Supplementary Information

Riccardo Pellarin and Amedeo Caflisch*

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Phone: (+41 44) 635 55 21
FAX: (+41 44) 635 68 62
email: caflisch@bioc.unizh.ch

* Corresponding author

February 22, 2006

February 22, 2006

Contents

1 Nomenclature

Energy terms

E^{vdW}	Van der Waals energy minimum for hydrophobic spheres
q	Dipole partial charge
E_b	Barrier height
E_π	Potential energy at the protected state π
E_β	Potential energy at the amyloid state β
dE	Potential energy difference between the π and the β states
$P(E_b, dE)$	CMAP dihedral potential
$C_{\phi_0}(\phi)$	Constraining dihedral potential
β-stable	Model with $dE \gtrapprox 0$ dihedral potential
β-unstable	Model with $dE \lessapprox -2.0$ dihedral potential

Monomer states and fibril morphology

π	Protected state
β	Amyloid state
m	Monomeric state
M	Micellar state
f	Monofilament state
F	Fibril state
$F(X_y)$	Fibril morphology symbol

System observables

n_p	Parallel polar contacts
n_h	Hydrophobic contacts

Kinetic observables

t_{lp}	Lag phase time obtained by exponential fitting
t_{50}	Lag phase measured at 50% amplitude (delay time)
k_e	Elongation rate
t_{50}^M	Time of micellization

Concentration and aggregation numbers

C	Total monomer concentration
C_M	Micelle concentration
C_M^r	Critical concentration of micelle formation
C_F^r	Critical concentration of fibril formation
N_T	Total number of simulated monomers
N_m	Number of dissociated monomers
N_M	Micelle aggregation number
N	Oligomer size
N^*	Nucleus size

Thermodynamics

$\Delta G_{\beta\pi}(N)$	Free energy difference between π and β states in an oligomer of size N
$\Delta G_{\beta\pi}(1)$, $\Delta G_{\beta\pi}$	Free energy difference between π and β in the monomeric form

2 Supplementary description of the model

The coarse-grained model developed for studying aggregation kinetics and thermodynamics is a compromise between the mesoscopic detail and the computational efficiency, which are conflicting requirements. Each monomer has internal flexibility and can interact through electrostatics and van der Waals forces with other monomers. Some of the model parameters are chosen to enforce a certain geometry (such as the bonds, the spheres radii and the angles). The strength of van der Waals and electrostatic energy terms are first varied to evaluate the effects on the formation of aggregates. Then the dihedral potential is changed to analyze different aggregation pathways and kinetics. The simplified model does not represent a particular protein; it is useful to understand different aggregation scenarios as observed for different amyloidogenic sequences and experimental conditions.

2.1 The force field

The monomer consists of 10 spherical beads, four of which represent the "backbone" (A2 A3 A6 A10) and six the "sidechains" (A1 A4 A5 A7 A8 A9) (Figure 1). The backbone consists of two identical dipoles with a partial charge q expressed in electronic units; this part of the monomer is designed to interact specifically by intermolecular dipole-dipole interactions. The larger beads represent the sidechains and interact by van der Waals forces. The energy E of the system is evaluated using the following force field formula:

$$\begin{aligned} E \;=\; & \sum_{\text{bonds}} k_b(l-l_0)^2 + \\ & \sum_{\text{angles}} k_a(\theta-\theta_0)^2 + \\ & \sum_{\text{dihedrals}} F(\phi) + \\ & \sum_{i,j} E_{ij}^{vdW}\left[\left(\frac{r_{ij}^{vdW}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{vdW}}{r_{ij}}\right)^{6}\right] + \\ & \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_m r_{ij}} \end{aligned} \tag{1}$$

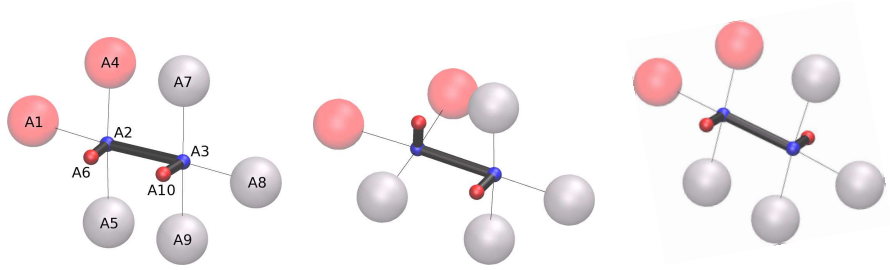

Figure 1: Model of the monomer: light gray spheres are hydrophilic and light red spheres are hydrophobic. The bold black bonds indicate the polar system. Here the positive charges are blue and negative are red. The monomer in the cis state is depicted in the left image, the +90 state in the center, the trans conformation on the right. The spheres drawn here do not reflect the actual van der Waals radii. Geometrical properties and force field parameters are described in the text. Labels indicate the sphere name.

where the sums are evaluated for all bonds, angles, dihedrals and sphere pairs i, j of the system. The variables l, θ and ϕ are the length of the bond, the angle and dihedral values, respectively, while r_{ij} is the distance between the sphere pair i, j. The values of the force constants k_b, k_a and the optimal distances l_0 and angles θ_0 are reported in Table 1. The "molecule" can change its conformation by rotation around the internal dihedral ϕ defined by the beads A6-A2-A3-A10. Depending on the simulation purpose the dihedral potential function $F(\phi)$ is either an harmonic function

$$F(\phi) \equiv C_{\phi_0}(\phi) = k_d(\phi - \phi_0)^2 \tag{2}$$

that restrains the value of the dihedral ϕ around the value ϕ_0, or a potential

$$F(\phi) \equiv P(\phi) \tag{3}$$

designed using the CMAP facility [1], with a grid size of 15 degrees. Several potentials P were investigated in the present work (see also Section 2.3).

The optimal van der Waals energy E^{vdW} and distance r^{vdW} as well as the partial charges q_i are listed in Table 2. The pair constants E_{ij}^{vdW} for the van der Waals interaction in Equation 1 are evaluated using the Lorentz-Berthelot mixing rules [2]. A dielectric constant $\epsilon_m = 1$ is used because the effects of the solvent are taken

Bond energy

Bead type	k_b ($kcal \cdot mol^{-1} \cdot \mathring{A}^{-2}$)	l_0 (Å)
A or C - B	1000.0	5.0
D - B	1000.0	2.0

Angle energy

Bead type	k_a ($kcal \cdot mol^{-1} \cdot rad^{-2}$)	θ_0 (degrees)
A or C or D - B - A or C or D	100.0	90.0

Table 1: Bonding parameter of the force field.

Name	Bead type	E^{vdW} [kcal/mol]	r^{vdW} [Å]	mass [a.u.]	charge q [e.u.]
A1	A	-0.1/-1.6 (∗)	2.5	500	0.0
A2	B	-0.1	2.0	500	0.29/0.52 (∗)
A3	B	-0.1	2.0	500	0.29/0.52 (∗)
A4	A	-0.1/-1.6 (∗)	2.5	500	0.0
A5	C	-0.1	2.5	500	0.0
A6	D	-0.1	2.0	500	-0.29/-0.52 (∗)
A7	C	-0.1	2.5	500	0.0
A8	C	-0.1	2.5	500	0.0
A9	C	-0.1	2.5	500	0.0
A10	D	-0.1	2.0	500	-0.29/-0.52 (∗)

Table 2: Nonbonding parameter of the force field. (∗) Variation of these parameters is investigated in Section 2.4.

into account implicitly by the nonbonding parameters E^{vdW} and q. Assuming an ellipsoidal symmetry, the volume of the monomer is 941 $\mathring{A}^3$, which roughly corresponds to the volume occupied by a peptide of 5 to 11 residues. The mass per bead is set to 500 a.u. This value corresponds to a mass of 4-5 residues, and is chosen to provide stability to the molecular dynamics simulations.

Two types of sidechains are defined: A2, A3 and A5 to A10 have a van der Waals energy minimum E^{vdW} of -0.1 kcal/mol, while A1 and A4, the hydrophobic

sidechains, have a much more favorable van der Waals energy minimum. These two sidechain types generate an "amphipathic" moment which allows the formation of amorphous aggregates such as micelles and the assembly of fibrils. Soreghan *et al.* have emphasized the surfactant properties of β-amyloid peptide and its capability to form solvent oriented structures [3]. For many amyloid proteins, amorphous on and off pathways intermediates have been detected [4–8]. More complex combinations of "sidechain" types could be envisaged for future investigations.

Given the geometry of the monomer (see Section 2.3) and the simplified force field, only two nonbonding parameters are relevant. The van der Waals energy minimum E^{vdW} of the hydrophobic beads A1 and A4 tunes the strength of the non-specific interaction while the partial charge q regulates the strength of the dipole-dipole interaction (see Section 2.4).

2.2 Simulation protocol

Simulations were performed at different temperature values (300-360 K), and concentrations (1.52 - 106.0 mM) using 125 monomers in a box with periodic boundary conditions. The size of the cubic simulation box defines the value of the concentration. A few runs were performed with 1000 monomers to investigate the "seeded" aggregation (see Section 4.2). The simulation protocol is the same for all runs. Monomers are initially placed in a cubic lattice. The system is then heated for 2 ps to the nominal temperature and equilibrated for 20 ps. In this first stage the integration time step is 2 fs and there is no shake constraint. The second stage is a more intensive equilibration; the monomer centers of mass are constrained at their position and simulated for 50 ns. The purpose is to equilibrate the dihedral degree of freedom. The third stage is the production, with previous constraints released. For the second and the third stages the time step is 50 fs, all bonds are restrained with SHAKE [9], and the leapfrog integrator is used for Langevin dynamics at a very low viscosity (0.01 ps^{-1}) which does not influence the thermodynamic properties. For all stages the cutoffs are set to 25 Å for the nonbonding list, 20.0 Å

and 18.0 Å for the nonbonding interactions cutoff and cuton, respectively, with a switching function [10].

2.3 Probing different monomer conformations

The two dipoles are orthogonal to each other in the states π_{90} and π_{-90} corresponding to $\phi = +90$ and -90, respectively. In these conformations the monomers cannot stack along a longitudinal axis, i.e., fibrils cannot be formed. These conformations represent the amyloid-protected state. The states β_0 and β_{180} correspond to cis ($\phi = 0$) and trans ($\phi = 180$) conformation, respectively. These two states can propagate a longitudinal stacking, namely they can form fibrils. The effects of different dihedral conformations were investigated by $1.5\mu s$ runs (125 monomers) with the harmonic potential C defined by the equation 2 (see Figure 2). Besides the π_{90}, π_{-90}, β_0 and β_{180} states mentioned above, conformations with a deviation of ± 15 and ± 30 degrees from cis or trans (noted as $\beta_{\pm 30}$, $\beta_{\pm 15}$, $\beta_{\pm 150}$,$\beta_{\pm 165}$ see Figure 2) were simulated to investigate the effect of monomer chirality on fibril structure. All simulations were performed at aggregation-promoting conditions (see Section 3.4), i.e., concentration of 20.88 mM, 310 K, and aggregation-promoting nonbonding parameters (see Section 2.4), i.e., $E^{vdW} = -1.3$ kcal/mol for the hydrophobic spheres and $q = 0.34$ electronic units.

For all monomeric conformations the system readily starts to form aggregates of different kinds. As mentioned above, conformations π_{90} and π_{-90} do not produce ordered aggregates but rather spherical micellar assemblies (see figure 4) where the hydrophobic spheres are partitioned into the core and the hydrophilic spheres are exposed. These micelles are in equilibrium with dissociated monomers and have an average size of 20-23 monomers (see Section 3.4). All other conformers associate into ordered structures, but interestingly only the chiral conformers produce fibril-like aggregates. The simulations where the monomers are constrained to be either in the state β_0 or β_{180} yield an ordered oligomeric assembly that resembles a disordered crystal lacking a precise cylindrical symmetry (see Figure 2.2) rather

than a fibril. The remaining simulations ($\beta_{\pm 15}, \beta_{\pm 30}, \beta_{\pm 165}, \beta_{\pm 150}$) yield a single cylindrical aggregate consisting of three to four filaments intertwined together, and assembled around the hydrophobic core (see Figure 2.1-3). These ordered aggregates display a twist clockwise or counter-clockwise depending on the chirality (in Figure 2 the circular arrows indicate the helicity of the fibril). These observations agree with a statistical mechanical model for the assembly of chiral molecules [11]. Taken together, the simulation results reveal a complex conformational scenario; in spite of a minimal set of system degrees of freedom, by far smaller than a real polypeptide, the morphological heterogeneity is noticeable. It is important to note that real fibrils do not display a large morphology assortment [12, 13]. In the case of $A\beta_{42}$ only two main morphologies are observed, and probably only two monomer conformations are favored among all possible to form fibrils [13].

