Amyloid fibril polymorphism is under kinetic control

Supplementary Information

Riccardo Pellarin Philipp Schütz Enrico Guarnera
Amedeo Caflisch*

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

Contents

1 Progress variables

Three progress variables are used to monitor the aggregation process: the size of the largest aggregate N_{la}, the number of monomers in the β-state within the largest aggregate N_{la}^{β}, and the number of protofilaments in the largest aggregate N_{la}^{pf}. N_{la} is defined as:

$$N_{la}(t) = \max_{I} N_I(t)$$

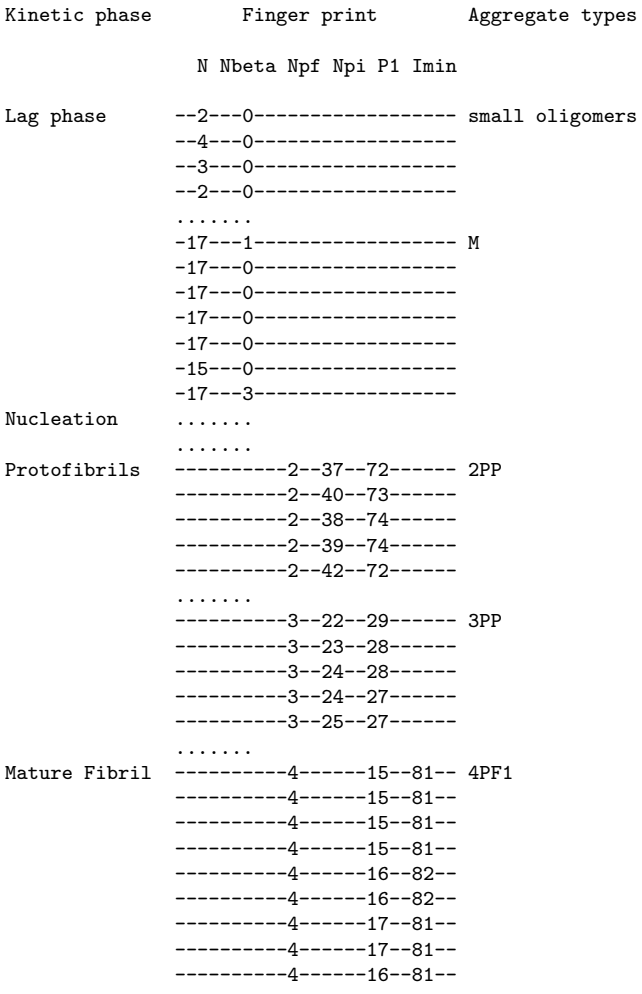
where N_I is the aggregation number of aggregate I present at time t in the simulation box. Note that the range of N_{la} is limited by the size of the simulated system ($1 \leq N_{la} \leq 125$). The number of π monomers within the largest aggregate N_{la}^{π} is the difference between N_{la} and N_{la}^{β}. The number of protofilaments within a single aggregate is calculated by counting the files of monomers in the β-state with intermolecular dipolar interactions. Let N_f be the number of such files present into a given aggregate, and $\omega_1, \ldots, \omega_{N_f}$ the number of monomers in each file (with $\omega_i > 10$ to reduce noise). The number of protofilaments in aggregate a, N_a^{pf}, is thus defined as:

$$N_a^{pf} = \frac{\left(\sum_{i=1}^{N_f} \omega_i\right)^2}{\sum_{i=1}^{N_f} \omega_i^2} \tag{1}$$

This definition prevents from counting small isolated files whose formation is a result of thermal fluctuations, enhancing the signal to noise ratio with respect to N_f. Two limiting cases are useful to explain this variable. In the case that all files have the same size (i.e., $\omega_1 = ... = \omega_{N_f}$), the protofilament number N_a^{pf} is equal to the number of files N_f. In the case where a single ω_i predominates ($\omega_i >> \omega_k$ for all k different from i) N_a^{pf} tends to 1. The number of protofilaments in the largest aggregate N_{la}^{pf} is thus the function N_a^{pf} applied to the largest of all aggregates present in the simulation volume.

