

UvA-DARE (Digital Academic Repository)

Primary Nucleation Kinetics of Short Fibril-Forming Amyloidogenic Peptides

Luiken, J.A.; Bolhuis, P.G.

DOI

[10.1021/acs.jpcc.5b05799](https://doi.org/10.1021/acs.jpcc.5b05799)

Publication date

2015

Document Version

Final published version

Published in

The journal of Physical Chemistry. B

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

Citation for published version (APA):

Luiken, J. A., & Bolhuis, P. G. (2015). Primary Nucleation Kinetics of Short Fibril-Forming Amyloidogenic Peptides. *The journal of Physical Chemistry. B*, 119(39), 12568-12579. <https://doi.org/10.1021/acs.jpcc.5b05799>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

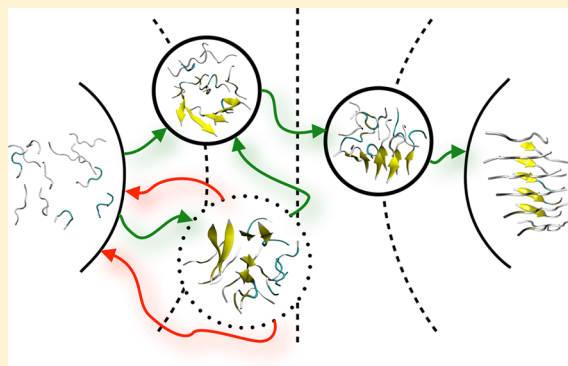
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Primary Nucleation Kinetics of Short Fibril-Forming Amyloidogenic Peptides

Jurriaan A. Luiken and Peter G. Bolhuis*

van 't Hoff Institute for Molecular Sciences, University of Amsterdam, P.O. Box 94157, 1090 GD Amsterdam, Netherlands

ABSTRACT: The primary nucleation step in amyloid fibril formation can, depending on the nature of peptide sequence, occur in one step, straight from a dilute solution, or in multiple steps, via oligomers or disordered aggregates. The precise kinetic pathways of these processes are poorly understood. Employing forward flux sampling and a midresolution coarse-grained force field, we analyzed the reactive pathways from the solvated state to the fibril nucleus for a system of 12 amyloidogenic peptides. In line with previous work, increasing the overall side-chain hydrophobicity switches the fibrillization mechanism from one- to two-step nucleation. Overall, in this mechanism, peptides first form dimers and trimers, which then grow into a β -sheet. This sheet serves as a template for nucleation of additional β sheets until the fibril nucleus is fully formed. Our simulations indicate that the presence of the recently predicted polymerized phase in the nucleation pathway of intermediately hydrophobic peptides slows down the dynamics of fibril formation considerably, which may influence the time scale on which toxic early oligomers exist. The structure of the amyloid fibrils was found to be strongly dependent on the relative hydrophobic strength of side chains along the sequence: β sheets in the fibril are oriented such that a core of the relative strongest hydrophobic residues is formed along the fibril axis.



INTRODUCTION

The self-assembly of amyloidogenic peptides into amyloid fibrils is interesting from a biological, medical, and material science viewpoint. Fibril formation occurs readily *in vivo* and *in vitro*¹ and is traditionally linked to neurodegenerative diseases such as type II diabetes, Parkinson's, Alzheimer's, and Huntington's disease.^{2–5} In the early stages of amyloid formation, proteins are known to aggregate into disordered oligomers. These early oligomers are considered to be toxic as they disrupt neuronal function.^{4,6} Amyloid fibrils, however, also play functional roles in some organisms, such as bacteria,^{7,8} and there has been increasing interest in the utilization of amyloids as biomaterials^{9–13} such as micrometer long fibrils formed from silk-based block copolymers,¹⁴ or strong fiber networks formed by proteins containing the IKVAV sequence.¹⁵

For the development of treatments against deposition diseases, or for tailoring materials with strong physical features, control of the aggregation is crucial. Such control requires a good understanding of the early steps of amyloid formation. Molecular simulation can, in principle, provide such knowledge. Straightforward molecular dynamics (MD) using all-atom force fields has been very useful in revealing the properties on the single peptide level, such as the dock-lock mechanism of amyloidogenic peptide association,^{16,17} however, as the time scales on which the various steps of amyloid nucleation itself take place range from seconds to years,¹⁸ the computational expense associated with straightforward MD simulations using all-atom force fields is prohibitively large. These long time scales are related to high free-energy barriers that are present in

the nucleation mechanism. There are two major solutions to this problem available in the literature. The first is to use enhanced sampling methods, such as replica exchange MD (REMD)¹⁹ or metadynamics²⁰ to overcome the free-energy barriers involved in fibril nucleation. As an example, Baftizadeh et al.^{21,22} have investigated the aggregation of polyvaline and amyloid- β using bias exchange metadynamics and found nonclassical nucleation behavior. REMD can be used for faster convergence toward equilibrium, but it is not easy to extract kinetics from REMD simulations. While, in principle, it would be possible to obtain insight into kinetic nucleation pathways from REMD simulations,²³ the transitions often occur at unnaturally high temperatures and may thus not correctly represent the process under ambient or physiological conditions. The use of transition path sampling^{24,25} or forward flux methodology²⁶ circumvents this problem and has, for example, enabled the study of the docking mechanism of peptides^{27–31} but has not yet been applied to larger amyloid systems.

The second solution to the time scale problem is to employ coarse-grained force fields, which contain less atomic and molecular detail but which allow much longer simulation time scales. Coarse-grained molecular simulations have provided insight into the early steps of amyloid formation.^{14,27,32–36} The midresolution model of Bereau and Deserno³⁷ provides the

Received: June 17, 2015

Revised: September 3, 2015

Published: September 4, 2015

necessary tools to simulate amyloid formation, such as directional hydrogen bond interactions, dipole interactions, and full sequence specificity; however, accessing the long time scales of amyloid fibril formation requires a combination with enhanced sampling techniques. Recently, we studied the aggregation thermodynamics of three short, amyloidogenic peptides (GNNQQNY, VEALYL, and KLVFFAE) with REMD simulations using the model by Bereau and Deserno.³³ These REMD simulations yielded insight into the structural properties of the liquid phase and the fibril phase, and lead to the prediction of a new intermediate in the nucleation process: the polymerized (or associated) peptide liquid phase. In this phase, small clusters of aligned peptides coexist that possibly prolong the existence of the toxic early oligomers. We tested the robustness of this prediction with statistical mechanical theory.³³

In this work we aim to gain insight into the kinetic pathways of the primary nucleation steps of the three amyloidogenic peptides GNNQQNY, VEALYL, and KLVFFAE. In addition, we are interested in the dynamical role of the polymerized phase in the nucleation pathway, which to our knowledge has not been previously addressed. We utilize the forward flux sampling method^{26,38,39} to investigate the nucleation dynamics of 12 amyloidogenic peptides in a dilute solution. For each of the three peptide sequences we compute the kinetics and the free energy and analyze the mechanism.

The remainder of the paper is organized as follows. In the **Methods** section we briefly introduce the forward flux sampling method, the simulation details, and the analysis techniques. In the **Results** section we present the results of the simulations for each peptide separately. We discuss and summarize the findings in the **Discussion** and **Conclusion** sections.

METHODS

Path Sampling. Amyloid fibril formation is a slow process, with time scales ranging from seconds to years. These long time scales are related to high free-energy barriers, rendering the fibril nucleation a rare event. Enhanced sampling methods can access such rare events but require the correct reaction coordinate, which is often elusive. Transition path sampling^{24,25,40–42} (TPS) was developed to overcome this problem. In TPS, trajectories are generated using a Monte Carlo sampling of trajectory space. TPS is initiated from a single trajectory connecting the initial and final states. The shooting algorithm perturbs the momenta on a given time slice along this trajectory, and new trajectories are generated by integrating the equations of motion both backward and forward in time. The transition path ensemble (TPE) consists of all generated trajectories connecting the initial and final state.

The transition interface sampling (TIS)^{43,44} method can efficiently compute the rate constants within the path sampling framework. In TIS (and also in forward flux sampling), the initial state *A* and final state *B* are separated in terms of an order parameter λ such that $\lambda < \lambda_A$ in region *A* and $\lambda > \lambda_B$ in region *B*. The boundaries of *A* and *B* are therefore defined by λ_A and λ_B (or λ_0 and λ_n , respectively). Between *A* and *B* several interfaces $i = 1, \dots, n-1$ are defined such that $\lambda_{i+1} > \lambda_i$. The TIS expression for the rate constant k_{AB} is given by

$$k_{AB} = \frac{\bar{\Phi}_A}{\bar{h}_A} P(\lambda_B | \lambda_A) \quad (1)$$

where $\bar{\Phi}_A$ is the time-average of the effective positive flux of trajectories from *A* that cross λ_A , \bar{h}_A is the time-average of the history-dependent function h_A , defined such that $h_A = 1$ if the system visited *A* more recently than *B* and 0 otherwise, and the crossing probability $P(\lambda_B | \lambda_A)$ is the conditional probability that a trajectory from *A* reaches *B* before it returns to *A*

$$P(\lambda_B | \lambda_A) = \prod_i P(\lambda_{i+1} | \lambda_i) \quad (2)$$

TIS samples the crossing probability by sampling pathways with the shooting algorithm for each interface.