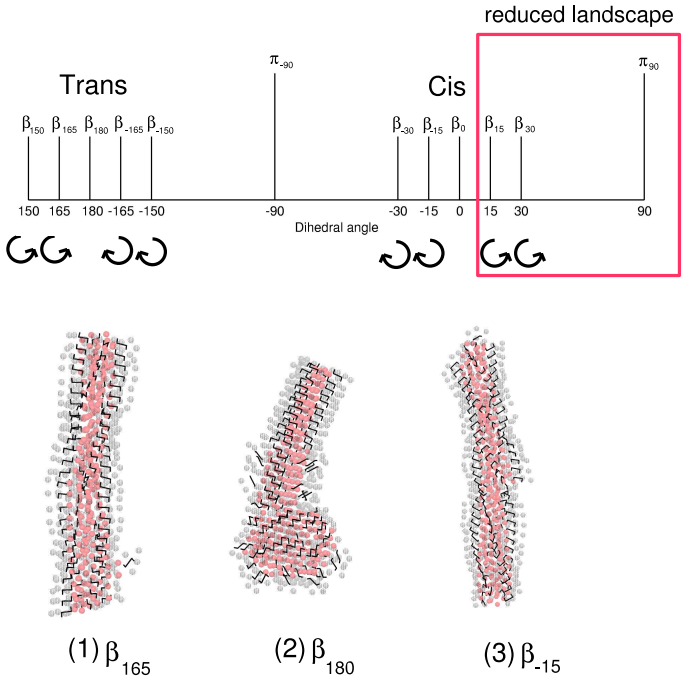

Figure 2: Effects of monomer dihedral angle on aggregation behaviour. Circular arrows indicate the helicity of the resulting fibril. The red box marks the range of dihedral angle values focussed upon by using the CMAP potential (see Figure 3). Fibrils resulting from the simulations of monomers in the state β_{165} (1), β_{180} (2), and β_{-15} (3). Black lines connect the spheres belonging to the polar system, light gray points are the hydrophilic spheres, and light red points are the hydrophobic spheres.

To prevent sampling of redundant, i.e., symmetry-related, conformations a value of 5.0 kcal/mol for all ϕ values outside the interval $7.5 - 97.5$ degrees was imposed by CMAP [1]. This procedure renders all conformations outside such interval inaccessible. Therefore the accessible ϕ-value interval includes the state π_{90}, which is ordered-aggregation protected, and the states β_{15} and β_{30} that form twisted fibrils (see Figure 2). It is now convenient to introduce a short notation for the two conformations: the *β-aggregation protected* state $\pi \equiv \pi_{90}$ and the *β-aggregation competent* state $\beta \equiv \beta_{15}$ or β_{30}.

A dihedral potential function can be introduced in the reduced region to explore different kinetic and thermodynamic properties of the monomer and to investigate their influence on fibril formation. The dihedral potential is defined by two parameters (Figure 3): E_π is the energy of the aggregation protected conformation for ϕ values ranging between 67.5 and 97.5 degrees, while E_b is the energy at the barrier, defined for the $52.5 < \phi < 67.5$ interval.

By fixing the reference state at $E_\beta = 0$ the dihedral potential is fully defined by $P(E_b; dE)$ where E_b is the potential value at the barrier, and $dE = E_\pi - E_\beta$ is the potential difference between the states π and β. As an example $P(1.0; -2.5)$ is

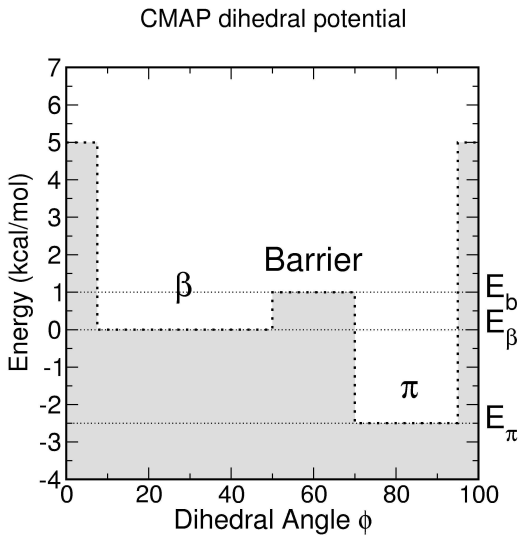

Figure 3: CMAP dihedral potential $P(1.0; -2.5)$. E_β is the energy for the aggregation-competent state, E_b is the energy at the barrier, and E_π is the energy at the protected state.

a potential with a 1.0 kcal/mol barrier for the $\beta \rightarrow \pi$ conversion, and a protected state π stabilized by -2.5 kcal/mol.

2.4 Probing nonbonding parameters

The effects of the variation of the two nonbonding parameters, E^{vdW} and q, are explained in this section. For three different potentials [$P(0;0)$, $P(1.0;-2.5)$ and $C_{+90}(\phi)$] simulations at different values of E^{vdW} for the hydrophobic spheres and q are performed (Table 3). There are different types of ordered and disordered aggregates of increasing complexity: micelles (M), single filament (f), non-twisted bundles of filaments ($F(I)$), twisted bundles of filaments ($F(II)$), and other ordered aggregates ($F(III)$) that cannot be classified in the previous two groups (Figure 4). The bundles (fibrils) can be of different sizes (two to four filaments, noted with a number, see caption of Table 3), but also a single fibril can present segments with a variable number of filaments.

All three investigated potentials show a common feature: the variation of the two parameters defines multiple phase change. At low q and marginally favorable E^{vdW} the monomers are dissociated. This corresponds to the top-right corner of the tables. An equilibrium of monomers with oligomers at the transition points is often observed, or eventually micelles with fibrils (as in the case of potential $P(1.0;-2.5)$, $q = 0.31$ and $E^{vdW} = -1.6$). At the bottom-left corner no coexistence is observed (pure ordered aggregates).

The restrained potential. The simulations performed with the $C_{+90}(\phi)$ potential are control simulations. As mentioned above, the π state ($\phi = +90$) cannot form any ordered aggregate but only micelles. As it is evident from Table 3, micelles formation occurs for $E^{vdW} \leq -1.0$ kcal/mol, weakly depending on the value of the charge q.

The β-unstable potential $P(1.0;-2.5)$. Favorable values of E^{vdW} and high values of q promote the ordered aggregation. $F(I)$ fibrils are observed at high q, together with single filaments f. The size of the bundle increases with more

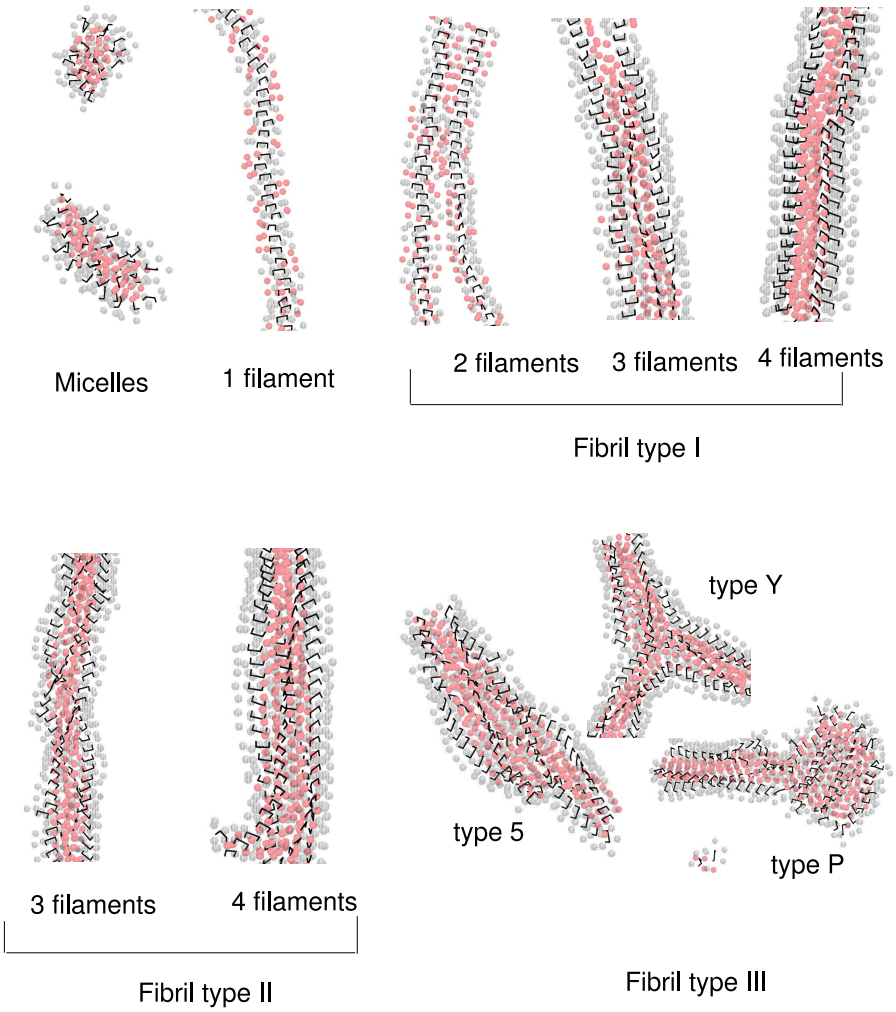

Figure 4: Different aggregation morphologies observed in simulations. The dipole system is indicated in black, the hydrophilic spheres are indicated in transparent gray, and the hydrophobic spheres are light red. Micelles are non-ordered spherical or elliptical metastable aggregates. Fibrils are multi-filament assemblies. Fibrils of "type I" are not twisted, while fibrils of "type II" are twisted. Fibrils of "type III" include everything that cannot be classified in the previous types: type III-5 is a bundle of 5 filaments, III-y is a cross of two or more fibrils, and type III-p is a planar fibril similar to a ribbon.

favorable E^{vdW}, i.e., by increasing the hydrophobicity. Twisted fibrils $F(II)$ are present at lower q, indicating that the twisting is a result of balancing of these two parameters.

The β-stable $P(0;0)$ potential. The region where the fibrillation occurs is larger than in the $P(1.0;-2.5)$ case, which is a consequence of the larger thermo-

Potential $C_{+90}(\phi)$

E^{vdW}	q=0.52	0.45	0.4	0.36	0.34	0.31	0.29
-0.1	m	m	m	m	m	m	m
-0.4	m	m	m	m	m	m	m
-0.7	m	m	m	m	m	m	m
-1.0	m+M	m+M	m	m	m	m	m
-1.3	M	M	m+M	m+M	m+M	m+M	m+M
-1.6	M	M	M	M	M	M	M

Potential $P(1.0;-2.5)$

E^{vdW}	q=0.52	0.45	0.4	0.36	0.34	0.31	0.29
-0.1	m+f:1	m	m	m	m	m	m
-0.4	F(I_2):1	m	m	m	m	m	m
-0.7	F(I_3):1	m	m	m	m	m	m
-1.0	F(I_3):2	F(II_3):2	m	m	m	m	m
-1.3	F(I_4):3	F(II_{3-4}):3	F(III_P):3	F(II_4):3	m+F(II_4):3	m+M	m+M
-1.6	F(III_P):3	F(II_{3-4}):4	F(III_Y):4	F(III_P):4	F(II_{3-4}):4	M+F(II_4):4	M

Potential $P(0;0)$

E^{vdW}	q=0.52	0.45	0.4	0.36	0.34	0.31	0.29
-0.1	f:1	m+f:1	m	m	m	m	m
-0.4	F(I_{2-3}):1	m+F(I_3):1	m+f:1	m	m	m	m
-0.7	F(I_{4-5}):1	m+F(I_3):1	m+F(I_3):1	m	m	m	m
-1.0	F(I_{2-4}):1	F(I_3):1	m+F(II_{3-4}):1	m+F(II_4):2	m+F(II_4):2	m+F(II_4):2	m+F(II_4):2
-1.3	F(II_{3-4}):1	F(I_{2-4}):1	F(II_4):2	F(III_5):2	m+F(II_4):3'	m+F(II_3):3'	m+F(II_{3-4}):3'
-1.6	F(III_P):2	F(III_P):2	F(III_P):2	F(II_{3-4}):3'	F(II_{3-4}):3'	F(II_3; III_P):3'	F(II_{3-4}):3'

Table 3: Effect of variation of the hydrophobic strength (E^{vdW} is the van der Waals potential energy well of spheres A1 and A4) and the charge q. Legend: (m) monomers, (M) micelles, (f) filaments, (F) fibrils. For the fibrils it is notated the type and the number of filaments: e.g. F(II_{3-4}) is a fibril of type II with 3 to 4 filaments. The number after the colon is the type of pathway followed by the simulation (see Figure 5). The sign plus "+" indicates coexistence of different phases. The concentration is 20.88 mM and the temperature is 310 K. The colors of the fields stand for the main aggregation morphology: green for micelles, blue for single filaments f, pink for $F(I)$, orange for $F(II)$, and red for $F(III)$.

dynamic accessibility of the β state. It is worth noting that chemically or mutation denaturated proteins are more susceptible to form fibrils than their standard conditions or wild type counterparts, respectively [14–16].

The aggregation process can be monitored by the number of parallel polar contacts and hydrophobic contacts along the trajectories. A parallel polar contact is formed whenever sphere 6 and 2 or sphere 10 and 3 of different monomers are closer than 5 Å. This selection of contacts defines the parallel aggregation, whereas the antiparallel is not observed because of the amphipathicity of the monomer (see above). A hydrophobic contact is formed whenever spheres 1 or 4 of different monomers are closer than 5 Å. The total number of polar n_p and hydrophobic n_h contacts is evaluated for each frame:

$$\begin{aligned} n_p &= \frac{1}{2}\sum_{i=A6}\sum_{j=A2}\delta(r_{ij}\leq 5) + \\ &\quad \frac{1}{2}\sum_{i=A10}\sum_{j=A3}\delta(r_{ij}\leq 5) \qquad (4) \\ n_h &= \frac{1}{2}\sum_{i=A1,A4}\sum_{j=A1,A4}\delta(r_{ij}\leq 5) \qquad (5) \end{aligned}$$

where, for instance, the summation index $i = A1, A4$ runs for all A1 and A4 spheres, and the function δ is equal to 1 if the distance between the two spheres r_{ij} is less than 5 Å. These quantities are used as progress variables of the aggregation process. A point in Figure 5 represents values of n_p and n_h of a single snapshot. Fibril formation corresponds to a trace of points that spans from the origin (monomeric state) to a maximal value of both variables that is around 225 for n_p and 600 for n_h in the simulation with 125 monomers. Given the geometry and size of the monomer the number of parallel polar and hydrophobic contacts per monomer incorporated into a fibril is about 2 and 5, respectively.