2 Aggregation state fingerprints

The number of protofilaments N_{pf} determines the entries of the fingerprint: N_{pf}^{la} =0 or 1, (N^{la}, N_{β}^{la}); N_{pf}^{la} =2 or 3, (N_{pf}^{la}, N_{π}, P_1); $N_{pf}^{la} \geq 4$, (N_{pf}^{la}, P_1, I_{min}). Example of fingerprint timeseries (dE=-2.5) with time from top to bottom. The values of P_1 and I_{min} are multiplied by 100 and 50, respectively, and discretized.



3 Stability of morphologies.

dE [kcal/mol]	Morphology	C_r [mM]	Δg [kcal/mol]
-2.5	4PF1	2.61 ± 0.12	-3.66
	4PF2-	2.05 ± 0.17	-3.81
	4PF2+	1.59 ± 0.14	-3.96
-2.25	4PF1	1.59 ± 0.15	-3.96
	4PF2-	1.21 ± 0.11	-4.13
	4PF2+	0.98 ± 0.08	-4.26
-2.0	4PF1	1.10 ± 0.09	-4.19
	4PF2-	0.96 ± 0.13	-4.28
	4PF2+	0.77 ± 0.14	-4.41
-1.5	4PF1	0.46 ± 0.07	-4.73
	4PF2-	0.42 ± 0.04	-4.79
	4PF2+	0.34 ± 0.06	-4.92

Table S1

Stability of morphologies calculated at different dE values. Morphology dependent critical concentration C_r and free energy of monomer association Δg are reported for four values of the amyloidogenic potential dE.

4 Simple examples to illustrate basin-isolation procedure.

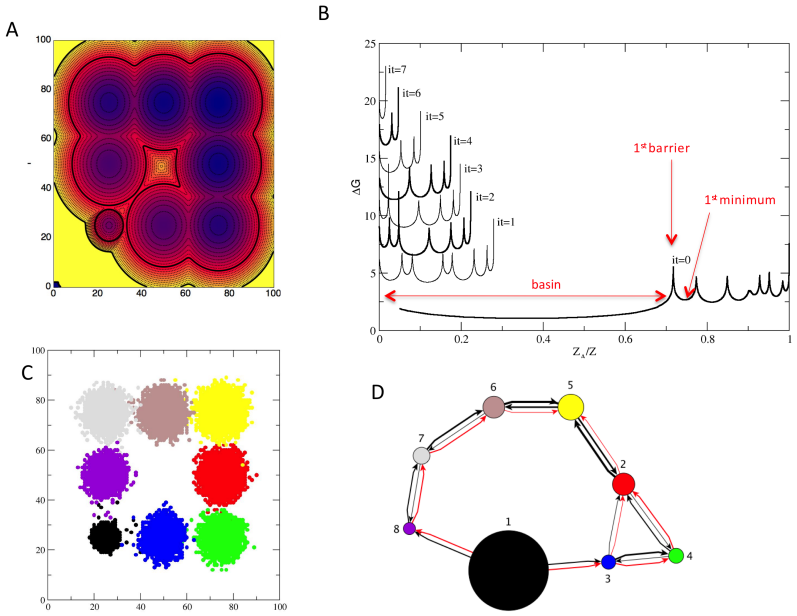


Fig. S1. **Simplified graph generated from a 2D-potential, example 1.** (A) An energy potential with 8 wells is created on a 100 × 100 grid of x and y values (every isoenergetic level corresponds to 1 kcal/mol). A total of 100 trajectories are run using the Metropolis Monte Carlo algorithm at 310 K starting from the x=25 and y=25 state (belonging to the bottom left basin). The simulations are stopped after a fixed amount of Monte Carlo iterations that doesn't allow a reversible exploration of the complete landscape to simulate an out of equilibrium process. From left to right and from bottom to top, the basins are increasingly more stable, and are separated by potential energy barriers of 3-4 kcal/mol, with the exception of the starting basin, that has a high exiting barrier (10 kcal/mol), to mimic a rare nucleation process followed by a downward exploration of basins, as in the case of fibril formation. The potential is shaped to have two parallel pathways from the starting basin to the most stable basin (top right). (B) The time series of the x-y values are coarse-grained for the mfpt cFEP calculation, and the iterative procedure of free energy basin detection described in the main text is applied. At iteration 0, the cFEP is calculated from the most populated x-y state. All the states with Z_A/Z lower than the value of the first barrier are assigned to basin 1 and removed. The new reference state is the first minimum after the first barrier. At every iteration the previously defined basin is removed, resulting in a reduced cFEP (the cFEPs are shifted upward along the y-axis to improve readability). (C) To test whether the state-basin assignment is correct, the x-y states sampled by the trajectories are colored according to the basin they belong to. The resulting plot reflects the original potential energy and the correct partitioning of states into basins. (D) The simplified graph of the inter-basin transitions is constructed based on the original x-y state trajectories. The area of the node is proportional to the statistical weight of the basin. The oriented black links correspond to the back and forward transitions, the red link is the difference between back and forward transition. The thickness of the links is proportional to the number of transitions. The topology of the simplified graph correctly describes the two parallel pathways, that start from basin 1 and end in basin 5. The statistical weight of the nodes is not proportional to the actual potential energy of the basins, as the trajectories are out of equilibrium. In fact, the most visited node is the basin 1, as it is the starting state and it has a high exiting barrier.

