy of 3.3 kcal/mol per chain and contains the global minimum, a disordered aggregate. The orange isosurface, at 5.8 kcal/mol per 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 seven representative structures along the path, with their corresponding free energies.

nucleus is even larger than $n = 18$, or that ordered structures are intrinsically metastable and committed to the disordered aggregate.

Three-Dimensional Picture of the Nucleation Process. Since the free energy profile as a function of a single coordinate measuring the number of β-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 antiparallel and parallel packing of β-strands inside a layer and the steric zipper packing of the β-strands in front of each other (CV-C, CV-E, and CV-H, see Methods). A volumetric representation of the free energy with respect to these three CVs is

presented in Figure 5. 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 per chain (light orange region labeled with 2 in Figure 5). It mostly consists of structures with a high content of antiparallel β-sheets, that seem to form more easily from the disordered aggregate. At variance with what is observed in the one-dimensional free energy landscape, in three dimensions a well-defined free energy minimum is present (labeled 3 in Figure 5). The minimum in this region is at 5.3 kcal/mol per chain, very close to the free energy in the plateau, but this region is separated from the rest by a barrier of at least 2 kcal/mol per 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 β-sheets.

Dependence on the Force Field. A critical issue in computer simulation is the choice of the force field; here we use Amber99.[40] In fact, it has been reported that the Amber99 force field has higher α-helical propensities relative to experiment.[58,59] In order to investigate the effect of the force field in our simulation, we compared the result of NPT simulations with the results from two different force fields, Amber99[40] and Amber99SB,[58] for some selected configurations. We considered one structure corresponding to the disordered aggregate, one to the plateau region rich in antiparallel β-sheets, and one to the parallel β-sheets minimum (see Figure 5). In all the simulations, the structures are very stable for all 100 ns of the simulations, indicating that the two force fields provide a qualitatively similar picture of the kinetics of the system on this time scale. However, the enthalpy of the disordered structure is significantly different when evaluated with the two force fields (see Figure 6). While with Amber99 the three structures have basically the same average enthalpy, with Amber99SB structures rich in β-sheets are favored by 7 kcal/mol per chain with respect to the disordered aggregate. The enthalpy difference between the two structures rich in β-sheets is instead approximately zero in both force fields (within the uncertainty of a relatively short 50 ns simulation).

DISCUSSION

The aggregation process of 18 chains of polyvaline has been investigated by MD 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 8 reaction coordinates. These have been chosen in an attempt of capturing the most important degrees of freedom associated with aggregation, especially the parallel and antiparallel arrangements of strands in a β-sheet layer (lateral packing) and the arrangement of β-strands in layers facing each other (frontal packing). Each of the 8 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 β 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 different primary sequences.

The free energy as a function of three coordinates, CV-C, CV-E, and CV-H (see Methods), shows that while structures rich in antiparallel β-sheets belong to the same free energy minimum of the disordered aggregate, structures with an even smaller fraction of β-sheets, but rich in parallel strands, form a well-defined minimum, separated from the disordered aggregate by a relatively high barrier. Comparison of Figures 4 and 5 shows that for this system projecting the free energy on a single "natural" reaction coordinate does not capture the qualitative features of the process.[60] The minimum free energy path is shown in Figure 5 as a red line. Clearly, at least in the space of the three variables chosen for the representation, the path is highly nontrivial: first the system forms antiparallel β-sheets that are favored over parallel ones when the structure
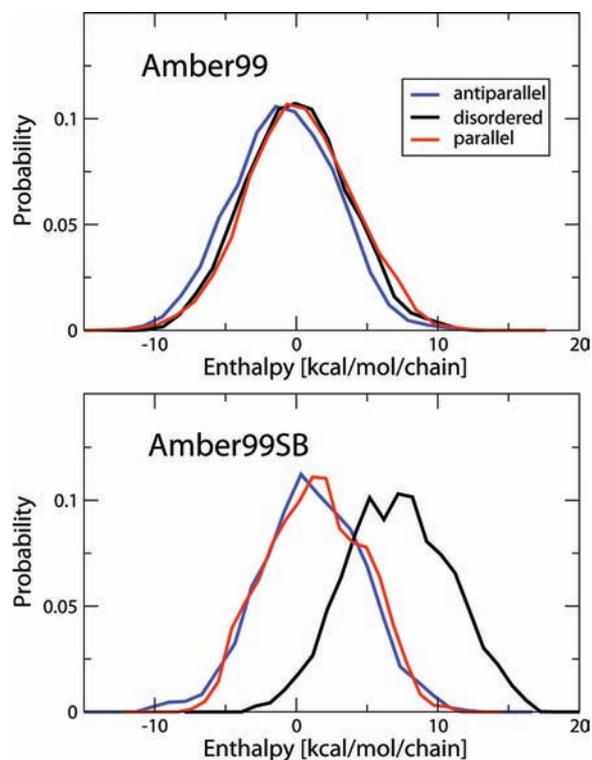