The left and right plot of figure 5 display the (n_p, n_h)-values of the snapshots saved along the simulations with the $P(0;0)$ and $P(1.0;-2.5)$ potentials respectively. Different pathways of fibril formation can be identified. The meaning of the

different pathways can be understood considering first the path number 2 which follows a straight line for both potential models. The variables under consideration depend linearly on each other in path 2, namely for any association event n_p, and n_h increase by a value that is constant along the aggregation process. The physical meaning of this behavior is that the fibril, once it has nucleated, progressively increases its size by absorbing monomers or small oligomers. During elongation the morphology of the progressing fibril it is similar to the final fibril.

There are different considerations for path 1, path 3-3' and path 4. Path 1 is a double stage transition for both potentials. In the first stage monomers form aggregates that maximize the number of polar interactions. These oligomers assemble by increasing hydrophobic contacts number in the second stage. They can be identified with single filaments. Only for the $P(1.0;-2.5)$ potential, paths of type 3 and 4 are observed. Clouds of points in the low n_p and high n_h can be interpretated as disordered aggregates. From visual examination, these on-pathway assemblies are micellar-like oligomers very similar to those obtained for the $C_{+90}(\phi)$ potential.

In the case of path 3' of $P(0;0)$ potential, these on-pathway micellar aggregates are not present: there is a faster transition from monomer state to fibril state. Path 3' is thus different from path 3. The pathway 4 is observed only at high hydrophobic strength; it is similar to pathway 3, but shifted towards the high hydrophobic contacts content. In other words it is a nucleation from a bigger micellar aggregate. Likewise the pathway 3' is present at high hydrophobic strength for the $P(0;0)$ potential. As mentioned above the micellar state is absent for this pathway, it is consequently equivalent to path 2, shifted towards high content of hydrophobic contacts. Pathways 3 and 4 are therefore qualitatively the same, and the same is valid for pathways 2 and 3'. This allows a characterization of the fibril formation pathways in three main classes (see Figure 6).

The path numbers are reported in Table 3 for all parameters pairs. The preferred path depends on the values of q and E^{vdW}. For potential $P(1.0;-2.5)$ the

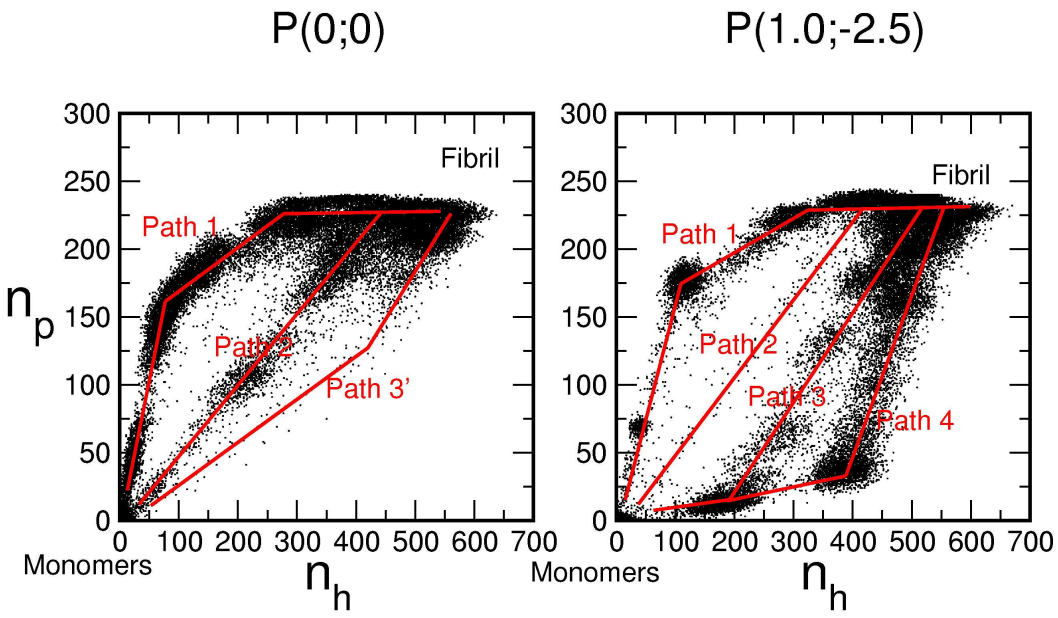

Figure 5: Pathways of fibril formations. The number of parallel polar interactions n_p and the number of hydrophobic interactions n_h for all trajectories defined in Table 3 for $P(0;0)$ and $P(1.0;-2.5)$ potentials. The clustering of points permits a classification of diverse aggregation pathways. Red lines are meant to guide the eyes.

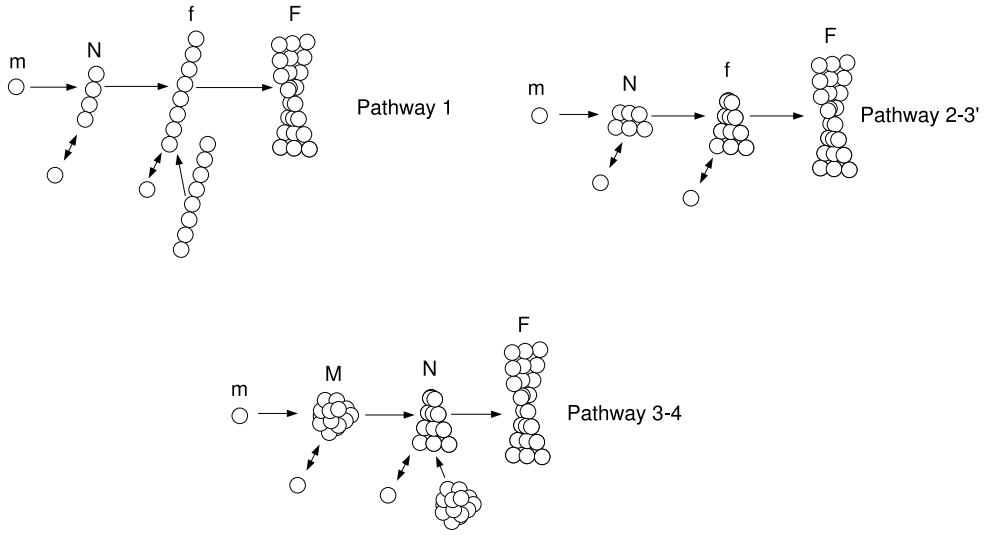

Figure 6: Pathways classification. Legend: (m) monomer, (M) micelle, (N) nucleus, (f) filament, (F) fibril.

predominant path changes from 1 to 4 by increasing the strength of the hydrophobic contacts. For the $P(0;0)$ potential the variation is influenced by both parameters.

These results allow us to describe the distinctive aggregation pathways for the

two potential models. For the β-**unstable** model $P(1.0; -2.5)$ the micellar state is an intermediate required for nucleating the ordered aggregation. Especially when the monomer specific affinity is low, i.e., when the charge q is small, fibril formation occurs via micellar intermediates. The hydrophobic interactions are essential for the fibril nucleation step. For the β-**stable** $P(0; 0)$ potential the structure interconversion is kinetically fast, the monomers depositing onto an ordered aggregate promptly convert its conformation to the β state. The elongation process is driven by polar interactions and hydrophobic interactions play a role only in the fibril morphology, i.e., filaments assembly.

With the values of $E^{vdW} = -1.3$ kcal/mol and $q = 0.34$ e.u. an equilibrium between monomers and fibrils is reached at the final stage of the β-unstable $P(1.0; -2.5)$ potential simulations (Table 3). This thermodynamic behavior is an important feature of a realistic model system. For this reason these values are adopted for all kinetic and thermodynamic analysis in the following sections and the main text.

3 Supplementary methods

In this section, and the main text, values of $E^{vdW} = -1.3$ kcal/mol and $q = 0.34$ e.u. are used for the van der Waals energy minimum and the partial charge, respectively.

3.1 Evaluation of system kinetics

The lag phase time t_{lp} is extracted from the exponential fitting of the time series of the number of parallel polar contacts n_p (defined by Equation 4). The fitting function is

$$n_p(t) = n_p(0) + [n_p(\infty) - n_p(0)](1 - e^{-k_e(t-t_{lp})})S(t, t_{lp}) \quad (6)$$

where $n_p(0)$ is the initial number of polar contacts, $n_p(\infty)$ the equilibrium value, and k_e the elongation rate. $S(t, t_{lp})$ is a switching function that is 0 for $t < t_{lp}$ and 1 for $t \geq t_{lp}$ and is needed to force the function $n_p(t)$ to have a constant value for $t < t_{lp}$ (Figure 7). The **lag phase** lasts during $t < t_{lp}$ while the **final fibril-monomer equilibrium** is established at times $t > 10t_e$ when the function $n_p(t)$ exceeds 90% of its maximal value $n_p(\infty)$, where $t_e = 1/k_e$. The **elongation** occurs during $t_{lp} \leq t \leq 10t_e$.

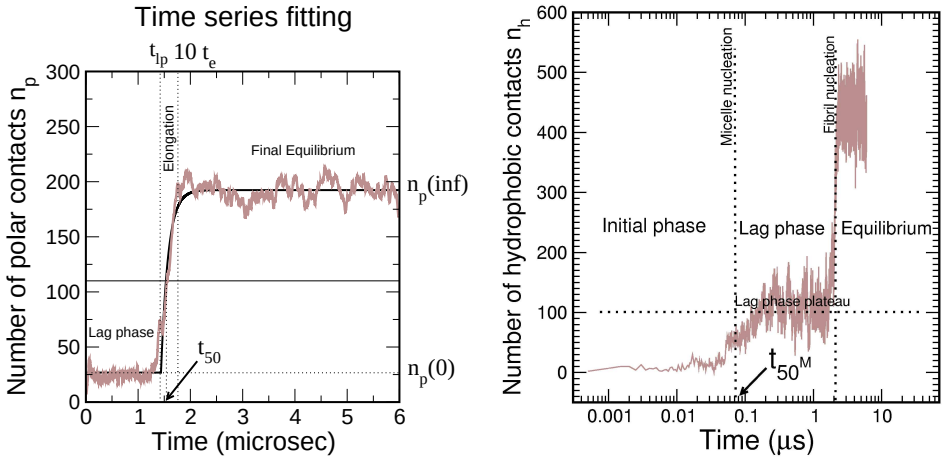

Figure 7: Example of the fitting of a time series of the number of parallel polar contacts n_p (left). Time series of the number of hydrophobic contacts n_h and the time of micelle formation t_{50}^M (right).

For the kinetic investigations of Sections 3.7 and 3.8 it is more appropriate to use a slightly different definition of the lag phase time, i.e., the time needed to reach 50% of the maximal amplitude t_{50} [17] (Figure 7). The t_{50} (termed delay time henceforth) is more robust than the lag phase time t_{lp}, especially for nucleation events with a short lag time. The lag phase time t_{lp} is used only for an exact measure of the lag phase, e.g. when thermodynamic properties at the lag phase are investigated (Sections 3.3, 3.4 and 3.5).

The number of parallel polar contacts n_p is not appropriate to monitor the nucleation kinetics of disordered aggregates such as micelles. For this purpose it is convenient to use the number of hydrophobic contacts n_h (defined in Equation 5). In analogy with t_{50}, one can define the *time of micellization* t_{50}^M as the time needed to reach 50% of lag phase plateau amplitude starting from $t = 0$ (see Figure 7). The t_{50}^M time can be evaluated only for β-unstable models and at low concentration, where the lag phase is long enough to separate the micellization from the fibril nucleation phase. The average value of t_{50}^M is 30 ns at C=8.5 mM (Figure 2 in the main text).

3.2 Clustering

A clustering algorithm is used to calculate the size of the oligomeric species along the simulations. It is based on the matrix of contacts D_{ij} between labeled monomers. Given a single frame of the simulation D_{ij} is equal to one if any sphere of monomer i is closer than 6.0 Å to any sphere of monomer j, otherwise it is zero. D_{ij} is equivalent to the first neighbor matrix, $d_{ij}^{(1)}$. The second neighbor matrix $d^{(2)}$ is constructed from $d^{(1)}$ including the neighbors of the first neighbors. The converged contact matrix $d_{ij}^{(\infty)}$ is defined by the following recursive sequence

$$
\begin{aligned}
d_{ij}^{(1)} &= D_{ij} \\
d_{ij}^{(n)} &= \begin{cases} 1 & \text{if} \quad d_{ij}^{(n-1)} = 1 \\ 1 & \text{if} \quad d_{ik}^{(n-1)} = 1 \quad \text{and} \quad d_{kj}^{(n-1)} = 1 \\ 0 & \text{otherwise} \end{cases} \\
d_{ij}^{(\infty)} &= \lim_{n \to \infty} d_{ij}^{(n)}
\end{aligned}
\tag{7}
$$

which yields a block matrix of ones and zeroes. Each block represents a cluster of monomers and contains first-neighbors, second neighbors, third and so on. This procedure is equivalent to a hierarchical clustering performed with a spanning tree technique [18]. With the converged contact matrix one can identify clusters (i.e. tagging each oligomer with an identification number), list monomers belonging to a specific oligomer and make statistics on the size of oligomers.