Forward Flux Sampling. The forward flux sampling (FFS)^{26,38} method was originally developed with nonequilibrium dynamics in mind,³⁹ for which time reversibility did not exist; however, the method is also valid for equilibrium dynamics. FFS employs the above rate constant expressions but is different from transition path sampling (TPS and TIS) in the sense that the method does not employ Markov Chain Monte Carlo and that the equations of motion are integrated forward in time only.

The direct FFS algorithm²⁶ is initiated with a straightforward MD run in region *A*, during which the order parameter λ is followed over time. Each time the trajectory crosses boundary λ_A in the positive direction the configuration is stored until a total of N_f configurations are generated. This procedure forms a set of starting configurations on λ_A for trial runs toward the next interface (λ_1) and simultaneously provides an estimate of the flux $\bar{\Phi}_A = N_f/\tau$, where τ is the total length of the trial run. In the next phase of the algorithm several trial runs are initiated from a randomly selected starting configuration on interface *i*. If the trajectory reaches interface λ_{i+1} , the end-point is stored and the run is terminated. The trial run is also terminated if the order parameter returns to state *A*. Thus, each trajectory is free to recross previous interfaces before crossing the next. This procedure is repeated until C_{i+1} configurations are generated on interface $i+1$, and the number of runs needed to achieve this threshold is stored as M_i . The algorithm is schematically shown in **Figure 1**. The crossing probability that a trajectory from *A* crosses interface λ_i and reaches λ_{i+1} before returning to *A* is calculated as $P(\lambda_{i+1} | \lambda_i) = C_{i+1}/M_i$. This procedure is repeated until interface λ_B is reached. The conditional probability that a trajectory from *A* reaches *B* before it returns to *A* is then calculated using eq 2. From eq 1 we then directly obtain an estimate for the rate constant k_{AB} . The transition paths are

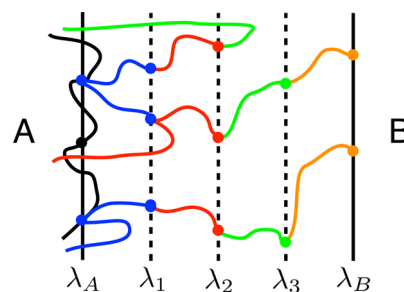


Figure 1. Schematic representation of the (direct) FFS algorithm with five interfaces. During an initial run (black line), configurations are generated on boundary *A*. From these configurations trial runs are initiated (colored lines) that either reach the next interface or return to region *A*. After generating a predefined number of configurations on each interface, the algorithm initiates trials from the next interface and so on until boundary *B* is reached.

extracted by backtracking all trajectories that reach boundary B and piecing the trial runs together.

Simulation Details. We employ a generic mid-resolution CG model developed by Bereau and Deserno.³⁷ This is an implicit solvent and sequence specific model, where each amino acid is represented by four beads: three for the backbone and one for the side-chain. The model includes interactions crucial to β -sheet stability such as back-bone hydrogen bonds and dipole interactions, and is parameterized such that both α -helical and β -sheet secondary structures are equally accessible. For a detailed description of the model we refer to ref 37. The integration was carried out by the ESPResSo⁴⁵ MD package using Langevin dynamics with a 1 fs time step. The FRESHS^{38,46,47} package (publicly available at <http://www.freshs.org/dw/doku.php>) was modified to interface with the peptide builder (publicly available on github: <https://github.com/tbereau/peptideB>), which feeds coarse-grained peptides into the MD integrator. We solvated 12 peptides in a simulation box with sides $L = 63.3$ Å, corresponding to a concentration of approximately 0.07 M. Periodic boundary conditions apply in all directions. We performed five independent forward ($A \rightarrow B$) and reverse ($B \rightarrow A$) FFS simulations for each peptide, and calculate the order parameter and save an output structure every 1 ps.

The proper choice of order parameters and the set of interfaces that separates the solvated and aggregated state requires substantial knowledge of the system, which is not always available a priori. Here we will make use of the results of REMD simulations of ref 33 using the same CG model, which indicates that the initial and final state are clearly separated in terms of the in-register contacts, N_{C_α} defined as when two C_α atoms from the same residue on different strands are within 6.5 Å of each other. The 12-peptide GNNQQNY system forms below the ordering temperature $T_o \approx 291$ K, defined by a sharp peak in the heat capacity, as obtained from the REMD simulations, a one-layered fibril consisting of around nine peptides with the remaining peptides arranged in a second layer parallel to the first. The second layer stabilizes the fibril structure, as is shown by the 20 peptide system in ref 33. The order parameter λ for GNNQQNY is defined as the number of in-register contacts in the largest cluster. For KLVFFAE the fibril state consists of two parallel aligned layers ($T_o \approx 249$ K), while for VEALYL the fibril consists of three parallel β sheets ($T_o \approx 307$ K). For these peptides we chose the order parameter λ to be the number of in-register contacts in the *two* largest clusters. For KLVFFAE and GNNQQNY, we set boundary B to $\lambda_B = 44$ in-register contacts and for VEALYL to $\lambda_B = 38$.

As previously mentioned, the FFS algorithm is initiated by performing a straightforward MD run that terminates when 150 initial configurations are generated on boundary A . Table 1 lists the threshold number of configurations C_i for each interface i required to initiate trials runs from i and the value of the order parameter at each interface. The FRESHS package imposes the condition $\lambda_{i+1} > \lambda_i$; therefore, for FFS simulations in the reversed direction ($B \rightarrow A$) we define an additional order parameter λ'

$$\lambda' = \lambda_A + \lambda_B - \lambda \quad (3)$$

Random output structures on boundary B from the forward simulations are used as starting configurations for the initial MD runs in region B .

Cluster Analysis. We obtain further insight into the aggregation process by following the formation and breaking

Table 1. Order Parameter Values λ_i and λ'_i and Number of Configurations (C_i) Required Per Interface (i) for GNNQQNY and KLVFFAE^a

i	C_i	λ_i	λ'_i
0	150	5 (A)	5 (B)
1	120	8	8
2	90	11	11
3	80	14	14
4	75	17	17
5	70	20	20
6	65	23	23
7	60	26	26
8	58	29	29
9	56	32	32
10	54	35	35
11	52	38	38
12	51	41	41
13	50	44 (B)	44 (A)

^aFor VEALYL the boundary of B is set to 38 in-register contacts (interface 11).

of correctly formed peptide clusters of size n , by performing a cluster analysis on the frames along each reactive trajectory. Two peptides belong to the same cluster if they form at least four in-register contacts with each other. The algorithm is initiated by looping over all clusters in the first frame of a trajectory and computing the size n of each cluster. In the next frame we determine whether each cluster has grown, remained constant, or shrunk, and we increment the corresponding histogram $h_{\text{grow}}(n)$, $h_{\text{same}}(n)$, or $h_{\text{shrink}}(n)$. This procedure is repeated until the end of the trajectory is reached. The probability that a cluster of size n grows, $P_{\text{grow}}(n)$, is then calculated using

$$P_{\text{grow}}(n) = \frac{h_{\text{grow}}(n)}{h_{\text{grow}}(n) + h_{\text{same}}(n) + h_{\text{shrink}}(n)} \quad (4)$$

$P_{\text{same}}(n)$ and $P_{\text{shrink}}(n)$ are calculated in a similar manner. In some cases it is more insightful to calculate the growth probability relative to the shrinking probability

$$P_{\text{grow}}(n) = \frac{h_{\text{grow}}(n)}{h_{\text{grow}}(n) + h_{\text{shrink}}(n)} \quad (5)$$

Although, in principle, it is possible to extract information on the critical nucleus size, n^* , from this analysis, as $P_{\text{shrink}} > P_{\text{grow}}$ for $n < n^*$ while $P_{\text{shrink}} < P_{\text{grow}}$ for $n > n^*$,⁴⁸ we analyze only the reactive pathways and therefore only obtain insight into the nucleation mechanism rather than the size of the critical nucleus. We note that here we are referring to the classical view of nucleation, in which the growth of a nucleus is assumed to be strictly one-, two-, or three-dimensional and the critical nucleus size is well-defined. Recent studies have shown, however, that the concept of a critical nucleus size breaks down due to the multidimensional nature of fibril growth.^{49,50} The emphasis of the cluster analysis is therefore on the evolution of clusters along the trajectory and to aid in the interpretation of the reactive pathways, rather than on the size of the critical nucleus.

The Committor. The committor $P_B(x)$ is the probability that a trajectory initiated from configuration x reaches boundary B before it reaches boundary A . This probability increases from 0 to 1 as the reaction progresses from the initial to final state; therefore, it is, in principle, the ideal reaction

coordinate and, as such, is useful to measure the correlation between the chosen order parameter and the true reaction coordinate.^{39,51,52} A committor value of ~ 0.5 has special significance, as here the configuration has an equal probability of reaching the initial or final state. These configurations comprise the transition state ensemble (TSE) and can be analyzed to obtain information on the reaction mechanism.

