Aggregate Geometry in Amyloid Fibril Nucleation

Anders Irbäck, Sigurður Æ. Jónsson, Niels Linnemann, Björn Linse, and Stefan Wallin

Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University,
Sölvegatan 14A, SE-223 62 Lund, Sweden
(Received 1 September 2012; published 28 January 2013)

We present and study a minimal structure-based model for the self-assembly of peptides into ordered β-sheet-rich fibrils. The peptides are represented by unit-length sticks on a cubic lattice and interact by hydrogen bonding and hydrophobicity forces. Using Monte Carlo simulations with $>10^5$ peptides, we show that fibril formation occurs with sigmoidal kinetics in the model. To determine the mechanism of fibril nucleation, we compute the joint distribution in length and width of the aggregates at equilibrium, using an efficient cluster move and flat-histogram techniques. This analysis, based on simulations with 256 peptides in which aggregates form and dissolve reversibly, shows that the main free-energy barriers that a nascent fibril has to overcome are associated with changes in width.

Many proteins and peptides share the ability to self-assemble into amyloid fibrils, aggregates with a cross-β structure and remarkable mechanical properties, that are associated with a range of disorders as well as with functional roles [1,2]. The formation of amyloid fibrils, usually monitored by thioflavin T (ThT) fluorescence, is known to occur with reproducible sigmoidal kinetics [3], indicating a nucleation-dependent process. A powerful method for interpreting the experimental kinetic profiles is by means of rate equations [4]. This approach can reveal some general properties of intermediate species participating in the growth process. It has proven useful for some related self-assembly phenomena as well, such as hemoglobin S aggregation [5] and microtubule assembly [6]. Another method to elucidate the mechanisms of amyloid formation is by phase equilibria analysis [7].

By coarse-grained structure-based approaches [8], additional insights have been gained into the nucleation of amyloid fibrils [9–14]. Fibrillation pathways involve, however, a host of different aggregated species of widely varying size, and studying the competition among these species without restrictive assumptions represents a challenge even in coarse-grained models.

In this Letter, we introduce a minimal structure-based model that describes amyloid fibril formation in terms of physically inspired peptide-peptide interactions and yet allows for representative sampling of the model state space for relatively large systems. Using flat-histogram methods [15,16] and an efficient cluster move resembling the Swendsen-Wang algorithm for spin systems [17], we determine equilibrium distributions in size and shape of the aggregated structures, in order to elucidate the free-energy landscape that a nascent fibril has to navigate.

We consider N identical peptides, represented by unit-length sticks on a periodic cubic lattice with dimensions L^3. We assume that the internal dynamics of a peptide are

fast compared to the time scales for fibril formation, and therefore can be averaged out.

Each peptide i is centered at a lattice site $\mathbf{r}_i$, and two peptides cannot simultaneously occupy the same site. Associated with each peptide are two unit vectors $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{p}}_i$ that can point in any of the six lattice directions [Fig. 1(a)]; $\hat{\mathbf{b}}_i$ represents the N-to-C backbone orientation, whereas $\pm\hat{\mathbf{p}}_i$ are the directions in which hydrogen bonds can form. The vectors $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{p}}_i$ are perpendicular, leaving a total of 24 possible orientations of a peptide. The vectors $\hat{\mathbf{s}}_i = \hat{\mathbf{b}}_i \times \hat{\mathbf{p}}_i$ and $-\hat{\mathbf{s}}_i$ represent side-chain directions. The $+\hat{\mathbf{s}}_i$ and $-\hat{\mathbf{s}}_i$ sides of a peptide are assumed to have different interaction properties and are referred to as hydrophobic and polar, respectively.

The energy function describing the interactions between the peptides is assumed pairwise additive, $E = \sum_{i<j} \epsilon_{ij}$, where $\epsilon_{ij} \leq 0$. The pair potential ϵ_{ij} is nonzero only if (i) peptides i and j are nearest neighbors on the lattice, and (ii) $\hat{\mathbf{b}}_i$ and $\hat{\mathbf{b}}_j$ are perpendicular to $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ and aligned either parallel or antiparallel to each other. When these conditions are met, we set $\epsilon_{ij} = -1$ except in the three cases illustrated in Fig. 1. The first two cases correspond to parallel [Fig. 1(b)] and antiparallel [Fig. 1(c)] β structure, respectively, and the third [Fig. 1(d)] to hydrophobic side-chain attraction. The corresponding interaction energies are given by

$$\epsilon_{ij} = \begin{cases} -(1 + a_{\mathrm{p}}) & \text{parallel } \beta \text{ structure} \\ -(1 + a_{\mathrm{ap}}) & \text{antiparallel } \beta \text{ structure} \\ -(1 + b) & \text{hydrophobic attraction.} \end{cases} \quad (1)$$