Figure 6. Histogram of averaged enthalpy for 50 ns of NPT simulation at 350 K and 1 bar for three relevant structures (see text). Before computing the histogram, a running average on a window of 0.5 ns was performed to reduce the magnitude of fluctuations and highlight the difference between different simulations. The upper panel shows the result obtained using the Amber99 force field.[40] The lower panel shows the result obtained using the Amber99SB force field.[58]

is highly disordered. Then, within a relatively large ordered nucleus formed mostly by antiparallel β-sheets, a few parallel β-sheets start appearing. The relevant nucleation process seems to happen at this point: When a sufficient number of parallel sheets are formed, the free energy finally starts to decrease toward a new minimum, in which parallel β-sheets are predominant. Interestingly, this conformational change from antiparallel to parallel β-sheet has been found also by Li et al.[33] and Nasica-Labouze et al.[23] 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 β-sheets, which is much larger in the structures close to the free energy minimum. The structures close to the barrier have only a few parallel β-sheets, surrounded by antiparallel β-sheets.

The question that immediately arises is whether the parallel β-sheet-rich structure corresponds indeed to the structure of a nascent amyloid. Clearly, its free energy is significantly higher than that of a disordered aggregate. This could be an effect of the finite size of the system: if the critical nucleus is approximately formed by 14 chains, with 18 chains the free energy has just started to decrease, and one can argue that adding more chains would lower the free energy of the ordered aggregate. In order to prove this claim, one should perform simulations on an even larger system, which are prohibitively expensive with the available computational resources. Moreover, our investigation on the dependence of our results on the choice of force field indicates that while the relative stability of structures rich in β-sheets is only marginally affected by the

choice of the force field, with Amber99SB the disordered aggregate is significantly enthalpically destabilized. Thus, the relative depths of the structured and the unstructured minima would probably be different from those reported here for a simulation fully performed with Amber99SB, with the parallel β-sheet-rich minimum further stabilized. It is worth noting that the PASTA[61] server for VAL8 predicts amyloids with parallel register of the β-sheets.