The procedure to calculate the committor for direct FFS is similar to that for branched growth FFS,⁵³ with the main difference that the number of trials fired varies per configuration rather than per interface. For each configuration j on interface λ_i we estimate the corresponding committor P_{Bj}^i from the sum of the committors of the configurations it connects to on the next interface, P_{Bj}^{i+1} , divided by the number of trials fired from this configuration, k_j

$$P_{Bj}^i = \frac{\sum_{m=1}^{C_{i+1}} P_{Bm}^{i+1}}{k_j} \quad (6)$$

We consider all configurations for which $0.35 < P_B(x) < 0.65$ to be part of the TSE. A good indication for a strong correlation of the order parameter with the true reaction coordinate is a highly peaked distribution of the order parameter in the TSE.^{25,42}

Free-Energy Analysis. The free-energy F is directly related to the distribution function $\rho(N_{C_a})$ through

$$F(N_{C_a}) = -k_B T \ln \rho(N_{C_a}) \quad (7)$$

In the case of a stable state B , $\rho(N_{C_a})$ is extracted from FFS simulations by weighting the average time all trial runs from the forward ($A \rightarrow B$) and reverse simulations ($B \rightarrow A$) spent at order parameter N_{C_a} .^{39,54}

$$\rho(N_{C_a}) = \rho_A \Phi_{A,0} \tau_+(N_{C_a}; \lambda_A) + \rho_B \Phi_{B,n} \tau_-(N_{C_a}; \lambda_B) \quad (8)$$

where ρ_A is the probability of finding the system in region A at steady state, $\Phi_{A,0}$ is the flux of trajectories from A that reaches interface λ_0 , and $\tau_+(N_{C_a}; \lambda_A)$ is the average time a trajectory originating from A has spent at order parameter N_{C_a} . Similarly, ρ_B is the steady-state probability of finding the system in region B , $\Phi_{B,n}$ is the flux of trajectories from B that reaches interface λ_n , and $\tau_-(N_{C_a}; \lambda_B)$ is the average time a trajectory originating from B has spent at order parameter N_{C_a} .

We extract the probabilities ρ_A and ρ_B from the rate constants using the relation

$$\rho_A k_{AB} = \rho_B k_{BA} \quad (9)$$

and the fact that $\rho_A + \rho_B \approx 1$ at steady state. The fluxes $\Phi_{A,0} = N_f/\tau$ and $\Phi_{B,0} = N_r/\tau$ are calculated during the initial straightforward MD simulation in the corresponding regions, where τ is the total simulation time of the initial run. The function $\tau_+(N_{C_a}; \lambda_A)$ is calculated using data from the forward simulations³⁹

$$\tau_+(N_{C_a}; \lambda_A) = \pi_+(N_{C_a}; \lambda_A) + \sum_{i=1}^{n-1} \pi_+(N_{C_a}; \lambda_i) \prod_{j=0}^{i-1} P(\lambda_{j+1} | \lambda_j) \quad (10)$$

where $\pi_+(N_{C_a}; \lambda_i)$ is the average time a run originating from interface λ_i spent with order parameter N_{C_a} . The function

$\tau_-(N_{C_a}; \lambda_A)$ is calculated in a similar fashion with data from the reverse simulations.

RESULTS

GNNQQNY. In ref 33, we inferred the nucleation behavior of the weakly hydrophobic (hydrophilic) GNNQQNY peptide on the basis of the aggregation thermodynamics. Using Wertheim's first-order perturbation theory and statistical associating fluid theory we showed that polymerization in the gas phase is possible for weakly hydrophobic peptides; however, in the REMD data, we did not find evidence of a full polymerization transition for the 12-peptide GNNQQNY system.

The polymerization transition is signified by an inflection point in the aggregation fraction Θ as a function of temperature. (See Figure 2.) Although clustering is possible

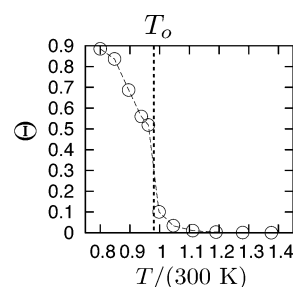


Figure 2. Plot of the aggregation fraction versus temperature.

in the gas phase ($\Theta > 0$ for $T > T_o$), there is no inflection point for $T > T_o$. We therefore hypothesize that a critical fibril nucleus forms during the polymerization process, thus allowing the system to crystallize before it can fully polymerize. On the basis of the thermodynamic behavior of GNNQQNY, we argued that slightly below the ordering temperature peptides nucleate in one step directly from the dilute solution.

Here we study the mechanism of fibril formation of GNNQQNY by performing FFS simulations. We set the simulation temperature to 285 K, slightly below T_o , and define boundaries $\lambda_A = 5$ and $\lambda_B = 44$ with $\Delta\lambda = 3$. In the first phase of FFS, a straightforward MD simulation, initiated from a random solvated state, is run until it crosses boundary A in the positive direction 150 times. (See Table 1.) In the second phase, paths are generated from these crossing points, for each interface, saving each crossing point of the next interface. As $C_B = 50$, we generate eventually 50 paths in the TPE. Because some of these paths are likely to be correlated, we perform five independent FFS simulations and combine the data. To calculate the steady-state probabilities ρ_A and ρ_B , the rate constant k_{BA} , and the free-energy curve, we initiate five reverse FFS simulations from boundary B using random output structures on boundary B from the forward simulations.

We computed the committor values for the forward transition $A \rightarrow B$ and plotted the distribution of the in-register contacts evaluated over the TSE in Figure 3a. Note that the probability of finding a transition state is peaked around 38 contacts, near boundary $\lambda_B = 44$. This suggests that the addition of peptides to large aggregates, consisting of ~ 38 contacts or more, is the rate-limiting step, and that state B , in fact, might be metastable. Indeed, the computed rate constants (see Table 2) are $k_{AB} = 4.79 \times 10^2 \text{ s}^{-1}$ and $k_{BA} = 7.61 \times 10^4 \text{ s}^{-1}$; that is, the reverse reaction proceeds significantly faster than the forward

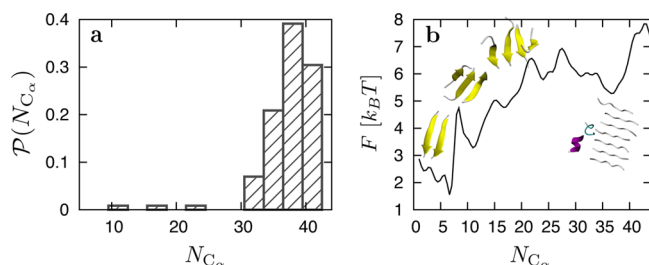


Figure 3. (a) Distribution of the order parameter of the TSE configurations for the $A \rightarrow B$ transition in the 12-peptide GNNQQNY system. (b) Free energy versus in-register contacts in the largest cluster.

Table 2. Crossing Probabilities Per Interface, the Flux (s^{-1}), and the Rate Constants (s^{-1}) for the Forward and Reverse FFS Simulations of All Peptides

i	GNNQQNY		VEALYL		KLVFFAE	
	$P(\lambda_{i+1} \lambda_i)$	$P(\lambda'_{i+1} \lambda'_i)$	$P(\lambda_{i+1} \lambda_i)$	$P(\lambda'_{i+1} \lambda'_i)$	$P(\lambda_{i+1} \lambda_i)$	$P(\lambda'_{i+1} \lambda'_i)$
1	0.1182	0.1872	0.2722	0.7057	0.1118	0.3414
2	0.4050	0.2376	0.4119	0.3058	0.2774	0.7531
3	0.3405	0.3550	0.3694	0.3220	0.2635	0.8603
4	0.2913	0.6084	0.2722	0.1831	0.3558	0.9558
5	0.2996	0.6805	0.4428	0.3696	0.2943	0.8712
6	0.2357	0.7501	0.4048	0.3844	0.4438	0.7073
7	0.3776	0.8401	0.5606	0.3483	0.3430	0.6093
8	0.3031	0.8374	0.5142	0.3549	0.2677	0.2092
9	0.6474	0.8282	0.5456	0.5779	0.2831	0.1179
10	0.6506	0.8930	0.5864	0.4793	0.4604	0.2043
11	0.7702	0.8734	0.5664	0.5128	0.4107	0.3527
12	0.8748	0.8151			0.4494	0.3150
Φ	2.89×10^7	6.31×10^7	1.55×10^7	2.59×10^7	8.84×10^7	3.21×10^6
k	4.79×10^2	7.61×10^4	1.39×10^3	90.1	11.7	68.4

reaction. The ratio of the rate constants $k_{AB}/k_{BA} = 0.006$ is equal to the equilibrium constant for the reaction. Using the well-known relation for the free-energy difference $\Delta F_{AB} = F_B - F_A = -k_B T \ln \frac{k_{AB}}{k_{BA}} = 5.07 k_B T$. The crossing probabilities, flux, and rate constants are listed in Table 2. The steady-state probability of finding the system in state A , given by

$$\rho_A = \frac{k_{BA}}{k_{AB} + k_{BA}} \approx 0.994 \quad (11)$$

and the probability of finding the system in state B , $\rho_B = 1 - \rho_A \approx 0.006$, reflect these results; at 285 K, large β -sheets of GNNQQNY peptides are at best metastable with respect to the dissolved A state.