The hydrophobic attraction is included because of evidence suggesting that a pairwise (steric zipper) β-sheet organization is a common architecture for the core of amyloid fibrils [18]. The b parameter must not be too large, in order for extended β sheets to form. Because the β sheets often are parallel in amyloid fibrils, we take

(a)

(b) (c) (d)
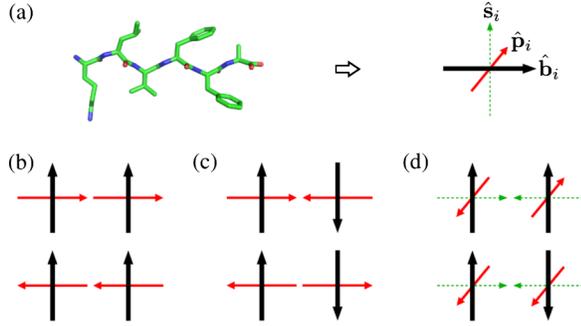
FIG. 1 (color online). Schematic illustration of the geometry and interactions of the model. (a) The orientation of a peptide is defined by the backbone vector $\hat{\mathbf{b}}_i$ (thick line) and the hydrogen-bond direction $\hat{\mathbf{p}}_i$ (thin line). The side-chain direction $\hat{\mathbf{s}}_i$ (dots) is given by $\hat{\mathbf{s}}_i = \hat{\mathbf{b}}_i \times \hat{\mathbf{p}}_i$. (b) Parallel β structure. (c) Antiparallel β structure. (d) Hydrophobic attraction. The β-structure definitions are such that a β sheet has one hydrophobic and one polar side.

$a_\mathrm{p} > a_\mathrm{ap}$, but the model can also be studied for $a_\mathrm{ap} > a_\mathrm{p}$. In what follows, for simplicity, we stick to a single parameter set, namely $a_\mathrm{p} = 5$, $a_\mathrm{ap} = 3$, and $b = 1$. With this choice, the parallel β-strand organization dominates, but the suppression of antiparallel strand pairs is not prohibitively strong.

We simulate the thermodynamics of this model using single-peptide as well as cluster moves. A cluster update makes it possible for aggregates to move without having to be first dissolved and then reassembled. To be able to also split and merge aggregates, we follow a stochastic Swendsen-Wang–type cluster construction procedure [17]. The construction is recursive and begins by picking a random first cluster member, i. Then, all peptides j interacting with peptide i ($\epsilon_{ij} < 0$) are identified and added to the cluster with probability $p_{ij} = 1 - e^{\beta\epsilon_{ij}}$, where $\beta = 1/k_BT$ is inverse temperature. This step is iterated until no cluster member has any unchecked interaction partner. Finally, the resulting cluster is subject to a trial rigid-body translation or rotation, drawn from a symmetric distribution, which is accepted whenever it does not cause any steric clashes. It can be verified that this algorithm fulfills detailed balance with respect to the canonical ensemble $p_\nu \propto e^{-\beta E_\nu}$.

To further enhance the sampling, we employ generalized-ensemble methods [15,16], along with reweighting techniques [19]. After estimating the density of states $g(E)$ by the Wang-Landau method [16], we simulate the ensemble $p_\nu \propto 1/g(E_\nu)$ [15], where the distribution of E is flat. This approach was recently used for atomic-level aggregation simulations [20] and is useful for the present system as well, which displays phase coexistence at the fibrillation temperature, T_m (see below). Our simulations sample a limited energy range, $E_\mathrm{min} < E \le 0$. The cutoff E_min is needed to prevent the formation of artificial cyclic aggregates, which otherwise may occur due to the periodic boundary conditions, but is sufficiently low to permit studies of temperatures in the fibrillar phase.

The above cluster update can be adapted for the generalized-ensemble simulations by adding an accept or reject step, with acceptance probability $p_\mathrm{acc}(\nu \to \nu') = \min[1, g(E_\nu)e^{-\beta E_\nu}/g(E_{\nu'})e^{-\beta E_{\nu'}}]$. Here, β changes its meaning to become a tunable algorithm parameter. We did not fine-tune β, but expect the optimal β to be in the neighborhood of $\beta_\mathrm{m} = 1/k_B T_\mathrm{m}$, as supported by preliminary runs.