3.3 Cluster size histogram

The above definition of oligomeric species allows the statistical analysis of cluster size. The probability that a monomer is aggregated in a cluster of size N is:

$$
p(N) = \langle \frac{1}{N_T} \sum_{i=1,N_T} \delta_{i,t}(N) \rangle_t \tag{8}
$$

where N_T is the total number of simulated monomers, $\delta_{i,t}(N)$ is equal to 1 if the monomer i at time t is embedded in a cluster of size N, and the angular brackets are the time average. This function (termed cluster size distribution) can be evaluated for the lag phase or the final monomer-fibril equilibrium. The elongation phase cannot be analyzed by $p(N)$ being an out of equilibrium dynamic process.

The peaks of the $p(N)$ distribution can be interpreted as stable oligomeric species. The monomer peak ranges from $N = 1$ to 7, the micellar peak from $N = 8$ to 60, and the fibril peak from $N = 61$ to 125 (Figure 8). The height of the peaks depends on the relative stability of the β-competent state as well as the total monomer concentration (Figure 8). For the β-stable potential $P(0;0)$ the

micelle peak is not observed at any concentration value. With incresing concentration the monomer and micelle peak distributions are skewed towards high N values because multi-monomer collisions and multi-micellar collisions, respectively, transiently generate oligomers of a larger size.

3.4 Phase diagrams and critical concentrations

Phase diagrams of temperature T and concentration C were calculated only for the potential $P(1.0; -2.5)$ because of their computational demand (about 2 weeks on 80 CPUs). The probability of a monomer being in the monomeric (m), micellar (M) and fibrillar (F) states are evaluated as cumulative sum of the probability $p(N)$:

$$p_m = \sum_{N=1}^{7} p(N) \tag{9}$$

$$p_M = \sum_{N=8}^{60} p(N) \tag{10}$$

$$p_F = \sum_{N=61}^{125} p(N) \tag{11}$$

In the simulated system three possible phases can coexist: monomeric, micellar (disordered oligomer), and fibril (Figure 9). The results are robust for a monomer-micelle threshold in the range 5-10 and a micelle-fibril threshold between 50 and 70. The C, T-diagram of the lag phase is similar to the one of the final equilibrium for large T values, where the fibrillization is inhibited. At equilibrium one has a triphasic diagram.

The probability distribution $p(N)$ can be used to evaluate the critical concentration of micelle formation C_M^r. The micelle aggregation number in the lag phase is defined as

$$N_M = \sum_{N=8}^{60} N p^{lp}(N) \tag{12}$$

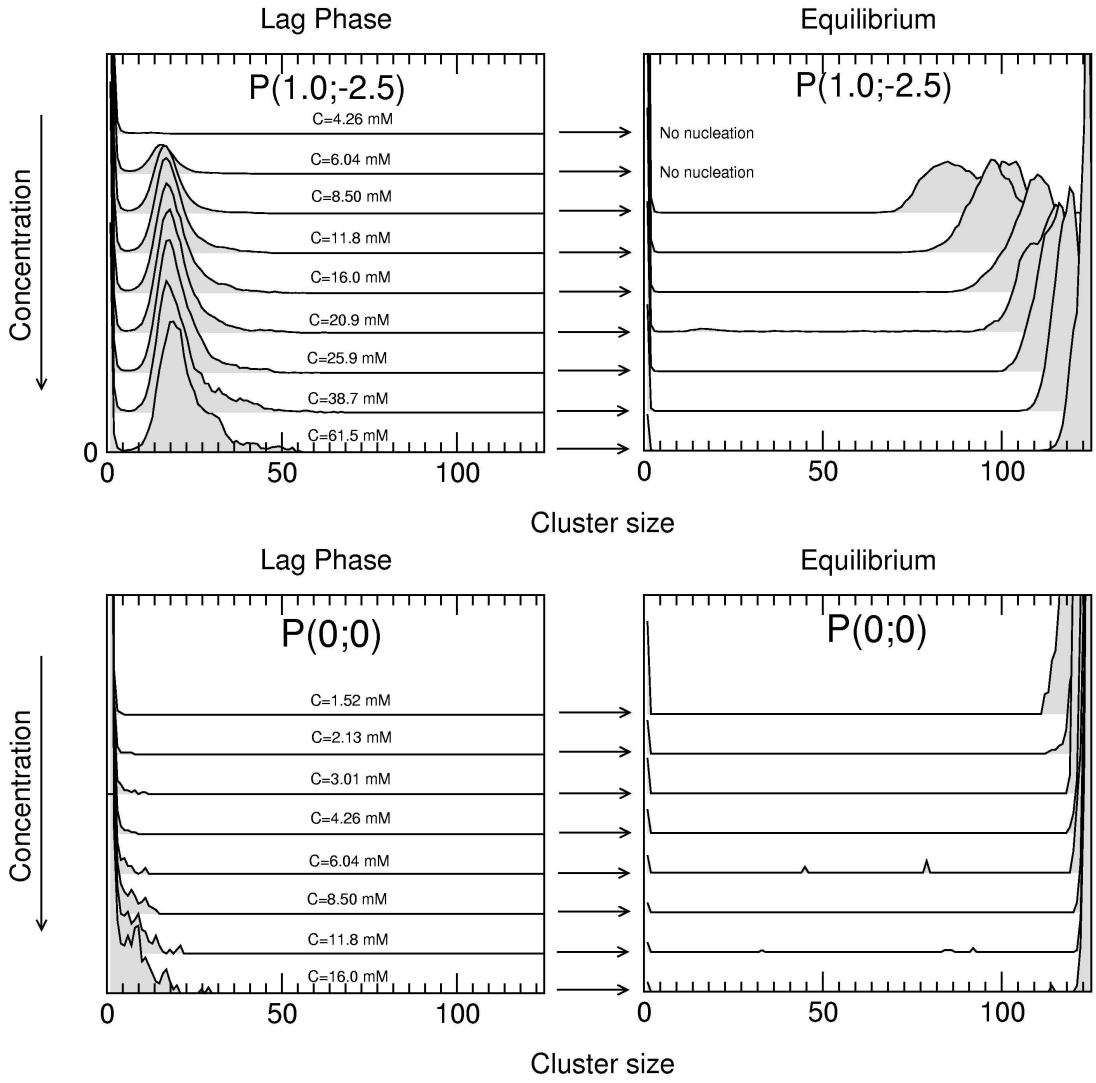

Figure 8: Cluster size histograms of the $P(1.0; -2.5)$ potential (top) and $P(0; 0)$ potential (bottom) calculated in the lag phase (left) and the final equilibrium (right). Histograms belonging to the same concentration are reported in the same row. The z-dimension represent the relative probability.

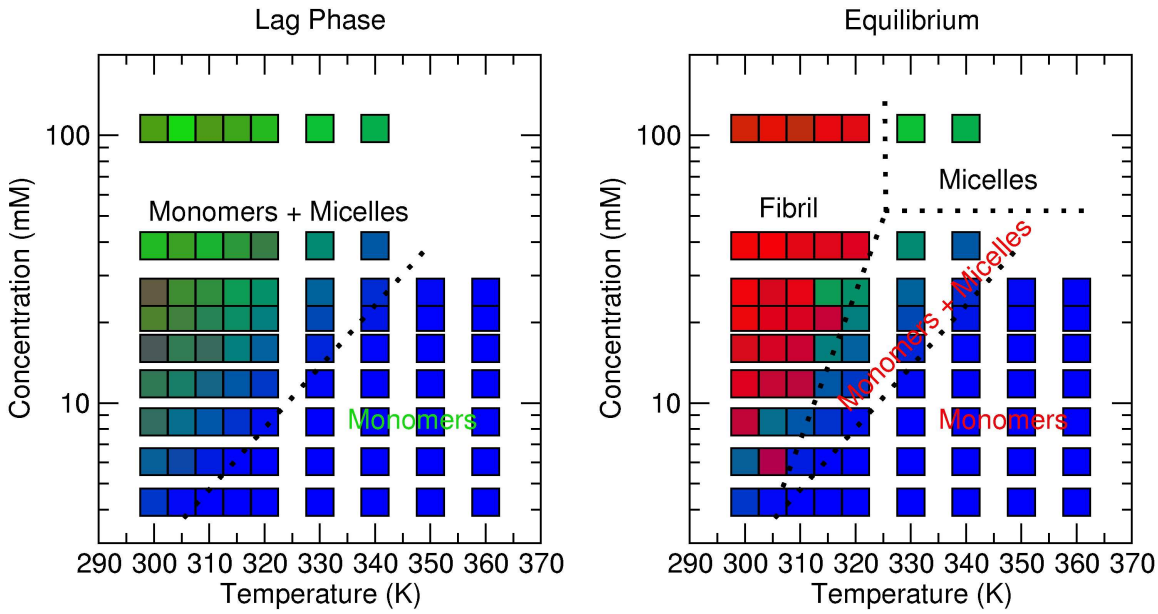

Figure 9: Phase diagram of the lag phase (left) and final equilibrium (right) for the potential $P(1.0; -2.5)$. Each color of the phase diagram is obtained by mixing the red, blue and green components according to the values of the calculated probabilities, i.e., red=p_F, blue=p_m, green=p_M (see Equations 9, 10 and 11).

where p^{lp} is the probability function evaluated only in the lag phase (where there is coexistence of micelles and monomers without fibrils). The number of micelles per simulation box is $N_T p_M N_M^{-1}$, where the total number of simulated monomers N_T is 125. The micelle concentration C_M is derived from the number of micelles in the simulation volume. The micelle aggregation number N_M and concentration are plotted in Figure 10 as a function of the total monomer concentration C for the potential $P(1; -2.5)$ in the lag phase. By extrapolating a linear fit of the concentration of micelles the critical concentration of micelle formation C_M^r can be evaluated. The value is C_M^r =4.36 mM.

Another important observable is the critical concentration of fibril formation C_F^r, that is obtained from the concentration of dispersed monomers in equilibrium with the final fibril. The number of dispersed monomers in the simulation box at the equilibrium phase is:

$$N_m = N_T p_m^{eq} = N_T \sum_{N=1}^{7} p^{eq}(N) \tag{13}$$

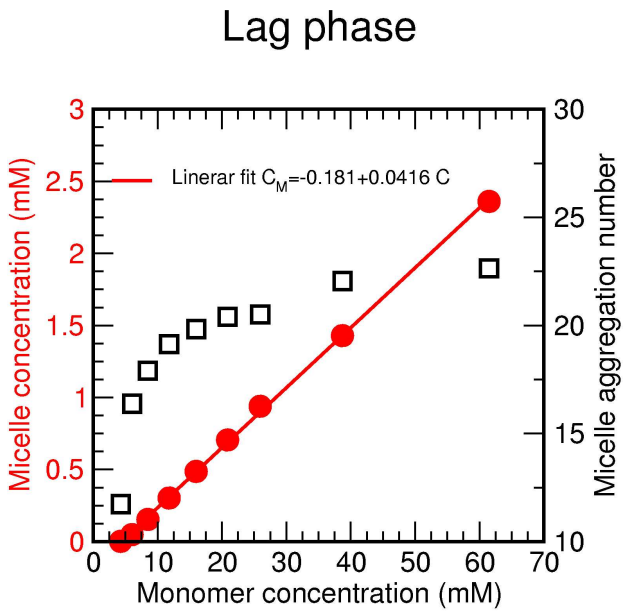

Figure 10: Lag phase of the P(1.0;-2.5) potential. Micelle concentration C_M (red circles) and the micelle aggregation number N_M (squares) as function of the total concentration of monomers C. The straight line is a linear fit whose parameters are reported in the graph.