The results of the free-energy analysis are shown in Figure 3b. We can identify several peaks belonging to the subsequent addition of peptides to the largest cluster. Clusters with size $n = 3$ contain at least eight contacts, and those with size $n = 4$ at least 12 contacts; therefore, the first and second peaks are likely due to the formation of trimers and tetramers. (See the insets in Figure 3b for simulation snapshots.) The top of the barrier is ~ 27 contacts, which corresponds to clusters consisting of four to six peptides. Such clusters are likely to act as a critical nucleus. The right-most metastable minimum around ~ 37 contacts is due to clusters of size $n = 7, 8$, and 9. Beyond this

point, β sheets become increasingly less stable. The metastable minimum is around 4 to $5k_B T$ above the minimum for state A , roughly in agreement with the free-energy difference based on the rate constants. The fact that this metastable minimum is not corresponding to the definition of state B ($\lambda \geq 44$) reveals the limitation of N_{C_a} to describe this state. Moreover, the location of the maximum in the free energy ($\lambda \approx 25$) does not agree with the distribution of the order parameter in the TSE in Figure 3a. Also, the height of the free-energy barrier (about $5k_B T$ and $2k_B T$ for the forward and backward transitions, respectively) is much lower than the rate constants suggest. This is indicative of an overlap of state B with a large part of the barrier in terms of the order parameter N_{C_a} .

By performing a time series cluster analysis of the TPE (see Methods section), we illustrate the decline in growth probability for these clusters. Figure 4a shows the growth

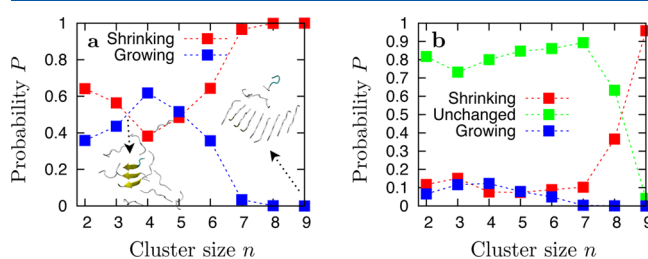


Figure 4. (a) Growth and shrinking of GNNQQNY clusters along the TPE. (b) Growing, shrinking, and stability of clusters.

probability relative to the shrinking probability. In the reactive nucleation paths it is likely that small peptide clusters $n < 4$ shrink on average, that is, $P_{\text{shrink}} > P_{\text{grow}}$. Clusters consisting of four or five peptides are most likely to grow, but the growth probability steeply declines for $n > 5$. This suggests that nuclei of four to five peptides are characteristic for a barrier crossing, but as was pointed out in the Methods section, they do not necessarily correspond to the critical nucleus because we take only reactive pathways into account. Indeed, this characteristic nucleus size does not correspond to the maximum in the free-energy barrier in Figure 3b. We included the fraction of unchanged clusters between snapshots in the time series cluster analysis in Figure 4b. Here we indeed observe a steep decline in the unchanged configurations for $n > 7$; larger clusters are thus more likely to shrink than to remain stable in the absence of free monomers. (We note that increasing the time between frames would decrease the unchanged fraction for all n .) We can attribute the decrease in stability to one or more of the following reasons: (1) The REMD data in Figure 2 indicate that at $T = 285$ K some peptides have begun assembling into the fibril; however, the system is not yet fully aggregated ($\Theta \approx 0.55$), and thus large β sheets ($n > 5$) may not be thermodynamically stable at this temperature; (2) the FFS algorithm drives the formation of a single large cluster rather than the formation of two layers, which was shown to be the thermodynamically most stable state at temperatures $T < T_0$ in ref 33. Therefore, state B as we defined it may not correctly represent the equilibrium structure of the fibril. Finally, (3) it may be due to a finite-size effect: The free monomer concentration declines as the cluster size increases. The time scale τ_{grow} on which a monomer docks to the fibril is inversely proportional to the monomer concentration, $\tau_{\text{grow}} \propto L^3/(N - n)$, and thus increases with increasing n . The time scale τ_{shrink}

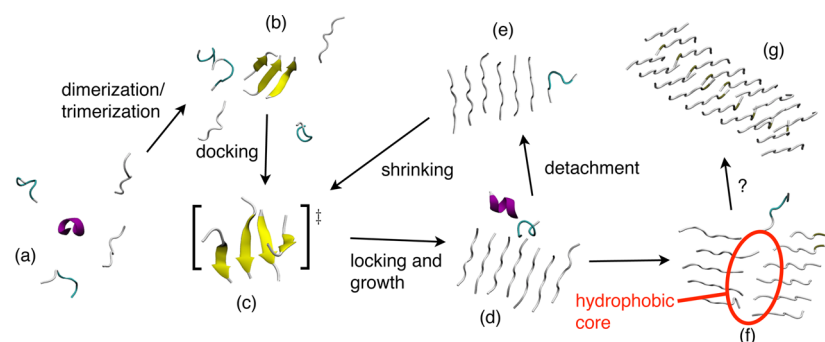


Figure 5. Schematic of the nucleation pathway of GNNQQNY as determined from FFS simulations.

on which peptides at the fibril ends detach, however, is independent of the monomer concentration; therefore, for large n the rate of detachment may exceed the rate of attachment ($\tau_{\text{grow}} > \tau_{\text{shrink}}$), and large β sheets may become more likely to shrink than grow in a given time frame.

By analyzing the reactive pathways of GNNQQNY we reconstructed the early aggregation mechanism and visualized it in Figure 5a–c. The initial state consists of random-coil peptides in solution (a). Then, the peptides form dimers and trimers via a docking mechanism²⁷ (b) until the critical nucleus size is reached (c). The local environment of aligned peptides acts as a template that helps peptides lock firmly into place, and the resulting β sheet grows through peptide addition at the ends of the sheet (d). This self-templated nucleation effect has been previously reported in ref 34 for nucleation inside oligomer droplets. The formation of the critical nucleus thus proceeds directly from solution, the hallmark of a one-step nucleation mechanism (1SN). For large clusters peptides detach more frequently due to the instability of large sheets or due to monomer depletion (e). Nevertheless, clusters are unlikely to become smaller than the critical nucleus size. The remaining free monomers form a second β -sheet that interacts laterally and aligns in-plane with the first sheet (f). In-plane aligned β -sheets of GNNQQNY peptides were observed in previous work,³⁷ where the driving force was found to be the attraction between C-terminal tyrosines. From Figure 5f we can infer that rather than growing a cluster larger than seven peptides the additional peptides prefer to form a second layer. This second layer provides additional stability through lateral hydrophobic interactions between tyrosine residues, thereby forming a hydrophobic core along the fibril axis. This structure may well be a transition state in the formation of the two-layered fibril (g). To test this hypothesis we performed additional FFS simulations using the number of in-register contacts in the two largest clusters (data not shown); however, here two independent β -sheets formed in solution rather than a two-layered fibril.

VEALYL. The VEALYL peptide is the most hydrophobic of the three peptides, and has the strongest overall side-chain interactions³⁷ compared with GNNQQNY and KLVFFAE. These interactions drive the formation of a dense liquid phase prior to forming fibrillar structures. We set the simulation temperature to 300 K, slightly below the ordering temperature of $T_o \approx 307$ K determined from the REMD simulations. The boundaries of the disordered droplet state A and the fibrillar state B are set to $\lambda_A = 5$ and $\lambda_B = 38$, respectively, with a spacing of $\Delta\lambda = 3$ between interfaces.

During the initial run in region A we found that peptides quickly coalesce into a dense disordered liquid of peptides,

driven by the strong hydrophobic interactions between the side chains, which suggests a two-step nucleation (2SN) mechanism. The distribution of the order parameter in the TSE of the $A \rightarrow B$ transition is plotted in Figure 6a and is peaked around

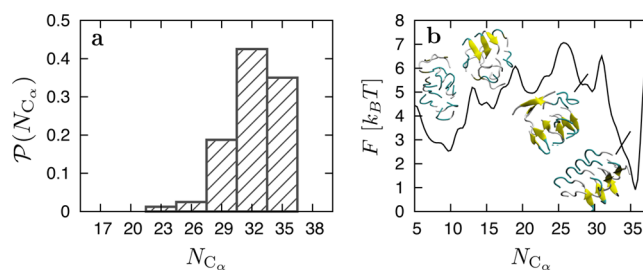


Figure 6. (a) Distribution of the order parameter of the TSE configurations for the $A \rightarrow B$ transition in the 12-peptide VEALYL system. (b) Free-energy versus in-register contacts in the two largest clusters. Simulation snapshots are shown as insets.

32 contacts. The fibril state of the 12-peptide VEALYL system is more stable than the disordered state. The forward and reverse reaction rate constants are $k_{AB} = 1.39 \times 10^3 \text{ s}^{-1}$ and $k_{BA} = 90.1 \text{ s}^{-1}$, respectively. Table 2 lists the crossing probabilities, flux, and rate constants. These rate constants translate into steady-state probabilities of $\rho_A = 0.0608$ and $\rho_B = 0.939$. The free energy of fibril formation $\Delta F_{AB} = -k_B T \ln \frac{k_{AB}}{k_{BA}} = -2.73 k_B T$ is indeed negative.