Using these methods, we studied the thermodynamics of the model for several different system sizes. Here, we focus on the results obtained for $N = 256$ and $L = 64$, corresponding to a peptide concentration of $\rho \approx 10^{-3}$ per unit volume. This system size would have been very time-consuming to study with standard Monte Carlo methods.

In our simulations, two distinct major phases occur: a high-energy phase dominated by small aggregates and a low-energy phase where large fibril-like aggregates are present. As displayed in Fig. 2(a), at the midpoint temperature, $T_\mathrm{m} \approx 0.6714$, the energy distribution is bimodal, showing that the two phases coexist. Figure 2(b) shows the aggregate mass distribution, $p(m)$, at T_m, which gives the probability for a peptide to be part of an aggregate with m peptides ($m = 1$ corresponds to free monomers). Like the energy distribution, $p(m)$ is bimodal. The mass fractions of aggregates with $m \le 6$, $6 < m \le 62$, and $m > 62$ are 81.4%, 1.9%, and 16.7%, respectively. Small aggregates are present in both phases and constitute a large fraction of
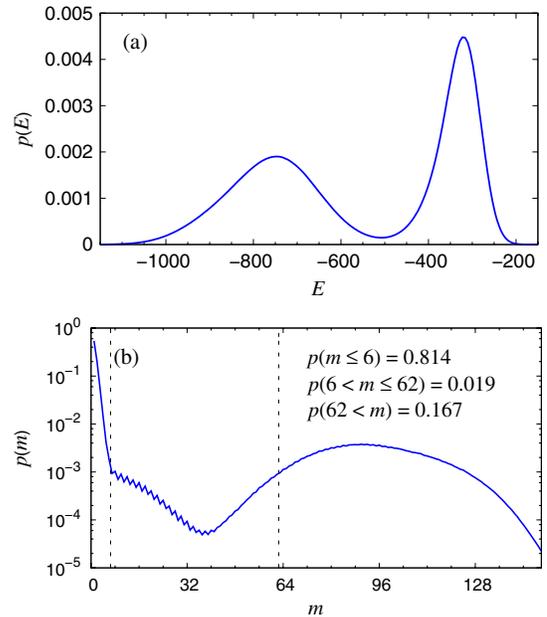
FIG. 2 (color online). Overall thermodynamic properties of the $N = 256$, $L = 64$ system at $T_\mathrm{m} \approx 0.6714$. (a) Energy distribution. The sampled range is $-1300 < E \le 0$. Smoothing was applied to remove short-scale irregularities in $g(E)$. (b) Mass fraction of aggregates with mass m, $p(m)$, against m. Summed probabilities for three regions in m are indicated. Statistical errors on both $p(E)$ and $p(m)$ are comparable to the line width.

the total mass at T_m (see also Fig. S1 in the Supplemental Material [21]).

At first glance, the bimodality of $p(m)$ may seem to indicate that fibril nucleation occurs when a critical mass is reached. However, this picture is geometrically incomplete, because the species involved are neither strictly one dimensional nor sharing one common shape, such as spherical. A simple but useful way to extend the analysis is via the inertia tensor. As measures of the length and width of an aggregate, we define $l = \sqrt{12\lambda_1 + 1}$ and $w = \sqrt{12\lambda_2 + 1}$, where $\lambda_1 \geq \lambda_2$ are eigenvalues of the inertia tensor. In our model, there is no interaction between longitudinally aligned peptides to support growth in a third dimension. With these definitions, for a rectangular aggregate, l and w are the numbers of peptide layers in the two directions.

Figure 3 shows the probability $p(l, w)$ for a peptide to be part of an aggregate with length l and width w, at T_m. Consistent with Fig. 2(b), $p(l, w)$ is highest for aggregates with l small and $w \approx 1$. Among larger aggregates, a clear preference can be seen for even over odd values of w, reflecting a pairwise β-sheet organization, although aggregates with six or more layers are severely constrained by finite-size effects. A second trend is that single-layer aggregates are shorter than two-layer ones, which in turn are shorter than those with four layers. We expect this trend to persist beyond the four-layer level if the system is sufficiently large. These overall features of $p(l, w)$ are likely to be quite robust, although the locations of the different maxima depend on both T and ρ.

The shape of $p(l, w)$ has implications for how fibrils nucleate and grow in the model. It suggests that the main free-energy barriers faced by a growing aggregate are associated with changes in width, and it must increase in width to be able to grow.