It has to be emphasized that, in spite of the fact that amyloid formation is considered a generic property of polypeptide chains,[7] sequence is known to affect the specific dynamics of amyloid formation.[24,62,63] For instance, at variance with many studies in implicit solvent[25,32,33,64] and explicit solvent[65] on the aggregation of amyloidogenic peptide segments, we did not observe any β-barrel-like structure. This difference could be a consequence of the specific and "trivial" primary sequence we use in this work. Moreover, the relative propensity of forming parallel or antiparallel β-strands might depend not only on the sequence but also on the length of the peptide.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information
Procedure to build the templates used in the definition of the CVs; criterion to choose the relevant CVs for cluster analysis together with a table that shows the results of all different trials for different combination of CVs; plot that shows the parallel growth of the biasing potential along each of the CVs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
laio@sissa.it
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(2) Hartley, D. M.; Walsh, D. M.; Ye, C. P.; Diehl, T.; Vasquez, S.; Vassilev, P. M.; Teplow, D. B.; Selkoe, D. J. *J. Neurosci.* **1999**, *19*, 8876−8884.
(3) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507−511.
(4) Brändén, C.; Tooze, J. *Introduction to protein structure*; Garland: New York, 1999.
(5) Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. *J. Mol. Biol.* **1997**, *273*, 729−739.
(6) Sipe, J. D.; Cohen, A. S. *J. Struct. Biol.* **2000**, *130*, 88−98.
(7) Fändrich, M.; Dobson, C. M. *EMBO J.* **2002**, *21*, 5682−5690.
(8) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773−778.
(9) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447*, 453−457.
(10) Rochet, J. C.; Lansbury, P. Jr. *Curr. Opin. Struct. Biol.* **2000**, *10*, 60−68.
(11) Liu, C.; Sawaya, M.; Cheng, P.; Zheng, J.; Nowick, J.; Eisenberg, D. *J. Am. Chem. Soc.* **2011**, *133*, 6736−6744.
(12) Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L. *Science* **2000**, *289*, 1317−1321.
(13) Ban, T.; Hoshino, M.; Takahashi, S.; Hamada, D.; Hasegawa, K.; Naiki, H.; Goto, Y. *J. Mol. Biol.* **2004**, *344*, 757−767.
(14) Krishnan, R.; Lindquist, S. L. *Nature* **2005**, *435*, 765−772.
(15) Bieschke, J.; Zhang, Q.; Powers, E. T.; Lerner, R. A.; Kelly, J. W. *Biochemistry* **2005**, *44*, 4977−4983.
(16) Lomakin, A.; Teplow, D. *Protein Pept. Lett.* **2005**, *13*, 247−254.
(17) Lomakin, A.; Teplow, D. B.; Benedek, G. B. *Methods Mol. Biol.* **2005**, *299*, 153−174.
(18) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111−116.
(19) Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *360*, 882−892.
(20) Ma, B.; Nussinov, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 445−452.
(21) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301−2310.
(22) Derreumaux, P. *Phys. Rev. Lett.* **2000**, *85*, 206−209.
(23) Nasica-Labouze, J.; Meli, M.; Derreumaux, P.; Colombo, G.; Mousseau, N. *PLoS Comput. Biol.* **2011**, *7*, e1002051.
(24) Lu, Y.; Derreumaux, P.; Guo, Z.; Mousseau, N.; Wei, G. *Proteins: Struct., Funct. Bioinf.* **2009**, *75*, 954−963.
(25) Wei, G.; Song, W.; Derreumaux, P.; Mousseau, N. *Front. Biosci.* **2008**, *13*, 5681−5692.
(26) Mousseau, N.; Derreumaux, P. *Acc. Chem. Res.* **2005**, *38*, 885−891.
(27) Wei, G.; Mousseau, N.; Derreumaux, P. *Biophys. J.* **2004**, *87*, 3648−3656.
(28) Favrin, G.; Irbäck, A.; Mohanty, S. *Biophys. J.* **2004**, *87*, 3657−3664.
(29) Nguyen, H.; Hall, C. *J. Am. Chem. Soc.* **2006**, *128*, 1890−1901.
(30) Kelley, N.; Vishal, V.; Krafft, G.; Pande, V. *J. Chem. Phys.* **2008**, *129*, 214707.
(31) Fawzi, N.; Okabe, Y.; Yap, E.; Head-Gordon, T. *J. Mol. Biol.* **2007**, *365*, 535−550.
(32) Irbäck, A.; Mitternacht, S. *Proteins* **2008**, *71*, 207−214.
(33) Li, D.; Mohanty, S.; Irbäck, A.; Huo, S. *PLoS Comput. Biol.* **2008**, *4*, e1000238.
(34) Auer, S.; Dobson, C. M.; Vendruscolo, M. *HFSP J.* **2007**, *1*, 137−146.
(35) Gnanakaran, S.; Nussinov, R.; Garcia, A. E. *J. Am. Chem. Soc.* **2006**, *128*, 2158−2159.
(36) Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14126−14131.
(37) Park, J.; Kahng, B.; Hwang, W. *PLoS Computat. Biol.* **2009**, *5*, e1000492.
(38) Röhrig, U. F.; Laio, A.; Tantalo, N.; Parrinello, M.; Petronzio, R. *Biophys. J.* **2006**, *91*, 3217−3229.
(39) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(40) Wang, J.; Cieplak, P.; Kollman, P. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(41) Dobson, C.; Karplus, M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 92−101.
(42) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(43) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
(44) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.
(45) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(46) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(47) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(48) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961.
(49) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9−12.
(50) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(51) Pietrucci, F.; Laio, A. *J. Chem. Theory Comput.* **2009**, *5*, 2197.

(52) Cossio, P.; Marinelli, F.; Laio, A.; Pietrucci, F. *J. Phys. Chem. B* **2010**, *114*, 3259−3265.

(53) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. *Curr. Phys. Chem.* **2012**, *2*, 79−91.

(54) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(55) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. *PLoS Comput. Biol.* **2009**, *5*, e1000452.

(56) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P. A.; Rosenberg, J. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(57) Biarnés, X.; Pietrucci, F.; Marinelli, F.; Laio, A. *Comput. Phys. Commun.* **2012**, *183*, 203−211.

(58) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct. Bioinf.* **2006**, *65*, 712−725.

(59) Best, R.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.

(60) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.

(61) Trovato, A.; Seno, F.; Tosatto, S. *Protein Eng. Des. Sel.* **2007**, *20*, 521−523.

(62) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L.; Dobson, C.; Vendruscolo, M.; Favrin, G. *PLoS Comput. Biol.* **2007**, *3*, e173.

(63) Reddy, G.; Straub, J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21459.

(64) Bellesia, G.; Shea, J. *J. Chem. Phys.* **2009**, *130*, 145103.

(65) De Simone, A.; Derreumaux, P. *J. Chem. Phys.* **2010**, *132*, 165103.