In Figure 6b we plotted the free energy as a function of the number of in-register contacts in the two largest clusters. The minimum at six to nine contacts corresponds to a disordered oligomeric state containing mostly monomers and one dimer, two dimers, or one trimer. The two peaks following this minimum correspond to the formation of a small β sheet consisting of four peptides. This β sheet forms a hydrophobic surface in the oligomer, which serves as a template for monomers to align and form the second β -sheet flat on the first sheet. This process is known as self-templated nucleation.³⁴ Subsequently, the remaining monomers align and a stable three-layered fibril is formed. The minimum at $N_{C_\alpha} = 35$ is roughly $2k_B T$ lower in free energy than the dissolved state, in agreement with the rate constant estimates. The structure of the fibril corresponds well with the thermodynamically most stable structure found from REMD simulations and consists of three sheets hooked into each other through their strongly hydrophobic C-terminal leucine and tyrosine residues. Several representative simulation snapshots are shown in Figure 6b. As in the case of GNNQQNY, the location of the maximum in the

free energy ($\lambda \approx 25$) does not agree with the distribution of the order parameter in the TSE in Figure 6a, and free-energy barrier (about $4 k_B T$ and $6 k_B T$ for the forward and backward transition, respectively) is much lower than the rate constants indicate. Again, this suggests a limitation of order parameter N_{Ca} .

A time series cluster analysis of all reactive pathways reveals that the growth probability of dimers and trimers exceeds the shrink probability (Figure 7a). The growth probability abruptly

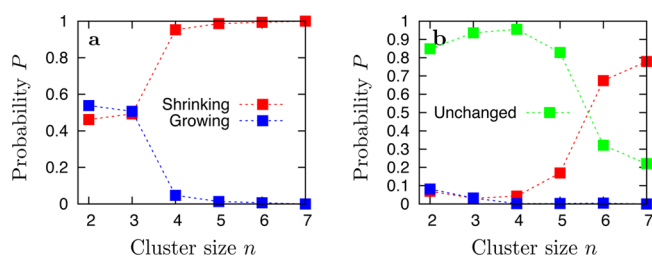


Figure 7. (a) Growth and shrinking of VEALYL clusters along the TPE. (b) Growing, shrinking, and stability of clusters.

decays to zero for $n \geq 4$, which confirms that in the reactive pathways the peptides form a second sheet rather than attaching to the first sheet. The fraction of clusters that remain of constant size (green line in Figure 7b) also indicates that the stability of larger clusters steeply declines with $n > 4$. We can infer that (1) the overall hydrophobicity of the chains drives a hydrophobic collapse into a disordered oligomer prior to forming the fibril (2SN) and (2) rather than forming a flat three-layered fibril, the β -sheets are oriented such that a hydrophobic core of leucine and tyrosine residues forms along the fibril axis; therefore, the structural properties of the fibril are determined by the relative side-chain hydrophobicity along the peptide chain.

On the basis of inspection of the reactive pathways we visualized the aggregation mechanism in Figure 8. First, freely soluble peptides (a) rapidly coalesce into a disordered oligomer (b). Inside the oligomer, peptides form persisting bonds through their strongly hydrophobic C-terminal residues and form a critical nucleus. The nucleus grows until a single β -sheet consisting of four peptides is formed (c). The formation of a second sheet proceeds more rapidly as the first sheet serves as a template for nearby peptides.³⁴ Because of this self-templated nucleation effect the fibril forms in a highly cooperative manner. The second sheet lies on top of the first sheet and is aligned parallel to maximize the hydrophobic interactions between the C-terminal residues (d). In the last step, the

remaining peptides align to form a third sheet (d) and the resulting fibril undergoes a structural change to maximize the hydrophobic interactions. The fibrillar structure consists of three sheets hooked into each other to form a hydrophobic core along the fibril axis. Subsequent growth would take place along the fibril axis. Indeed, the REMD simulations in ref 33 for 20 peptides show more elongated structures.

KLVFFAE. KLVFFAE peptides have an overall hydrophobic strength in between that of VEALYL and GNNQQNY. We initially performed FFS simulations of KLVFFAE peptides at 240 K, slightly below the ordering temperature $T_o \approx 249$ K determined from REMD simulations, using various order parameters (in-register contacts, total contacts, in-register contacts in largest cluster(s), and the nematic order parameter P_2 ^{55,56}) to drive the formation of the thermodynamically most stable structure at this temperature, the two-layered fibril. We found that, similar to VEALYL, the peptides quickly collapse into a liquid disordered phase, driven by strong side-chain interactions. Subsequently, the peptides polymerize to form a liquid of randomly oriented dimers, trimers, and so on: a polymerized or associated liquid.³³ This polymerized liquid appears to be almost frozen as the small clusters are kinetically arrested. In the thermodynamically preferred pathway these clusters must either dissolve and reassemble in the correct way or diffuse and align to form the fibril; however, the extremely slow intrinsic dynamics of the polymerized liquid at 240 K would render such a pathway too long for path simulations. At this temperature, the polymerized phase is essentially a kinetic trap and the system is not able to escape within a reasonable (simulation) time. To force the system to follow an accessible kinetic pathway of fibril formation, we therefore raised the simulation temperature to 285 K. This reduced the degree of polymerization and led to faster dynamics in the oligomer droplet (Figure 9). Because the fibrillar aggregate consists of two parallel β -sheets, we set the in-register contacts in the two largest clusters as the order parameter. The initial and final state are defined by boundaries $\lambda_A = 5$ and $\lambda_B = 44$, respectively, with a spacing of $\Delta\lambda = 3$. Using the number of in-register contacts in the two largest clusters as the order parameter, the system was able to reach the final interface consistently.

From an analysis of the committor for the $A \rightarrow B$ transition we again obtain the distribution of the order parameter in the TSE, shown in Figure 10a. As expected, the TSE structures are located on the interfaces near boundary B. Table 2 lists the computed crossing probabilities, flux, and rate constants. The reverse reaction rate $k_{BA} = 68.4 \text{ s}^{-1}$ is now approximately 6 times faster than the forward reaction rate $k_{AB} = 11.7 \text{ s}^{-1}$, and the corresponding steady-state probabilities are $\rho_A = 0.854$ and



Figure 8. Schematic of the nucleation pathway of VEALYL peptides.

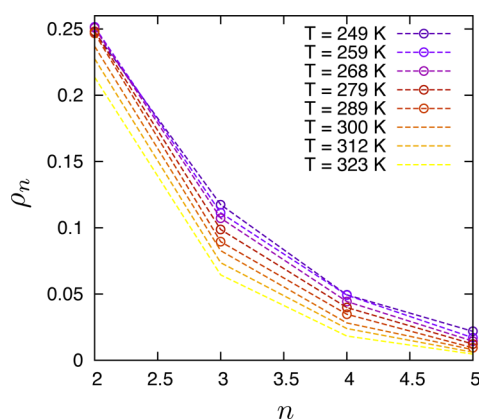


Figure 9. Cluster size distribution at various temperatures obtained from the REMD simulations in ref 33. The occurrence of small clusters decreases with increasing temperature.

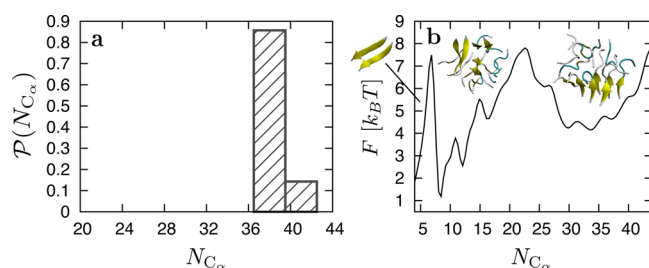


Figure 10. (a) Distribution of the order parameter of the TSE configurations for $A \rightarrow B$ transition of the 12-peptide KLVFFAE system. (b) Free energy versus in-register contacts in the two largest clusters.

$\rho_B = 0.146$. The free energy for fibril formation is $\Delta F_{AB} = -k_B T \ln \frac{k_{AB}}{k_{BA}} = +1.77 k_B T$, indicating that the two-layered fibril is indeed thermodynamically metastable at 285 K.