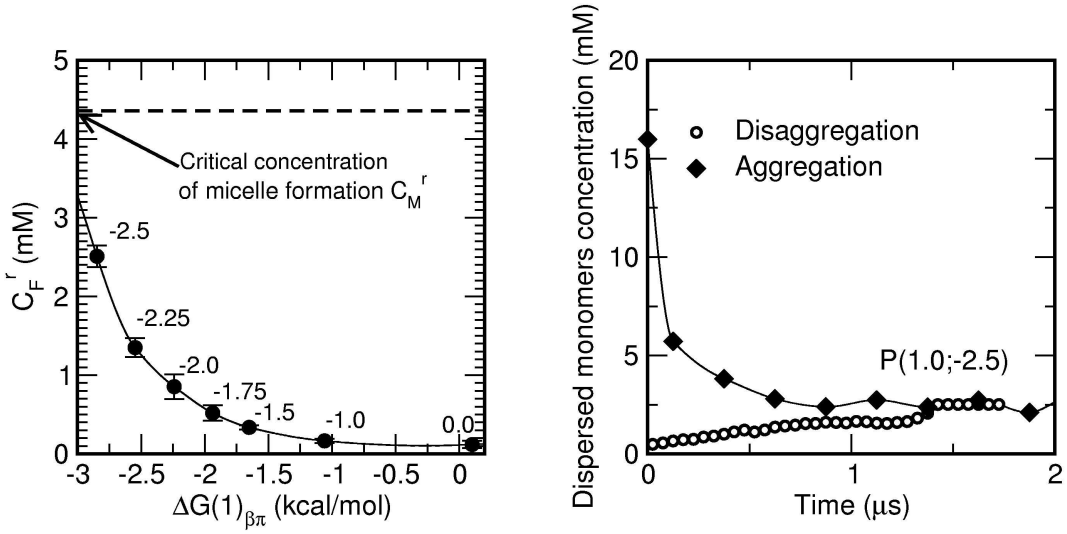

Figure 11: (Left) critical concentration of fibril formation C_F^r as a function of the monomeric $\Delta G_{\beta\pi}(1)$. The numbers displayed near the data points are the value of dE for the potential $P(1.0; dE)$. The critical concentration of micelle formation is indicated by a dashed line. (Right) Validation of the critical concentration of fibril formation, for the potential $P(1.0; -2.5)$. The dispersed monomer concentration is monitored along time for the aggregation process (filled diamonds) and along a simulation of disaggregation started from a previously formed fibril (empty circles).

where p^{eq} is the probability function evaluated at the equilibrium (where there is the coexistence of monomers and fibrils). The value of C_F^r can be evaluated by dividing N_m by the simulation volume. In Figure 11, C_F^r is displayed for different monomer potentials P and plotted against relative stability of the protected state $\Delta G_{\pi\beta}(1)$ (Table 4). Smaller differences in free energy result in lower critical concentration. In other words β-stable models are more reactive and shift the reaction towards the fibril formation. The critical concentration of micelle formation is always higher than the critical concentration of fibril formation; for this reason the micelles disappear at the monomer-fibril equilibrium.

The critical concentration of fibril formation is validated by additional simulations (Figure 11 right plot). The dispersed monomer concentration is evaluated dynamically for an aggregation trajectory (potential $P(1.0; -2.5)$, C=16.0 mM). The final concentration is equal to the predicted critical concentration of 2.5 mM (Figure 11 left plot, dE=-2.5). The reverse reaction is also performed to test if such C_F^r value is approachable also from the disaggregation direction. A fibril, previously prepared at C=16.0 mM, is simulated at the critical concentration of fibril formation (C=C_F^r=2.5 mM). The fibril progressively disassembles and the dispersed monomer concentration increases to the value of 2.5 mM. The same combination of forward and reverse reactions were used to experimentally test the robustness of the critical concentration of fibril formation for the β-amyloid peptide [19].

3.5 Aggregation process

Six monomeric states can be defined as a combination of β or π, and the three oligomeric species m, M, or F: $\pi m, \beta m, \pi M, \beta M, \pi F, \beta F$. As an example πF is the state for a monomer in the π conformation within a fibril. The transition matrix $T_{ij}(\Delta t)$ is defined as

$$T_{ij}(\Delta t) = p(i|j, \Delta t) \tag{14}$$

where i and j are two of the six states and $p(i|j, \Delta t)$ is the conditioned probability

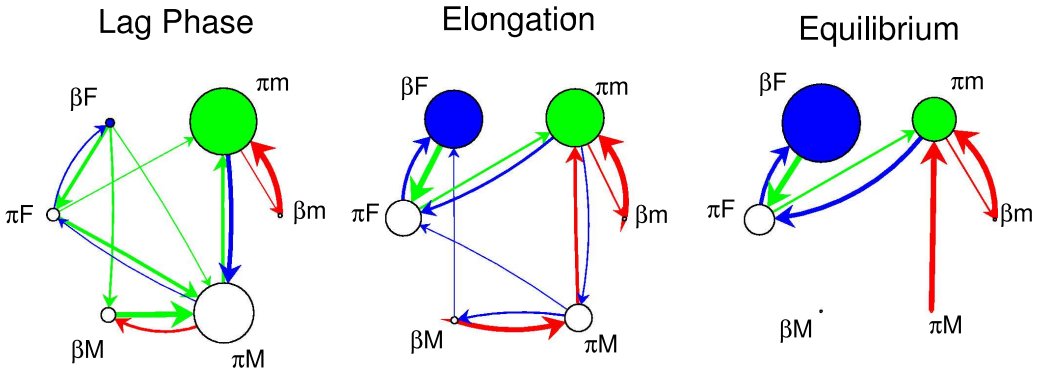

Figure 12: Simplified network representation of the transition matrix T_{ij} in the lag phase, elongation and equilibrium time regimes. The size of the nodes represents the state self transition T_{ii}, and the size of the links is the cross transitions T_{ij} with $i \neq j$. Assuming the initial state πm (green) and the final state βF (blue), blue arrows indicate the pathways leading form πm to βF and green arrows from βF to πm. Only transitions with probability greater than 0.05 are showed. The transition matrix is evaluated on the potential $P(1; -2.5)$ at concentration 11.8 mM and on 15 independent simulations of 12μs each.

of jumping to the j state from state i in a time Δt. The time Δt is chosen as the smallest available time in the simulation (the time of coordinate saving, 0.5 ns) to resolve the fastest events. From the simplified network representation of the transition matrix (Figure 12) it is clear that for the potential model $P(1.0; -2.5)$ the lag phase consists of a micellar association equilibrium $\pi m \leftrightarrow \pi M$ as well as intramonomer interconversions $\pi m \leftrightarrow \beta m$ and $\pi M \leftrightarrow \beta M$. Furthermore, in the lag phase the fibril state F is mainly accessible through the micellar state M. Therefore, in this time regime the fibril state is a transient ordered oligomer in equilibrium with micelles. At the final equilibrium micelles are very unstable and the main pathway is $\beta m \rightarrow \pi m \leftrightarrow \pi F \leftrightarrow \beta F$, indicating that monomers are attached to the fibril in the protected conformation π before assuming the amyloid conformation β. Furthermore, the βm state is off-pathway. This mechanism is consistent with kinetic experiments on radiolabeled $A\beta 40$ peptides where the transition from soluble to amyloid-like conformation of the peptide was suggested to be mediated by interaction with the fibril template (dock-lock mechanism) [20]. The equilibrium $\pi F \leftrightarrow \beta F$ reflects coexistence of monomers π and monomers β in the fibril.

From trajectory visualization it is clear that the conformations β and π populate different domains of the fibril; monomers β are found mainly in the central region of the fibril, whereas monomers π populate the disordered caps. Therefore, isolated monomers in equilibrium with the fibril are continously attaching to and detaching from the caps of the fibril in the β-protected conformation π. The growing phase is regulated by monomer addition, rather than oligomer addition. Collins *et al.* [21], using a combination of kinetic measures, reported a monomer addition growing for the yeast prion.

3.6 $\pi - \beta$ free energy difference

The aggregation number N is a natural progress variable to monitor the polymerization progress of an oligomer. The clustering procedure introduced in Section 3.2, can be used to calculate the free energy difference between the state π and the state β of a monomer belonging to an oligomer of size N:

$$\Delta G_{\beta\pi}(N) = G_{\pi}(N) - G_{\beta}(N) = -\langle kT \log \left(\frac{N_{\pi}(N)}{N_{\beta}(N)} \right) \rangle_N \qquad (15)$$

where $N_{\pi}(N)$ and $N_{\beta}(N)$ are the number of π-monomers and β-monomers, respectively, present in an oligomer of size N, and the angular bracket is the average over all oligomers of size N. This function of the number of monomers does not depend on the concentration because it is an intrinsic property of the oligomer and independent of the surrounding environment (Figure 13.A-B). The lack of dependence on concentration allows the evaluation of the function $\Delta G_{\beta\pi}$ using simulation data at different values of concentration (Figure 13.C).

The models with dihedral energy difference $dE < -2.5$ kcal/mol are not observed to nucleate, even at high concentration (C=61 mM) and long simulation times (16 μs); their aggregation number (or cluster size) N does not exceed $60-70$ (empty circles in Figure 13.C). The $dE = -2.5$ and -2.25 kcal/mol potentials, which are the most β-unstable potentials still capable of fibril formation, have a $\Delta G_{\beta\pi}(N) = 0$ at $N \approx 45$ and 30, respectively. Interestingly, these values roughly

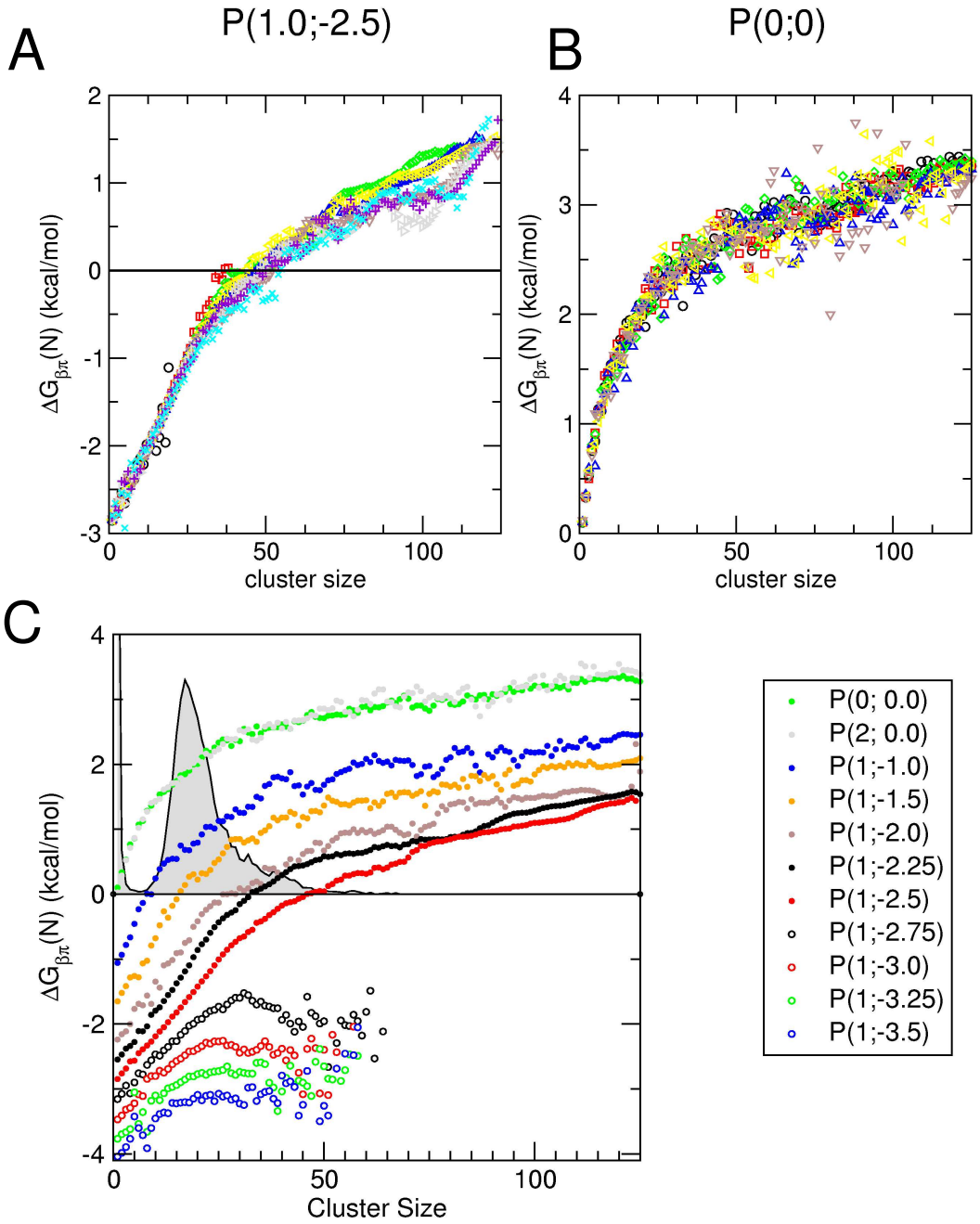

Figure 13: $\Delta G_{\beta\pi}(N)$ evaluated at different concentrations (different colors and symbols) for the potential $P(1;-2.5)$ (A) and $P(0;0)$ (B). (C) Variation of the $\Delta G_{\beta\pi}(N)$ evaluated for different potentials. The gray area in the background is the normalized cluster distribution during the lag phase.

Barrier height	
Potential	$\Delta G_{\beta\pi}(1)$ (kcal/mol)
$P(4;0)$	-0.165
$P(3;0)$	-0.0776
$P(2;0)$	-0.0255
$P(1;0)$	0.0844
$P(0;0)$	0.101

$\pi - \beta$ Relative stability	
Potential	$\Delta G_{\beta\pi}(1)$ (kcal/mol)
$P(1;-0.5)$	-0.452
$P(1;-1.0)$	-1.05
$P(1;-1.5)$	-1.65
$P(1;-1.75)$	-1.94
$P(1;-2.0)$	-2.24
$P(1;-2.25)$	-2.54
$P(1;-2.5)$	-2.84
$P(1;-2.75)$	-3.16
$P(1;-3.0)$	-3.47
$P(1;-3.25)$	-3.77
$P(1;-3.5)$	-4.04

Table 4: Free energy difference of isolated monomers $\Delta G_{\beta\pi}(1)$ for all investigated potential models.

correspond to the estimated nucleus sizes of 40 for the $dE = -2.5$ model and 27 for $dE = -2.25$ (see Figure 6 of the main text) indicating that the oligomeric size N_0 at which $\Delta G_{\beta\pi}(N_0) = 0$ identifies thermodynamically the nucleus size, in a way which is consistent with the probabilistic definition of nucleus exposed in Section 3.11. Comparing the $\Delta G_{\beta\pi}(N)$ and the cluster size distribution at lag phase (Figure 13.C), it is revealed that the values of N_0 for nucleating models are in the range where the cluster size distribution has a statistically significant probability. For nucleating β-unstable models $dE = -2.5$ and -2.25 kcal/mol the N_0 values are located at the micelle right tail, indicating that the nucleation step is initiated in an oligomer with a size larger than a micelle. The values of the monomeric free energy difference $\Delta G_{\beta\pi}(1)$ show a shift to more pronounced stabilization of the π state with slightly more negative values than dE (Table 4).