Having examined the thermodynamics of the model, we now turn to the aggregation kinetics, studied using constant-temperature Monte Carlo dynamics. Because of evidence that amyloid growth occurs by monomer addition [22], here we use single-peptide moves only. The simulations start from random initial conditions and the temperature is $T = 0.66$.
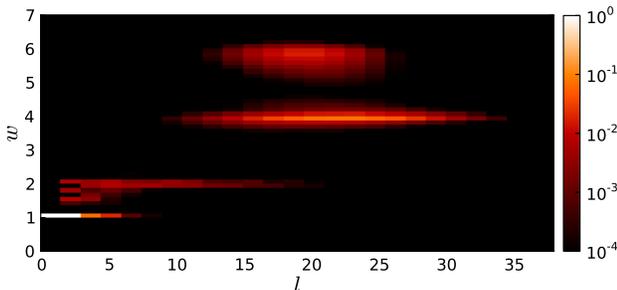
We first stick to the system size $N = 256$ and $L = 64$, which is useful for examining the formation of individual fibrils. Figure 4(a) shows the mass of the largest aggregate, $m_0(t)$, against Monte Carlo time t, in two representative runs. Both runs exhibit an apparent waiting phase before a large aggregate suddenly appears. Unlike aggregates occurring in the initial phase, this large aggregate is stable to dissolution. Near the jump in mass, a switch occurs in the width of the largest aggregate, w_0. With a tiny fraction of exceptions, w_0 is below 3.5 before and above 3.5 after the switch point. Interestingly, as indicated in Fig. 4(a), this switch in width occurs immediately before the sharp increase in mass. This suggests that the change in width is a critical event that renders the aggregate growth competent. This finding matches perfectly with the shape of the $p(l, w)$ distribution (Fig. 3).

The kinetics can also be studied for much larger systems, which makes it possible to test in a direct manner whether or not the model captures the sigmoidal behavior observed experimentally. Figure 4(b) shows the time evolution of the mass fraction of fibril-like aggregates, $x_f(t)$, in 10
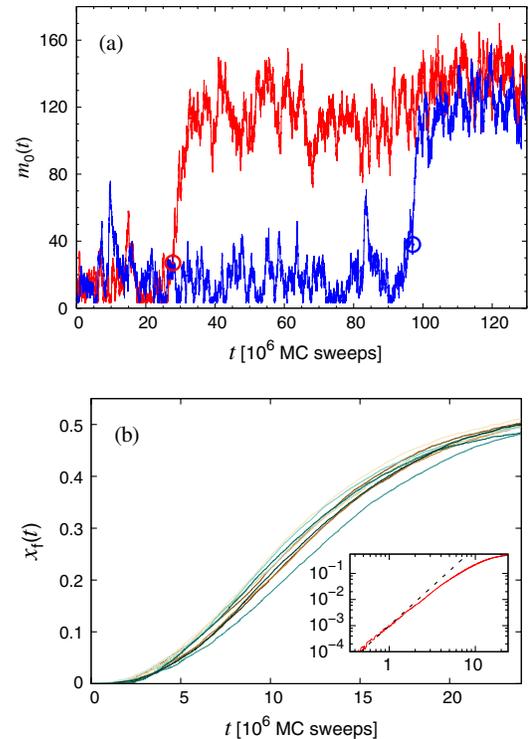
FIG. 4 (color online). Monte Carlo (MC) kinetics at $T = 0.66$. (a) Mass of the largest aggregate, $m_0(t)$, against time t, in two representative runs with $N = 256$ and $L = 64$. Circles indicate where the width of the largest aggregate, w_0, switches from $w_0 < 3.5$ to $w_0 > 3.5$. Time is given in sweeps, where one sweep consists of N single-peptide updates. (b) Mass fraction of fibril-like aggregates, $x_f(t)$, in 10 independent runs with $N = 131072$, $L = 512$. Inset: Log-log plot of the average over the 10 runs versus t. The dashed line corresponds to a cubic growth, $x_f(t) \propto t^3$.

FIG. 3 (color online). Mass fraction of aggregates with length l and width w, $p(l, w)$, at T_m for $N = 256$ and $L = 64$.

independent runs for $N = 131072$ and $L = 512$ (same concentration as before). We define an aggregate as fibril-like if $w > 3.5$, w being the width, which ensures stability to dissolution.