Figure 10b shows the calculated free energy as a function of the number of in-register contacts in the two largest clusters. We can identify various free-energy minima. The left-most minimum at $N_C = 4$ contacts corresponds to a disordered oligomer droplet consisting of one dimer and 10 monomers. A large peak separating the first and second minimum is likely to correspond to the formation of the critical nucleus of three aligned peptides, in line with results from higher resolution MD simulations in previous work.⁵⁷ Two possible oligomer compositions contribute to the minimum at $N_C = 8$: a droplet comprising a trimer and nine monomers or containing at least two dimers. The minimum at $N_C = 12$ is due to oligomers containing either one tetramer and eight monomers or a trimer and at least one dimer. Oligomer structures corresponding to the minimum at 16 contacts consist of either (1) one pentamer and seven monomers, (2) a tetramer and at least one dimer, or (3) at least two trimers. The free-energy minima at $N_C = 8, 12$, and 16 contacts therefore correspond to different configurations in the polymerized phase. The right-most minimum contains contributions from oligomer configurations with $29 \lesssim N_{C\alpha} \lesssim 37$ contacts in the two largest clusters. These highly ordered configurations generally consist of at least two aligned tetramers (see Figure 10b, inset) and are therefore close in structure to the fibril state. The minimum is $\sim 2k_B T$ above the free energy of the disordered state, consistent with the rate constant estimates. The free-energy maximum at $\lambda \approx 20$

disagrees with the distribution of the order parameter in the TSE in Figure 6a, and the free-energy barrier (about $6 k_B T$ and $4 k_B T$ for the forward and backward transitions, respectively) is again much lower than the rate constants indicate, which is deemed a limitation of the choice of the order parameter $N_{C\alpha}$.

Figure 11 shows the fraction of growing and shrinking clusters as a function of the cluster size in the TPE. The growth

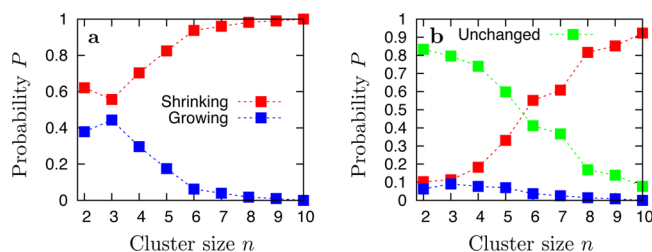


Figure 11. (a) Growth and shrinking of KLVFFAE clusters along the TPE. (b) Growing, shrinking, and stability of clusters.

probability of the trimer is (only slightly) lower than that of the dimer (Figure 11a), suggesting that the trimer indeed acts as the critical nucleus ($n^* = 3$). Again, we stress that only the reactive paths are taken into account here. We can attribute the fact that the shrinking probability is always greater than the growth probability to the elevated simulation temperature. The stability of clusters (Figure 11b, green line) further decays for $n > 4$ and plateaus around $n = 8$, indicating that these cluster sizes are not very stable.

Inspection of the reactive pathways in the TPE led to a visualization of the various steps in the aggregation mechanism in Figure 12. First, hydrophobic side-chain interactions drive the formation of a disordered liquid droplet (b) from solvated peptides (a). Polymerization of peptides in the droplet leads to the formation of a polymerized peptide liquid (c). The small clusters in this polymerized phase must either dissolve and reassemble or diffuse and align to form a single large β -sheet consisting of four or peptide peptides (d). Because of the slow intrinsic dynamics of the polymerized liquid, both of these processes are slow. Subsequently, peptides align with the first sheet to form a second sheet, with the remaining peptides still dissolved in the droplet (e). This conformation maximizes the contact surface between the strongly hydrophobic phenylalanine residues at the centers of the chains, thereby forming a hydrophobic core along the fibril axis. The remaining peptides align to form a flat two- or three-layered fibril (f), with layers sliding and rotating relative to each other (g), which is likely due to the elevated simulation temperature.

DISCUSSION

We performed forward flux sampling simulations of the fibril-forming peptide sequences GNNQQNY, KLVFFAE, and VEALYL using the midresolution coarse-grained model of Berau and Deserno.³⁷ In agreement with previous work, we found that the fibril nucleation behavior is indeed in large part determined by the overall side-chain hydrophobicity: the nucleation mechanism switches from one-step (1SN) to two-step nucleation (2SN) with increasing hydrophobic strength. The details of the mechanism and the fibrillar structure depend on the relative side-chain hydrophobicity along the peptide sequence. Previous work by Auer et al. has shown that the nucleation mechanism for a fixed sequence is in first instance determined by the temperature and peptide concentration.⁵⁸

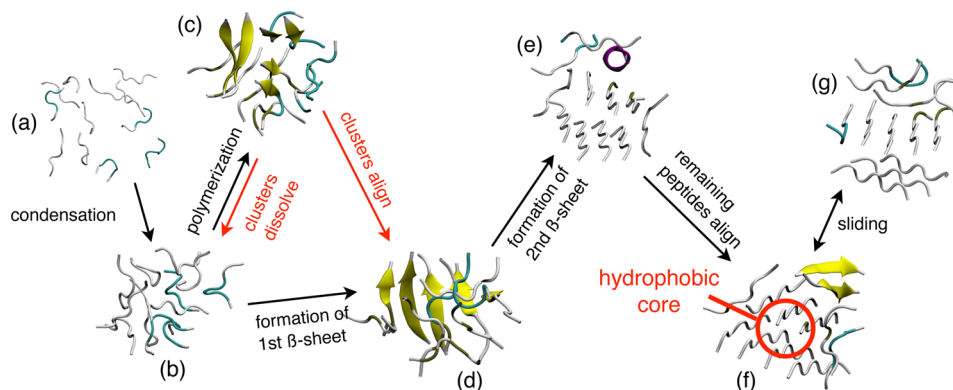


Figure 12. Schematic of the nucleation pathway of KLVFFAE. The polymerized phase (c) is an off-pathway intermediate that requires a slow alignment or dissolution of the clusters to grow into an ordered fibril.

These authors find that 1SN and 2SN can both occur depending on the temperature and concentration. In contrast, our simulations have been performed at a fixed temperature and concentration, and we instead varied the side-chain hydrophobicity by simulating various sequences. The qualitative equilibrium phase behavior will change by altering the sequence.³³ In addition, the dominant nucleation mechanisms may change upon varying the peptide concentration and system temperature.

For the weakly hydrophobic GNNQQNY peptide, where fibril formation is dominated by hydrogen-bond formation rather than hydrophobic interactions, the tyrosine residues drive the docking of peptides directly from solution (one-step nucleation) until a four peptide β -sheet nucleus is formed, after which the sheet is more likely to grow than shrink; however, we hardly observe clusters with more than seven peptides due to monomer depletion or the low stability of single-layer sheets. Indeed we find that the formation of the two-layered fibril is often preceded by the formation of small β -sheets that align in-plane with the tyrosine residues, forming a hydrophobic core along the fibril axis. The GNNQQNY results are in good agreement with previous experimental work⁵⁹ and simulations using atomistic and other CG models,^{60–62} where amyloid formation was found to proceed in a single step from solution (1SN) with a critical nucleus of four to six peptides. Our results for the GNNQQNY peptide should be qualitatively similar to the phase diagram in ref 58, and indeed we find 1SN to be dominant. We expect that at (much) higher concentration or lower temperature also a 2SN mechanism can be competitive.

The VEALYL peptides form fibrils via a two-step nucleation mechanism (2SN) in which hydrophobic interactions drive the formation of a disordered oligomer droplet. Inside this oligomer, peptides align to form a nucleus: a dimer or a trimer, which grows into the first four-peptide β -sheet. This sheet serves as a hydrophobic surface for the peptides still dissolved in the droplet, and a second and third β -sheet forms parallel to the first via self-templated nucleation. The fibril undergoes a structural change to maximize hydrophobic interactions between the C-terminal leucine and tyrosine residues, resulting into three β -sheets hooked into each other through the C-terminal residues, which form a hydrophobic core along the fibril axis and allows further growth.

Solid-state NMR spectroscopy in combination with atomistic MD simulations in recent work⁶³ reveal that VEALYL peptides arrange in antiparallel β -sheets stacked parallel in a steric zipper structure. These authors have identified several aggregated

states: disordered oligomeric, microcrystalline and fibrillar. We have identified the oligomeric intermediate in our simulations of the VEALYL peptide, but due to the lack of side-chain detail and electrostatic interactions on glutamic acid in the CG model we did not observe antiparallel β -sheets or steric zipper interfaces. However, earlier work on the related peptide LVEALYL revealed parallel β -sheets,⁶⁴ showing that the stability of different morphologies can be rather subtle, and can be influenced by e.g. the protonation state of the residues.

The kinetic pathway of KLVFFAE fibril formation was accessible only for a temperature above the ordering temperature due to the slow dynamics in the low-temperature polymerized liquid oligomer. The reaction scheme of KLVFFAE suggests two possible pathways. At the higher temperature fibrils form via a 2SN route in which peptides first condense into a disordered liquid, followed by the formation of a three peptide β -sheet nucleus, a full first β -sheet, and finally cooperative growth of a second sheet onto the template of the first. In the low-temperature pathway (below T_o), peptides aggregate along a condensation, polymerization, and ordering mechanism (2SN + polymerization). After coalescence into disordered oligomer droplets KLVFFAE peptides form a polymerized phase consisting of several slowly diffusing peptide clusters.³³ Any subsequent conversion of the polymerized droplet into a fibril is very slow as it requires the clusters to diffuse and align or dissolve and reassemble to form the first β -sheet.