3.7 Concentration influence on kinetics of fibril formation

The time series of the number of parallel polar contacts n_p for different potential models and at different concentration values are shown in Figure 15. The analysis of the concentration dependence of the delay time t_{50} and the elongation rate k_e is reported in Figure 14 C-D. Potentials $P(1;-2.5)$, $P(1;-2.25)$, $P(2;0)$ and $P(0;0)$ were analyzed. Elongation rate k_e is evaluated by fitting with the Equation 6 the

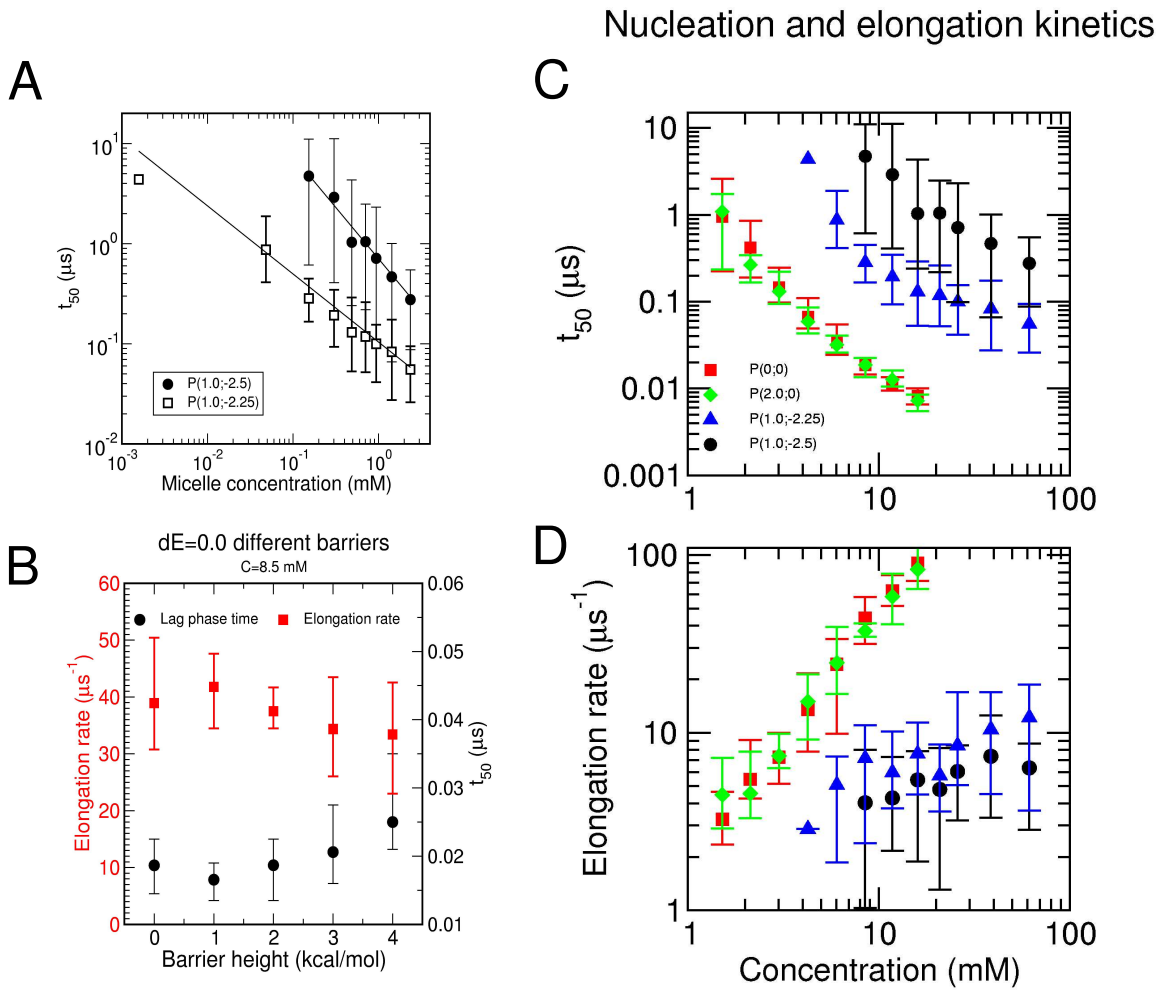

Figure 14: (A) Influence of the micelle concentration on the delay time t_{50} for β-unstable potentials $P(1.0;-2.5)$ and $P(1.0;-2.25)$. The power law fits are reported as continuous lines. The isolated data point at the lowest micelle concentration was not used for fitting for $P(1.0;-2.25)$. The error bars represent the minimum and the maximum value. (B) Effect on the delay time t_{50} (black circles) and the elongation rate k_e (red squares) of the barrier height variation for the potential $P(E_b;0)$. (C) Effect of concentration on delay time t_{50} for four potential models: $P(1.0;-2.5)$ black circles, $P(1.0;-2.25)$ blue triangles, $P(2.0;0)$ green diamonds, $P(0;0)$ red squares. The symbols represent the average value calculated on 15 simulations of $P(1.0;-2.5)$ and 10 simulations (all the others). The error bars represent the minimum and the maximum value. (D) Effect of concentration on the elongation rate k_e. Symbols and error bars as in (C).

C (mM)	P(0;0)	P(2;0)	P(1;-2.25)	P(1;-2.5)
1.52	10/10 (3.0 μs)	10/10 (3.0 μs)	n.a.	n.a.
2.13	10/10 (3.0 μs)	10/10 (3.0 μs)	n.a.	n.a.
3.01	10/10 (1.5 μs)	10/10 (1.5 μs)	n.a.	n.a.
4.26	10/10 (1.5 μs)	10/10 (1.5 μs)	1/10 (6.0 μs)	0/15 (3.0 μs)
6.04	10/10 (1.5 μs)	10/10 (1.5 μs)	10/10 (6.0 μs)	0/15 (6.0 μs)
8.50	10/10 (1.5 μs)	10/10 (1.5 μs)	10/10 (6.0 μs)	10/15 (12.0 μs)
11.8	10/10 (1.5 μs)	10/10 (1.5 μs)	10/10 (6.0 μs)	15/15 (12.0 μs)
16.0	10/10 (1.5 μs)	10/10 (1.5 μs)	10/10 (6.0 μs)	15/15 (12.0 μs)
20.9	n.a.	n.a.	10/10 (6.0 μs)	15/15 (3.0 μs)
25.9	n.a.	n.a.	10/10 (6.0 μs)	15/15 (3.0 μs)
38.7	n.a.	n.a.	10/10 (6.0 μs)	15/15 (3.0 μs)
61.5	n.a.	n.a.	10/10 (6.0 μs)	10/10 (3.0 μs)

Table 5: Table of all performed simulations for the concentration analysis of figure 14.C-D. The ratios indicate the number of nucleating trajectory over the number of independent simulations. The time reported in the brackets is the simulated time. (n.a.) the simulations were not performed at this concentration.

n_p time series of 15 independent simulations for the potential $P(1;-2.5)$, and 10 simulations for each of the remaining potentials. The delay time t_{50} is evaluated from the n_p time series as described in Section 3.1. Some of the $P(1;-2.5)$ runs were prolonged up to 12 μs (30 days on an Athlon 2800 GHz) to increase the number of nucleation events (Table 5).

The concentration dependence of the delay time and the elongation rate can be fitted by a power law $t_{50} = A_{50}C^{\gamma_{50}}$ and $k_e = A_e C^{\gamma_e}$, respectively, where C is the total monomer concentration. Results of the fit are reported in Table 6. Interestingly, the dependence of the rate of elongation on the concentration decreases significantly by increasing the stability of the protected state π. The reduced concentration dependence originates from competitive polymerizations, i.e., the elongation of the fibril and the presence of micelles. Furthermore, the concentration dependence of the delay time ($\gamma_{50} \neq 0$) for β-unstable potentials indicates that micelles promote the nucleation.

The nucleus sizes N^* can be extracted from the parameter γ_{50}, being $N^* = -2\gamma_{50}$, provided that *(a) the monomer concentration changes only by addition to and subtraction from polymers longer than the seed, (b) polymer formation by*

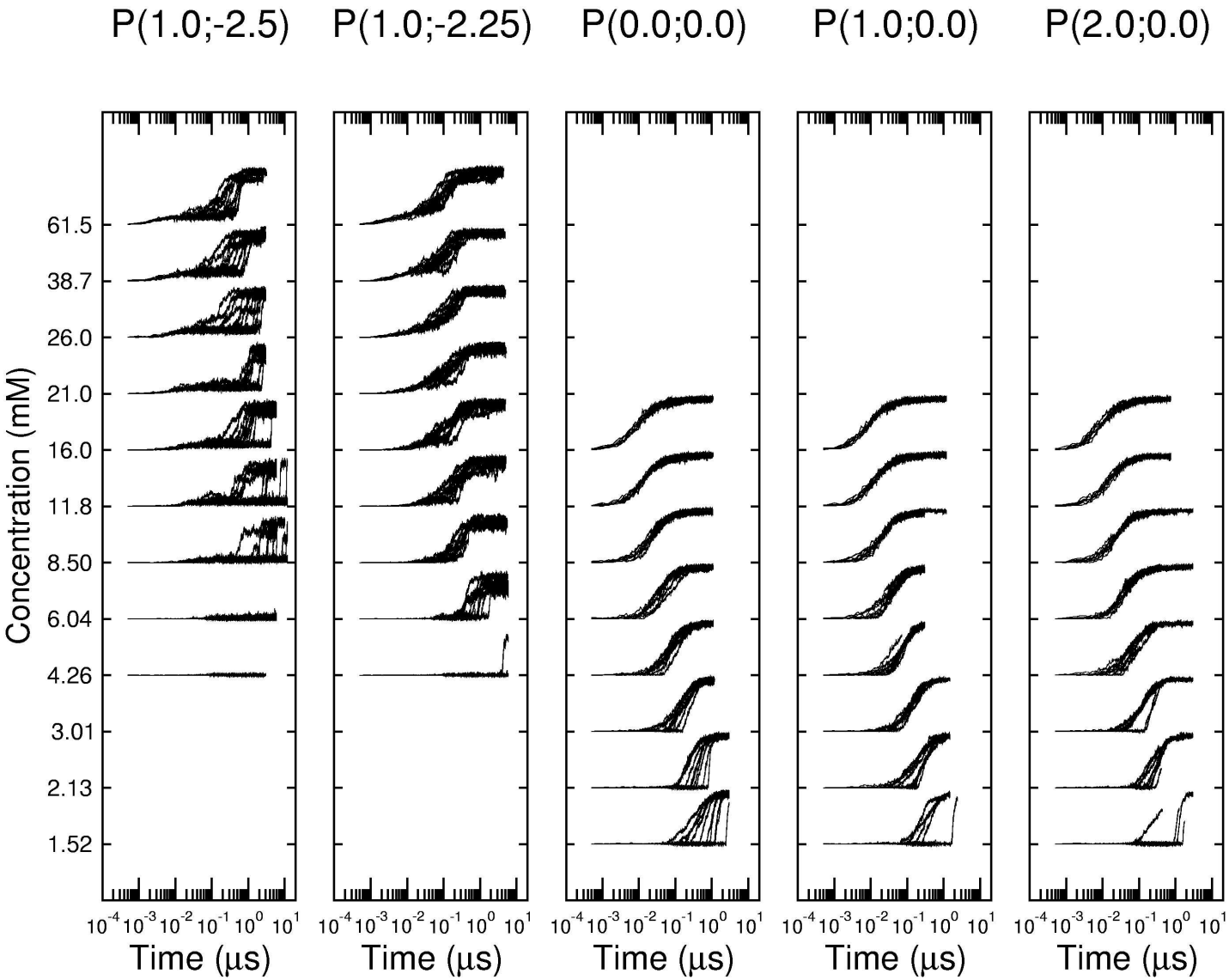

Figure 15: Concentration dependence for the time series of the number of parallel polar contacts n_p normalized to the maximum value. Time series belonging to the same concentration are reported in the same row. The five plots correspond to five different potential models.