Comparison of these 10 runs [Fig. 4(b)] shows that, for this system size, the kinetics are indeed sigmoidal and highly reproducible, as observed in bulk experiments [3]. At the end of the runs, the simulation box typically contains between 40 and 50 spontaneously formed fibrils (see Fig. S2 in the Supplemental Material [21]), with an average length and width of $l \sim 210$ and $w \sim 7$, respectively. Inspection shows that the nucleation of new fibrils stops after roughly 10^7 Monte Carlo sweeps. Existing fibrils continue to grow beyond that point, but eventually $x_f(t)$ levels off, due to monomer depletion.

Our kinetic simulations, which do not include fragmentation events, may be compared to the classical Oosawa theory for homogeneous polymerization [23]. In particular, this theory predicts the initial growth to be quadratic in time. Our data [Fig. 4(b)] are well described by a cubic growth for small t, $x_f(t) \propto t^3$. That the exponent appears to be different than it is in the Oosowa theory is expected, because nucleation involves more than a single step in our model.

In this Letter, we have presented a simple model for amyloid formation, where the nucleation of fibrils can be studied without any prior assumptions on the structure of the aggregates involved. The formation of aggregated structures with a few β-sheet layers has been observed in many previous simulations, also at the atomic level [24,25]. Here, we have used systems much larger than in previous studies, to be able to examine the interplay between aggregate length and width in fibril nucleation. Our study shows that in this model the width of a growing aggregate plays a key role; to reach a given length, a minimum width is required, and to increase in width the aggregate has to overcome major free-energy barriers. Because of these barriers, the nucleation of a fibril occurs in distinct steps. The present study focused on the spontaneous aggregation of free peptides, but the model may also be useful for studying surface-catalyzed aggregation and the effects of a confining geometry.

As in any model, simplicity is both a strength and a limitation. The final width of a growing fibril depends most likely on geometric factors left out in our model, such as twist. The question of what sets the final width is therefore beyond the scope of the present work. The assumption that internal degrees of freedom can be integrated out may be a good approximation for small flexible peptides, but is clearly poorly justified for a folded protein that has to partially unfold before aggregation takes place. Our model further ignores any possible cooperativity of the interactions involved [26,27]. In our model, aggregation is a highly cooperative process, although driven entirely by pairwise additive interactions.

[1] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).

[2] T. P. J. Knowles and M. J. Buehler, Nat. Nanotechnol. **6**, 469 (2011).

[3] E. Hellstrand, B. Boland, D. M. Walsh, and S. Linse, ACS Chem. Neurosci. **1**, 13 (2010).

[4] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).

[5] F. A. Ferrone, J. Hofrichter, and W. Eaton, J. Mol. Biol. **183**, 611 (1985).

[6] H. Flyvbjerg, E. Jobs, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **93**, 5975 (1996).

[7] J. D. Schmit, K. Ghosh, and K. Dill, Biophys. J. **100**, 450 (2011).

[8] C. Wu and J.-E. Shea, Curr. Opin. Struct. Biol. **21**, 209 (2011).

[9] R. Pellarin, E. Guarnera, and A. Caflisch, J. Mol. Biol. **374**, 917 (2007).

[10] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, Phys. Rev. Lett. **101**, 258101 (2008).

[11] M. S. Li, D. K. Klimov, J. E. Straub, and D. Thirumalai, J. Chem. Phys. **129**, 175101 (2008).

[12] J. Zhang and M. Muthukumar, J. Chem. Phys. **130**, 035102 (2009).

[13] S. Auer and D. Kashchiev, Phys. Rev. Lett. **104**, 168105 (2010).

[14] N. S. Bieler, T. P. J. Knowles, D. Frenkel, and R. Vácha, PLoS Comput. Biol. **8**, e1002692 (2012).

[15] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).

[16] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[17] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[18] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg, Nature (London) **447**, 453 (2007).

[19] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[20] S. Æ. Jónsson, S. Mohanty, and A. Irbäck, J. Chem. Phys. **135**, 125102 (2011).

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.110.058101 for illustrations of $p(E, m)$ and a typical final configuration from the kinetic runs.

[22] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, PLoS Biol. **2**, e321 (2004).

[23] F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10 (1962).

[24] D. Li, S. Mohanty, A. Irbäck, and S. Huo, PLoS Comput. Biol. **4**, e1000238 (2008).

[25] M. Cheon, I. Chang, and C. K. Hall, Biophys. J. **101**, 2493 (2011).

[26] K. Tsemekhman, L. Goldschmidt, D. Eisenberg, and D. Baker, Protein Sci. **16**, 761 (2007).

[27] B. Linse and S. Linse, Mol. Biosyst. **7**, 2296 (2011).