The fibrillar structure of the amyloid- β_{16-22} (KLVFFAE) segment has been studied using solid-state NMR spectroscopy⁶⁵ and MC simulations⁶⁶ and was found to be mostly in antiparallel arrangement, whereas here we exclusively find parallel β -sheets. We have attributed this to the lack of electrostatic interactions at the N- and C-terminal residues of KLVFFAE in the model by Bereau and Deserno³⁷ in previous work.³³ We would like to stress that the major prediction of our work is not the precise orientation of the peptides but rather the variety of primary nucleation routes that peptides can take depending on the sequence, and elucidation of the nucleation pathway involving the kinetically trapped intermediate state. The two-step nucleation mechanism of KLVFFAE was also clearly observed in experimental work employing fluorescence imaging,⁶⁷ where a hydrophobic collapse drives the formation of densely packed aggregates: a state critical for amyloid nucleation.⁶⁸ In addition, the size of the critical nucleus corresponds with previous simulation results;⁵⁷ however, we must be careful to draw such a conclusion because here the

nuclei growth analysis was performed on the reactive pathways only.

The FFS results presented in this work in combination with our previous REMD simulations indicate that an increase in the relative hydrophobic to hydrogen bond strength switches the nucleation pathway from 1SN to 2SN. The relative side-chain hydrophobicity along the peptide chain (i.e., the sequence) plays an important role in the details of the reaction mechanism and the structural properties of the fibril: There is a strong tendency of all peptides to form a hydrophobic core along the fibril axis. The side-chain hydrophobicity is what differentiates peptides, whereas the hydrogen bond contribution can be considered sequence-independent.

The main novel prediction of this work concerns the additional polymerization step that can occur in the 2SN mechanism for intermediate hydrophobicity. Hydrogen bond formation not only stabilizes the final fibrillar assemblies but also induces polymerization in disordered oligomers.³³ Our FFS simulations of the low-temperature pathway for KLVFFAE fibril formation show that this polymerization transition can indeed dramatically slow down the nucleation process. This slowdown is caused by the relative stability of the peptide clusters that need to align or dissolve before a stable fibrillar structure can form. Polymerization of disordered oligomers may thus lead to an increased lifetime of early oligomers, which are known to be involved in neurodegenerative disorders such as Alzheimer's disease.^{4,69}

The Bereau–Deserno model in combination with FFS simulations has been successful in elucidating the nucleation mechanism as a function of the side-chain hydrophobicity. We note that the reaction kinetics have been derived using an implicit-solvent four-beads-per-amino-acid CG model; therefore, the free-energy landscape is smoothed and the time scales presented will be faster than those obtained experimentally or from atomistic simulations. While a simple scaling of the kinetics is probably not possible, the model is nevertheless able to give qualitative insight into the nucleation mechanism. The predicted rate constants are in line with the generic conclusion that KLVFFAE yields the slowest nucleation kinetics; however, molecular simulations using a higher resolution model are recommended to resolve the structural details of the polymerized phase and the various aggregated states of amyloid fibrils. Finally, in FFS, the equations of motion are integrated forward in time only and might lead to unwanted dependence on the initial configurations. This is especially the case when the order parameters used to describe the interfaces are not optimal. The discrepancy between the location of the TSE of the forward transition and the maximum of the free energy reveals that this might indeed be the case. Time-reversible methods such as TPS^{24,25,40–42} or TIS^{43,44} may ameliorate this situation in a future study.

CONCLUSIONS

We predicted the kinetic primary nucleation pathways of short amyloidogenic peptides for varying side-chain hydrophobicity using a sequence specific off-lattice coarse-grained model. In line with previous work we find a switch from the 1SN to the 2SN mechanism with increasing hydrophobicity. For hydrophilic GNNQQNY peptides that grow via the 1SN mechanism, we found evidence of a critical nucleus of four peptides, however, larger fibrils grow increasingly unstable, most likely due to the instability of a single-layer fibril and the chosen order parameter. The VEALYL and KLVFFAE peptides nucleate via

the two-step route. We provide a molecular view of the nucleation pathway and find that it occurs via a self-templating process. The structure of the amyloid fibrils was found to be strongly dependent on the relative hydrophobic strength of side-chains along the sequence: β -sheets in the fibril are oriented such that a core of the relative strongest hydrophobic residues is formed along the fibril axis. Finally, and perhaps most importantly, we find that the intermediate strength KLVFFAE peptide becomes trapped along the 2SN route in a polymerized intermediate phase that is intrinsically slow. Future experimental kinetic studies on these or similar peptides should be able to test our hypotheses.

AUTHOR INFORMATION

Corresponding Author

*E-mail: p.g.bolhuis@uva.nl

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is part of the research programme VICI 700.58.442, which is financed by The Netherlands Organization for Scientific Research (NWO).

REFERENCES

- (1) Chiti, F.; Dobson, C. M. Amyloid Formation by Globular Proteins under Native Conditions. *Nat. Chem. Biol.* **2009**, *5*, 15–22.
- (2) Haass, C.; Selkoe, D. J. Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid Beta-Peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101–112.
- (3) Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M. Designing Conditions for In Vitro Formation of Amyloid Protofilaments and Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3590–3594.
- (4) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
- (5) Dobson, C. M. Protein Aggregation and its Consequences for Human Disease. *Protein Pept. Lett.* **2006**, *13*, 219–227.
- (6) Ferreira, S. T.; Klein, W. L. The A β Oligomer Hypothesis for Synapse Failure and Memory Loss in Alzheimer's Disease. *Neurobiol. Learn. Mem.* **2011**, *96*, 529–543.
- (7) Blanco, L. P.; Evans, M. L.; Smith, D. R.; Badtke, M. P.; Chapman, M. R. Diversity, Biogenesis and Function of Microbial Amyloids. *Trends Microbiol.* **2012**, *20*, 66–73.
- (8) Knowles, T. P. J.; Buehler, M. J. Nanomechanics of Functional and Pathological Amyloid Materials. *Nat. Nanotechnol.* **2011**, *6*, 469–479.
- (9) Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **2003**, *21*, 1171–1178.
- (10) Scheibel, T.; Parthasarathy, R.; Sawicki, G.; Lin, X.; Jaeger, H.; Lindquist, S. Conducting Nanowires Built by Controlled Self-Assembly of Amyloid Fibers and Selective Metal Deposition. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 4527–4532.
- (11) Channon, K. J.; Devlin, G. L.; MacPhee, C. E. Efficient Energy Transfer within Self-Assembling Peptide Fibers: A Route to Light-Harvesting Nanomaterials. *J. Am. Chem. Soc.* **2009**, *131*, 12520–12521.
- (12) Mezzenga, R.; Schurtenberger, P.; Burbidge, A.; Michel, M. Understanding Foods as Soft Materials. *Nat. Mater.* **2005**, *4*, 729–740.
- (13) Jones, O. G.; Mezzenga, R. Inhibiting, Promoting, and Preserving Stability of Functional Protein Fibrils. *Soft Matter* **2012**, *8*, 876–895.
- (14) Schor, M.; Bolhuis, P. G. The Self-Assembly Mechanism of Fibril-Forming Silk-Based Block Copolymers. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10457–10467.
- (15) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. Selective Differentiation of Neural

Progenitor Cells by High-Epitope Density Nanofibers. *Science* **2004**, *303*, 1352–1355.

(16) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer Adds to Preformed Structured Oligomers of A Beta-Peptides by a Two-Stage Dock-Lock Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 111–116.

(17) Reddy, G.; Straub, J. E.; Thirumalai, D. Dynamics of Locking of Peptides Onto Growing Amyloid Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11948–11953.

(18) Kashchiv, D.; Cabriolu, R.; Auer, S. Confounding the Paradigm: Peculiarities of Amyloid Fibril Nucleation. *J. Am. Chem. Soc.* **2013**, *135*, 1531–1539.

(19) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(20) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(21) Baftizadeh, F.; Biarnes, X.; Pietrucci, F.; Affinito, F.; Laio, A. Multidimensional View of Amyloid Fibril Nucleation in Atomistic Detail. *J. Am. Chem. Soc.* **2012**, *134*, 3886–3894.

(22) Baftizadeh, F.; Pietrucci, F.; Biarnes, X.; Laio, A. Nucleation Process of a Fibril Precursor in the C-Terminal Segment of Amyloid-beta. *Phys. Rev. Lett.* **2013**, *110*, 168103.

(23) Buchete, N. V.; Hummer, G. Peptide Folding Kinetics From Replica Exchange Molecular Dynamics. *Phys. Rev. E* **2008**, *77*, 030902.

(24) Dellago, C.; Bolhuis, P. G. Transition Path Sampling and the Calculation of Rate Constants. *J. Chem. Phys.* **1998**, *108*, 1964–1977.

(25) Bolhuis, P. G.; Chandler, D.; Dellago, C. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

(26) Allen, R. J.; Frenkel, D.; ten Wolde, P. R. Simulating Rare Events in Equilibrium or Nonequilibrium Stochastic Systems. *J. Chem. Phys.* **2006**, *124*, 024102.

(27) Schor, M.; Vreede, J.; Bolhuis, P. G. Elucidating the Locking Mechanism of Peptides Onto Growing Amyloid Fibrils Through Transition Path Sampling. *Biophys. J.* **2012**, *103*, 1296–1304.

(28) Reddy, A. S.; Chopra, M.; de Pablo, J. J. GNNQQNY - Investigation of Early Steps During Amyloid Formation. *Biophys. J.* **2010**, *98*, 1038–1045.