Potential	A_{50}	γ_{50}	A_e	γ_e	A^*_{50}	γ^*_{50}
P(0;0)	2.73	-2.52	2.34	1.32	-	-
P(2;0)	1.75	-2.40	2.32	1.29	-	-
P(1;-2.25)	1.93	-0.916	2.24	0.409	0.104	-0.680
P(1;-2.5)	225.8	-1.80	2.45	0.258	0.708	-1.02

Table 6: Resulting fit parameters of the power law regression for the concentration dependence of kinetic observables. The delay time t_{50} and the elongation rate k_e were fitted for C greater that 8.5 mM for the $P(1.0;-2.25)$ potential (see Figure 14.C-D). The values of A^*_{50} and γ^*_{50} were obtained by fitting to the micelle concentration (Figure 14.A).

seed production is irreversible, and (c) the seed precursor is in pre-equilibrium with monomers [22]. These three assumptions are valid for β-stable models where only fibril and monomer species are produced, as demonstrated in Section 3.3. For $P(0;0)$ the nucleus size N^* is about 5, a value close to the one calculated with the probability of fibril formation (see Section 3.11 and Figure 6 of the main text). β-stable models, involving a small nucleus size, share the downhill polymerization mechanism described for the partially denaturated transthyretin [17]. On the other hand, the β-unstable models show a strong cohexistence of micellar and fibrillar oligomers in the lag and the elongation phases (see Section 3.5), therefore the hypothesis (a) cannot be fulfilled. Assuming that the nucleation process is first order to the micelle concentration, one can fit the delay time with a power law $t_{50} = A^*_{50} C_M^{\gamma^*_{50}}$, where C_M is the micelle concentration (see Figure 14.A and Section 3.4 for micelle concentration evaluation). The nucleus size $N^* = -2\gamma^*_{50}$, expressed in micelle units, is 1.36 for $P(1.0;-2.25)$ and 2.04 for $P(1.0;-2.5)$. Given the average aggregation number per micelle of 17.5 at $C = 8.5$ mM (see Figure 10) a value of 23.8 monomers and 35.7 monomers involved in the nucleation is obtained for $P(1.0;-2.25)$ and $P(1.0;-2.5)$, respectively. Strikingly, very similar values are obtained using the probability of fibril formation (see Figure 6 of the main text).

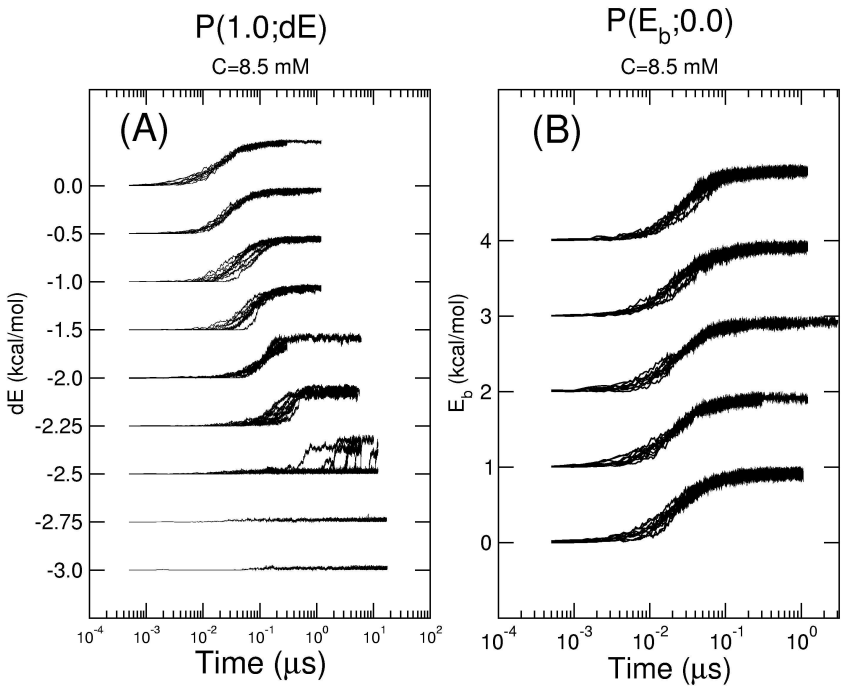

Figure 16: Stability, i.e., π-β free energy difference (A) and barrier (B) influence on the time series of the number of parallel polar contacts n_p normalized to the maximum value which corresponds to fibril. In plot (A) each row corresponds to a different value of dE (with constant E_b of 1.0 kcal/mol) while in plot (B) each row corresponds to a different value of E_b with constant $dE = 0$ kcal/mol.

3.8 Monomer energy landscape influence on the kinetics of fibril formation

To monitor the effects of the monomer energy surface a series of runs were performed at the concentration of 8.5mM. The left plot of Figure 16 displays the change of kinetics upon variation of the stability of the β-state without changing the $\beta \rightarrow \pi$ barrier E_b. The resulting rates and lag phase times are reported in Figure 2 of the main text. The right plot of Figure 16 shows the effects of variation of the barrier, keeping constant the stability of the β-state (see also Figure 14.B). No appreciable trend for the elongation and nucleation kinetics is observed for different values of E_b. These simulation results indicate that the aggregation kinetics of the model are mainly influenced by the relative stability of the β-aggregation prone state with negligible contribution of the $\beta \rightarrow \pi$ barrier.

3.9 Pathways of oligomeric aggregation

Since individual oligomers change their composition of monomers (i.e., in a given time interval an oligomer can absorb or release monomers to the solvent), it is crucial to define criteria for identifying oligomers along the simulation. Given the converged contact matrix at time t, $d^{(\infty)}(t)$, all oligomers at time t can be labeled: $A_1^t, A_2^t, ..., A_{n_t}^t$, where n_t is the number of oligomers. Each oligomer A_k^t has a size N_k^t and contains a list of tagged monomers $m_k^{1,t}, m_k^{2,t}, ..., m_k^{N_k^t,t}$. The time evolution of a single oligomer A_k^t at time $t+\tau$ is evaluated by comparing the monomer composition of every single oligomer present at time $t+\tau$. The similarity between two oligomers is defined as

$$S(A_k^t, A_l^{t+\tau}) = \sum_{i=1,N_k^t} \sum_{j=1,N_l^{t+\tau}} \delta(m_k^{i,t}, m_l^{j,t+\tau}) \tag{16}$$

where δ is the Kronecker function which is 1 if the compared monomers are the same. The time evolution $A_{k'}^{t+\tau}$ of oligomer A_k^t is defined as the oligomer with highest similarity:

$$S(A_k^t, A_{k'}^{t+\tau}) = \max_{l=1,n_{t+\tau}} (S(A_k^t, A_l^{t+\tau})) \tag{17}$$

If two or more $A_{k'}^{t+\tau}$ fulfill this equation, then the first labeled oligomer is chosen. A_k^t is then forwardly linked to $A_{k'}^{t+\tau}$, or equivalently $A_{k'}^{t+\tau}$ is assigned to the temporal successor of A_k^t. It is worth noting that:

1) *the temporal successor of an oligomer at time t is the oligomer at time* $t+\tau$ *that shares the highest number of monomers;*

2) *each oligomer can have a single temporal successor;*

3) *many oligomers can be forwardly linked to the same successor.*

By iterating this procedure, one can build the pathway of individual oligomers. Thus, the simulation trajectory is mapped to a network of temporally linked nodes (oligomers).

One natural definition for τ is the time difference between frames of the sim-

ulation, in our case $\tau = 0.5ns$, so that τ is smaller than the average life time of oligomers. The life time of an oligomer is the time needed to completely recycle monomers or to dissolve the oligomer. If this requirement for the time τ is not fulfilled the similarity can be ill-defined. Fibrils have a life time that is by far larger than τ being in the microsecond timescale (see Section 4.1). Metastable oligomers such as disordered aggregates or micelles have an estimated life time in the ten-nanoseconds scale. Unstable oligomers created by occasional collision of monomers have a life time slightly larger than τ.

3.10 β-subdomains time evolution and nucleus definition.

An important issue is the quantitative characterization of the nucleus. In literature, a nucleus is often defined as *the smallest marginally stable structured aggregate* [6]. In the framework of MD simulations a useful definition would be *the oligomer that has the same probability to either progress to a fibril or regress to the disordered state*. An analogous definition was applied to the folding transition state ensemble of two-state folders [23]. Using only the aggregation number N to distinguish the oligomers, this definition is problematic since an oligomer of given size can have a high morphological heterogeneity. It can contain no ordered aggregates, one subdomain ordered, two disjoint subdomains, and even more complicated features. A β-subdomain is a portion of the oligomer made of interacting β-monomers and the surrounding π-monomers can be considered as a local perturbation on the β-domains. With the previous nucleus definition it is possible to follow the dynamic evolution of β-subdomains in the context of their constituting oligomers. As a consequence the similarity procedure explained in Section 3.9 is applied to β-monomers only to identify the pathways of β-subdomains.

We define as progress variable of the ordered polymerization process, the aggregation number of the β-subdomain:

$$N(A_k^t) = N_k^t \tag{18}$$

where now A_k^t is an oligomer containing only β-monomers. A single trajectory is a collection of many independent pathways for the labeled β-subdomains (Figure 17). The pathway (A) is an unproductive event; first a β region appears in a π-only oligomer, it persists for a period of time, and disappears. (B) is a productive event; the appeared β-region spreads irreversibly into a fibril. Oligomers and their β-subdomains can interact in many different ways. Pathway (C) is a merging event while in (D) a β-region splitting is depicted. Pathway (E) is a combination of merging and division, while (F) is an interaction between oligomers that does not involve their β-subdomains.

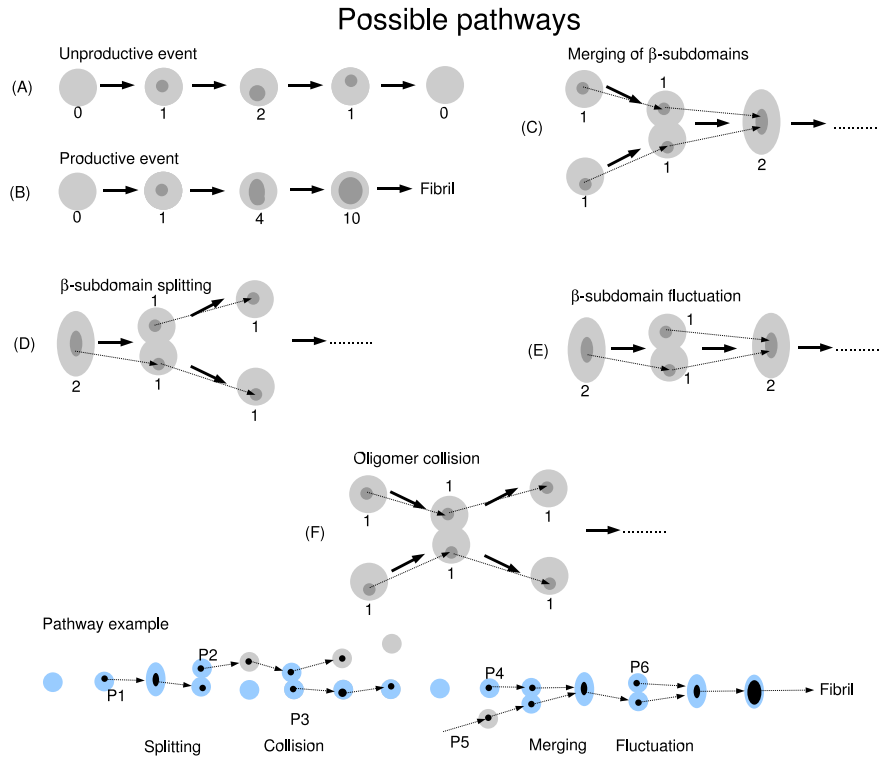

Figure 17: Classification of different pathways. The light gray regions schematize an oligomer, while the dark gray regions are the included β-subdomains. Bold arrows indicate the time evolution of entire oligomers, while the thin arrows are the time evolution of β-subdomains as defined by the similarity procedure (Equations 16 and 17). In the pathway example at the bottom, the dynamic evolution of an entire cluster (blue circles) and its β-subdomains (black dots) is depicted. Gray shaded regions are clusters interacting with the blue one.

3.11 Pathways analysis and probability of fibril formation

It is convenient to include an abstract state A_0, whose aggregation number is zero, to describe the beginning or the end of β-subdomain pathways. Another important state is the fibril state A_F; it is defined as the β-subdomain with aggregation number greater than 60. Given A_0 and A_F, an unproductive pathway is the trajectory of a β-subdomain that starts from A_0, and returns back to A_0, while a productive pathway is the trajectory of a β-subdomain that starts from A_0 and ends with the fibril state A_F. In the example of Figure 17, $P1$, $P2$ and $P3$ are unproductive while $P4$, $P5$ and $P6$ are productive. Given a set of trajectories, one can collect all productive and unproductive pathways P_i and define the probability of fibril formation of a β-subdomain A_{N_β} of size N_β as

$$p_{Ff}(N_\beta) = \frac{1}{M(A_{N_\beta})} \sum_{P_i \ni A_{N_\beta}} F(P_i) \tag{19}$$

where $M(A_{N_\beta})$ is the number of times that an aggregate of size N_β occurred in the simulations set. The sum runs over all P_is pathways that contain an aggregate A_{N_β}, and $F(P_i)$ is equal to 1 if the pathway P_i is productive, and is 0 otherwise. In this way the nucleus is unequivocally defined as *the oligomer containing a β-subdomain of size N^*_β with a probability of fibril formation $p_{Ff}(N^*_\beta)$ equal to* 0.5.

To finally characterize the nucleus it is useful to calculate its total aggregation number N^*. The average number of β-monomers as a function of the oligomer size N can be obtained by taking the exponential function on both side of Equation 15:

$$\frac{N_\beta(N)}{N_\pi(N)} = \exp\left[\frac{\Delta G_{\beta\pi}(N)}{kT}\right] \tag{20}$$

Using the fact that $N_\pi(N) = N - N_\beta(N)$ one has

$$N_\beta(N) = \frac{N \exp\left[kT^{-1}\Delta G_{\beta\pi}(N)\right]}{1 + \exp\left[kT^{-1}\Delta G_{\beta\pi}(N)\right]} \tag{21}$$

the average aggregation number N of an oligomer containing a β-subdomain with size N_β is obtained by numerical inversion of the function $N_\beta(N)$, $N = N(N_\beta)$ (Figure 6 of the main text), and the nucleus aggregation number is $N^* = N(N_\beta^*)$.

4 Supplementary analysis

4.1 Molecular Recycling

A molecular recycling mechanism has been observed by a combination of NMR spectroscopy and mass spectroscopy for an amyloid fibril formed from an SH3 domain [24]. To evaluate the recycling time of the coarse-grained model, simulations of mature fibrils in equilibrium with dispersed monomers are analyzed for the potential $P(1.0; -2.5)$ at all concentration values. The number of the unrecycled monomers $N_u(t)$ is defined as follows. First, all monomers belonging to the fibril at time $t = 0$ are labeled and counted. Then, at all times $t > 0$ the monomers that never detached from the fibril are counted and the resulting number is $N_u(t)$. In two of nine simulations, $N_u(t)$ goes to zero within 4 μs, which shows that all monomers initially belonging to the fibril have been recycled (Figure 18).

The number of unrecycled monomers $N_u(t)$ can be fitted with an exponential function:

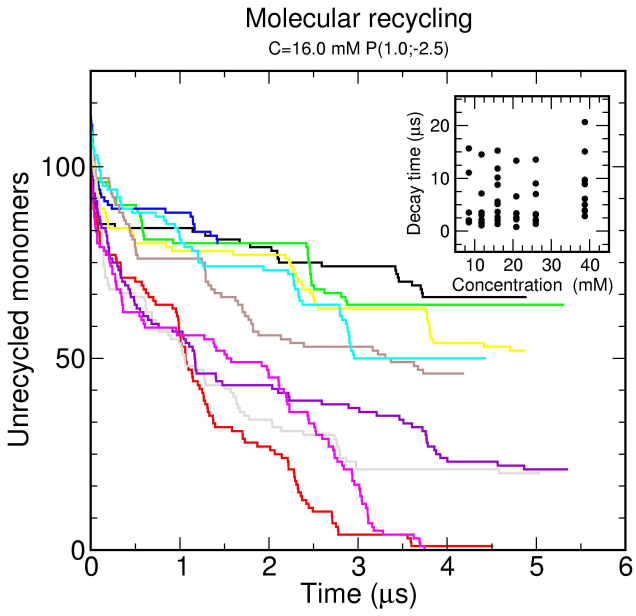

Figure 18: Number of unrecycled monomers N_u as a function of time for nine simulations started from a preformed equilibrated fibril. Simulations were performed at total concentration $C = 16.0$ mM and for the potential $P(1.0; -2.5)$. The values of decay time τ at different concentrations are reported in the inset.

$$N_u(t) = N_u(0)e^{-t/\tau} \tag{22}$$

where $N_u(0)$ is the initial value of monomers belonging to the fibril and τ is the time for the decay. In the inset of Figure 18, τ values for all concentrations and all independent simulations are reported. The decay times do not depend on the total concentration at which the fibril was formed.

4.2 Seeding

Fibril formation generally occurs via nucleation-dependent oligomerization with a lag time required for nucleus formation. This lag time can be abolished by using a seed, i.e., a small preformed fibril. To further evaluate the coarse-grained model and to validate the nucleus definition of Section 3.10, a seeding experiment is performed in *silico*. For the potential $P(1.0; -2.5)$ the smallest oligomer with high

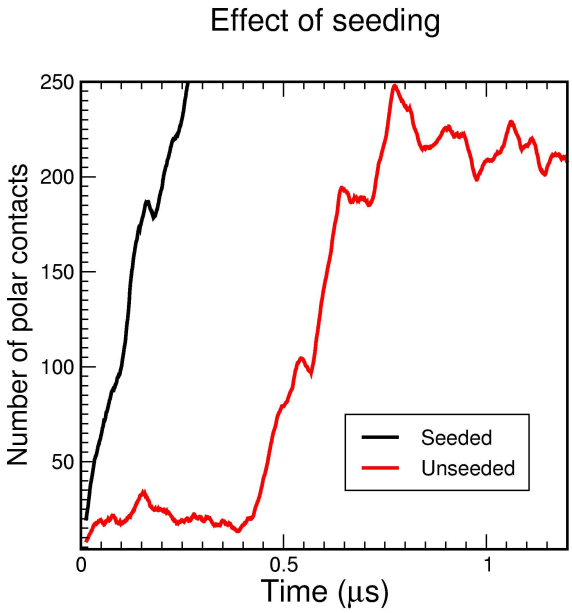

Figure 19: Number of polar contacts n_p along the time for a spontaneous (red line) and seeded trajectory both at 8.5 mM and $P(1.0; -2.5)$ potential. The unseeded simulation shown here is the fastest nucleation observed at this concentration while nucleation in the other runs are about one order of magnitude slower (see Figure 16 left). The spontaneous trajectory reaches a plateau at about 225 polar contacts because it was run with 125 monomers, whereas the seeded trajectory was run with a total of 1000 monomers.

probability of fibril formation is isolated from a trajectory at 38.7 mM. This post-critical oligomer consists of 60 monomers and has a probability of fibril formation of 98% according to the definition of Section 3.10 and to Figure 4 of the main text. The oligomer is introduced in a box with 940 dispersed monomers at a total concentration of 8.5 mM. This is the lowest concentration that displayed a nucleation for this potential (see Figure 15 and Table 5). The average lag phase for the spontaneous nucleation is around 5 μs (see Figure 14.C) and the minimal lag phase time observed is 0.4 μs (Figure 19). Strikingly, the lag phase is completely abolished in the seeded simulation (Figures 19 and 20).

On the other hand for the $P(0;0)$ potential seeding does not influence the kinetics (data not shown), as expected for downhill polymerization [17].

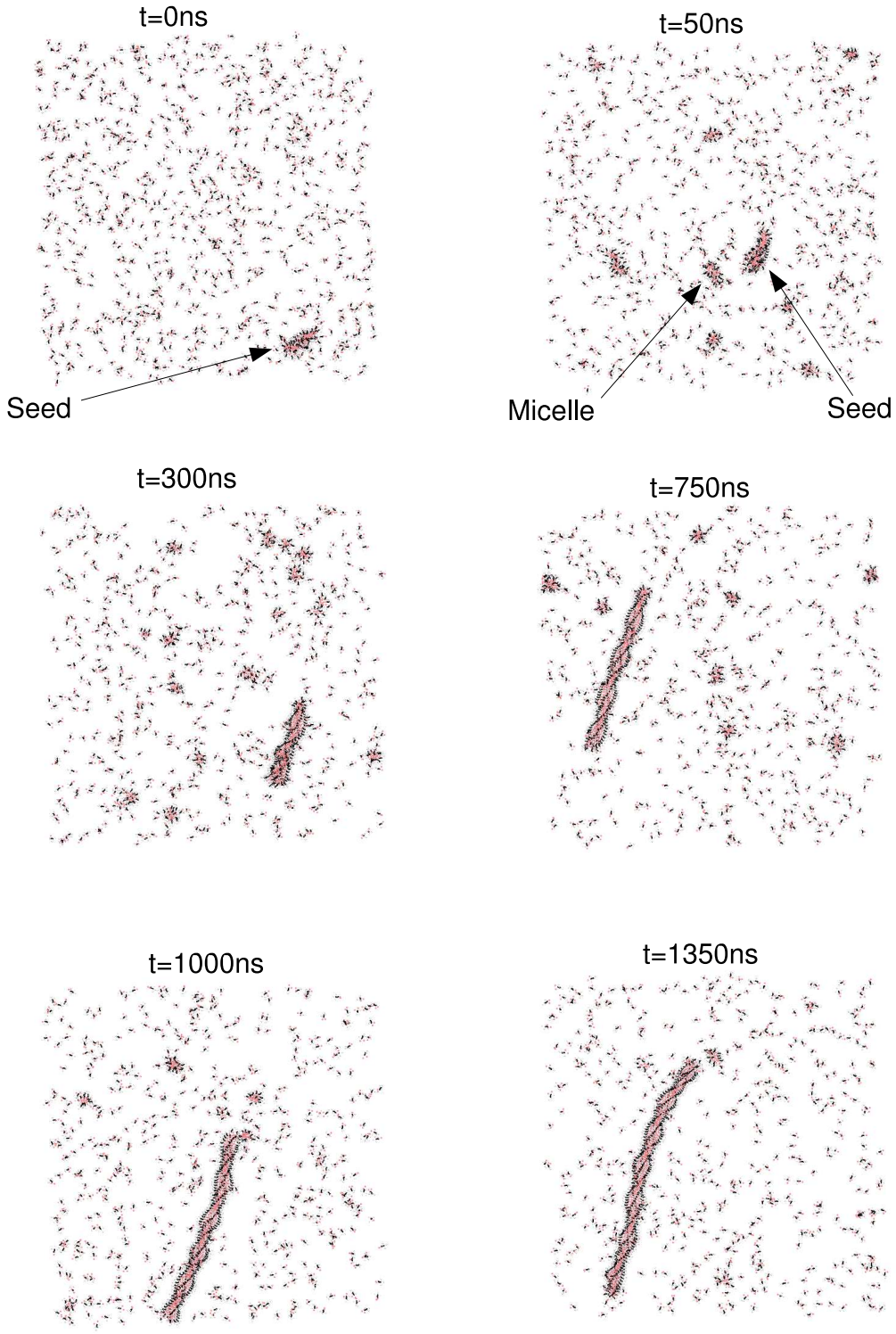

Figure 20: Six illustrative snapshots of the seeding trajectory. At the start ($t = 0$ ns) there are 60 monomers in the post-critical oligomer (i.e. the seed) and 940 monodispersed monomers. The total concentration is 8.5 mM. Within the first 50 ns micelles are nucleated, and progressively disappear during the fibril elongation phase.

References

1. MacKerell, A. D. J., Feig, M. & Brooks, C. L. r. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* **126**, 698–699 (2004).

2. Hansen, J.-P. & McDonald, I. R. *Theory of simple liquids* (Academic Press, San Diego, 1996).

3. Soreghan, B., Kosmoski, J. & Glabe, C. Surfactant properties of Alzheimer's A beta peptides and the mechanism of amyloid aggregation. *J Biol Chem* **269**, 28551–28554 (1994).

4. Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* **289**, 1317–1321 (2000).

5. Yong, W. *et al.* Structure determination of micelle-like intermediates in amyloid beta -protein fibril assembly by using small angle neutron scattering. *Proc Natl Acad Sci U S A* **99**, 150–154 (2002).

6. Rhoades, E. & Gafni, A. Micelle formation by a fragment of human islet amyloid polypeptide. *Biophys J* **84**, 3480–3487 (2003).

7. Bitan, G. *et al.* Amyloid beta -protein (Abeta) assembly: Abeta 40 and Abeta 42 oligomerize through distinct pathways. *Proc Natl Acad Sci U S A* **100**, 330–335 (2003).

8. Sabatè, R. & Estelrich, J. Evidence of the existence of micelles in the fibrillogenesis of β-amyloid peptide. *J Phys Chem* **109**, 11027–11032 (2005).

9. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equation of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341 (1977).

10. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

11. Aggeli, A. *et al.* Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta -sheet tapes, ribbons, fibrils, and fibers. *Proc Natl Acad Sci U S A* **98**, 11857–11862 (2001).

12. Goldsbury, C. S. *et al.* Polymorphic fibrillar assembly of human amylin. *J Struct Biol* **119**, 17–27 (1997).

13. Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. *Science* **307**, 262–265 (2005).

14. Chiti, F. *et al.* Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J* **19**, 1441–1449 (2000).

15. Chiti, F. *et al.* Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF N-terminal domain. *Protein Sci* **10**, 2541–2547 (2001).

16. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329–332 (1999).

17. Hurshman, A. R., White, J. T., Powers, E. T. & Kelly, J. W. Transthyretin aggregation under partially denaturing conditions is a downhill polymerization. *Biochemistry* **43**, 7365–7381 (2004).

18. Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).

19. O'Nuallain, B., Shivaprasad, S., Kheterpal, I. & Wetzel, R. Thermodynamics of A beta(1-40) amyloid fibril elongation. *Biochemistry* **44**, 12709–12718 (2005).

20. Esler, W. P. *et al.* Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry* **39**, 6288–6295 (2000).

21. Collins, S. R., Douglass, A., Vale, R. D. & Weissman, J. S. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol* **2**, e321 (2004).

22. Goldstein, R. F. & Stryer, L. Cooperative polymerization reactions. Analytical approximations, numerical examples, and experimental strategy. *Biophys J* **50**, 583–599 (1986).

23. Du, R., Pande, V. S., Grosberg, A. Y., Tanaka, T. & Shakhnovich, E. I. On the transition coordinate for protein folding. *J. Chem. Phys.* **108**, 334–350 (1998).

24. Carulla, N. *et al.* Molecular recycling within amyloid fibrils. *Nature* **436**, 554–558 (2005).