(29) Gsponer, J.; Haberthur, U.; Cafilisch, A. The Role of Side-Chain Interactions in the Early Steps of Aggregation: Molecular Dynamics Simulations of an Amyloid-Forming Peptide from the Yeast Prion Sup35. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5154–5159.

(30) De Simone, A.; Esposito, L.; Pedone, C.; Vitagliano, L. Insights into Stability and Toxicity of Amyloid-Like Oligomers by Replica Exchange Molecular Dynamics Analyses. *Biophys. J.* **2008**, *95*, 1965–1973.

(31) Strodel, B.; Whittleston, C. S.; Wales, D. J. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(32) Lu, Y.; Derreumaux, P.; Guo, Z.; Mousseau, N.; Wei, G. Thermodynamics and Dynamics of Amyloid Peptide Oligomerization are Sequence Dependent. *Proteins: Struct., Funct., Genet.* **2009**, *75*, 954–963.

(33) Luiken, J. A.; Bolhuis, P. G. Prediction of a Stable Associated Liquid of Short Amyloidogenic Peptides. *Phys. Chem. Chem. Phys.* **2015**, *17*, 10556–10567.

(34) Auer, S.; Dobson, C. M.; Vendruscolo, M.; Maritan, A. Self-Templated Nucleation in Peptide and Protein Aggregation. *Phys. Rev. Lett.* **2008**, *101*, 258101.

(35) Auer, S.; Meersman, F.; Dobson, C. M.; Vendruscolo, M. A Generic Mechanism of Emergence of Amyloid Protofilaments from Disordered Oligomer Aggregates. *PLoS Comput. Biol.* **2008**, *4*, e1000222.

(36) Qi, X.; Hong, L.; Zhang, Y. A Variational Model for Oligomer-Formation Process of GNNQQNY Peptide from Yeast Prion Protein Sup35. *Biophys. J.* **2012**, *102*, 597–605.

(37) Bereau, T.; Deserno, M. J. Generic Coarse-Grained Model for Protein Folding and Aggregation. *J. Chem. Phys.* **2009**, *130*, 235106.

(38) Allen, R. J.; Warren, P. B.; ten Wolde, P. R. Sampling Rare Switching Events in Biochemical Networks. *Phys. Rev. Lett.* **2005**, *94*, 018104.

(39) Allen, R. J.; Valeriani, C.; Rein ten Wolde, P. Forward Flux Sampling for Rare Event Simulations. *J. Phys.: Condens. Matter* **2009**, *21*, 463102.

(40) Dellago, C.; Bolhuis, P. G.; Chandler, D. On the Calculation of Reaction Rate Constants in the Transition Path Ensemble. *J. Chem. Phys.* **1999**, *110*, 6617–6625.

(41) Bolhuis, P. G.; Dellago, C.; Geissler, P. L.; Chandler, D. Transition Path Sampling: Throwing Ropes over Mountains in the Dark. *J. Phys.: Condens. Matter* **2000**, *12*, A147–A152.

(42) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv. Chem. Phys.* **2002**, *123*, 1–78.

(43) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A Novel Path Sampling Method for the Calculation of Rate Constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774.

(44) van Erp, T. S.; Bolhuis, P. G. Elaborating Transition Interface Sampling Methods. *J. Comput. Phys.* **2005**, *205*, 157–181.

(45) Limbach, H.; Arnold, A.; Mann, B. A.; Holm, C. ESPResSo - an Extensible Simulation Package for Research on Soft Matter Systems. *Comput. Phys. Commun.* **2006**, *174*, 704–727.

(46) Berryman, J. T.; Schilling, T. Sampling Rare Events in Nonequilibrium and Nonstationary Systems. *J. Chem. Phys.* **2010**, *133*, 244101.

(47) Kratzer, K.; Arnold, A.; Allen, R. J. Automatic, Optimized Interface Placement in Forward Flux Sampling Simulations. *J. Chem. Phys.* **2013**, *138*, 164112.

(48) Auer, S.; Dobson, C. M.; Vendruscolo, M. Characterization of the Nucleation Barriers for Protein Aggregation and Amyloid Formation. *HFSP J.* **2007**, *1*, 137–146.

(49) Bingham, R. J.; Rizzi, L. G.; Cabriolu, R.; Auer, S. Non-monotonic Supersaturation Dependence of the Nucleus Size of Crystals with Anisotropically Interacting Molecules. *J. Chem. Phys.* **2013**, *139*, 241101.

(50) Cabriolu, R.; Kashchiv, D.; Auer, S. Breakdown of Nucleation Theory for Crystals with Strongly Anisotropic Interactions between Molecules. *J. Chem. Phys.* **2012**, *137*, 204903.

(51) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Illustration of Transition Path Theory on a Collection of Simple Examples. *J. Chem. Phys.* **2006**, *125*, 084110.

(52) Weinan, E.; Vanden-Eijnden, E. Towards a Theory of Transition Paths. *J. Stat. Phys.* **2006**, *123*, 503–523.

(53) Borrero, E. E.; Escobedo, F. A. Reaction Coordinates and Transition Pathways of Rare Events via Forward Flux Sampling. *J. Chem. Phys.* **2007**, *127*, 164101.

(54) Valeriani, C.; Allen, R. J.; Morelli, M. J.; Frenkel, D.; Rein ten Wolde, P. Computing Stationary Distributions in Equilibrium and Nonequilibrium Systems with Forward Flux Sampling. *J. Chem. Phys.* **2007**, *127*, 114109.

(55) Cecchini, M.; Rao, F.; Seeber, M.; Cafilisch, A. J. Replica Exchange Molecular Dynamics Simulations of Amyloid Peptide Aggregation. *J. Chem. Phys.* **2004**, *121*, 10748–10756.

(56) Cecchini, M.; Curcio, R.; Pappalardo, M.; Melki, R.; Cafilisch, A. A Molecular Dynamics Approach to the Structural Characterization of Amyloid Aggregation. *J. Mol. Biol.* **2006**, *357*, 1306–1321.

(57) Mompeán, M.; González, C.; Lomba, E.; Laurents, D. V. Combining Classical MD and QM Calculations to Elucidate Complex System Nucleation: a Twisted, Three-Stranded, Parallel Beta-Sheet Seeds Amyloid Fibril Conception. *J. Phys. Chem. B* **2014**, *118*, 7312–7316.

(58) Auer, S.; Ricchiuto, P.; Kashchiv, D. Two-step Nucleation of Amyloid Fibrils: Omnipresent or Not? *J. Mol. Biol.* **2012**, *422*, 723–730.

(59) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A.; Riek, C.; Grothe, R.; Eisenberg, D. Structure of the Cross- β Spine of Amyloid-Like Fibrils. *Nature* **2005**, *435*, 773–778.

(60) Barz, B.; Wales, D. J.; Strodel, B. A Kinetic Approach to the Sequence-Aggregation Relationship in Disease-Related Protein Assembly. *J. Phys. Chem. B* **2014**, *118*, 1003–1011.

(61) Nasica-Labouze, J.; Meli, M.; Derreumaux, P. A Multiscale Approach to Characterize the Early Aggregation Steps of the Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35. *PLoS Comput. Biol.* **2011**, *7*, e1002051.

(62) Nasica-Labouze, J.; Mousseau, N. Kinetics of Amyloid Aggregation: a Study of the GNNQQNY Prion Sequence. *PLoS Comput. Biol.* **2012**, *8*, e1002782.

(63) Matthes, D.; Daebel, V.; Meyenberg, K.; Riedel, D.; Heim, G.; Diederichsen, U.; Lange, A.; de Groot, B. L. Spontaneous Aggregation of the Insulin-Derived Steric Zipper Peptide VEALYL Results in Different Aggregation Forms with Common Features. *J. Mol. Biol.* **2014**, *426*, 362–376.

(64) Ivanova, M. I.; Sievers, S. A.; Sawaya, M. R.; Wall, J. S.; Eisenberg, D. Molecular Basis for Insulin Fibril Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18990–18995.

(65) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid Fibril Formation by A β _{16–22}, a Seven-Residue Fragment of the Alzheimer's β -Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry* **2000**, *39*, 13748–13759.

(66) Favrin, G.; Irbäck, A.; Mohanty, S. Oligomerization of Amyloid A beta(16–22) Peptides Using Hydrogen Bonds and Hydrophobicity Forces. *Biophys. J.* **2004**, *87*, 3657–3664.

(67) Liang, Y.; Lynn, D. G.; Berland, K. M. Direct Observation of Nucleation and Growth in Amyloid Self-Assembly. *J. Am. Chem. Soc.* **2010**, *132*, 6306–6308.

(68) Liang, Y.; Pingali, S. V.; Jogalekar, A. S.; Snyder, J. P.; Thiagarajan, P.; Lynn, D. G. Cross-strand Pairing and Amyloid Assembly. *Biochemistry* **2008**, *47*, 10018–10026.

(69) Walsh, D. M.; Selkoe, D. J. A-Beta Oligomers - A Decade of Discovery. *J. Neurochem.* **2007**, *101*, 1172–1184.