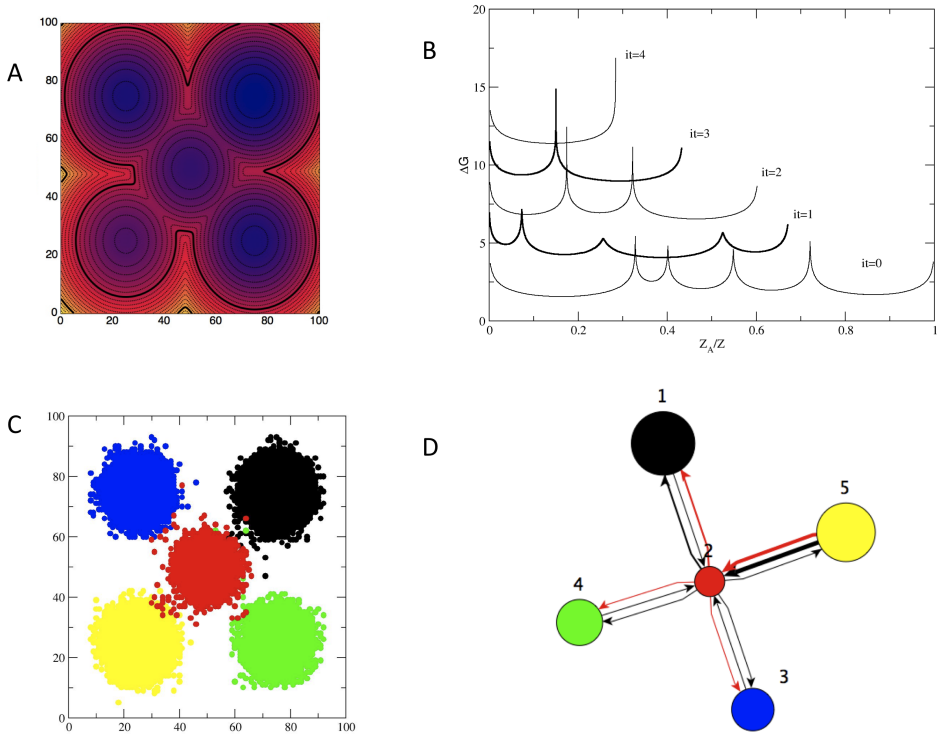


Fig. S2. **Simplified graph generated from a 2D-potential, example 2.** (A) In this second example the 2D potential has 5 wells arranged in a cross shape. From left to right and from bottom to top, the basins are increasingly more stable, and the barriers are about 5 kcal/mol. A total of 100 trajectories are run using the Metropolis Monte Carlo algorithm at 310 K from the starting state (x=25 and y=25, belonging to the bottom left basin). (B) The iterative procedure of free energy basin detection is applied. (C) The explored states are colored according to the mfpt cFEP basins they belong to. (D) The simplified graph of the inter-basin transitions. The area of the node is proportional to the statistical weight of the basin. The oriented black links correspond to the back and forward transitions, the red link is the difference between back and forward transition. The reaction starts from basin 5 and proceeds through basin 2 (the central well of the potential) to basin 1. Basin 1 is the most populated and is isolated first by the iterative procedure.

5 Structural characterization of fibril morphologies.

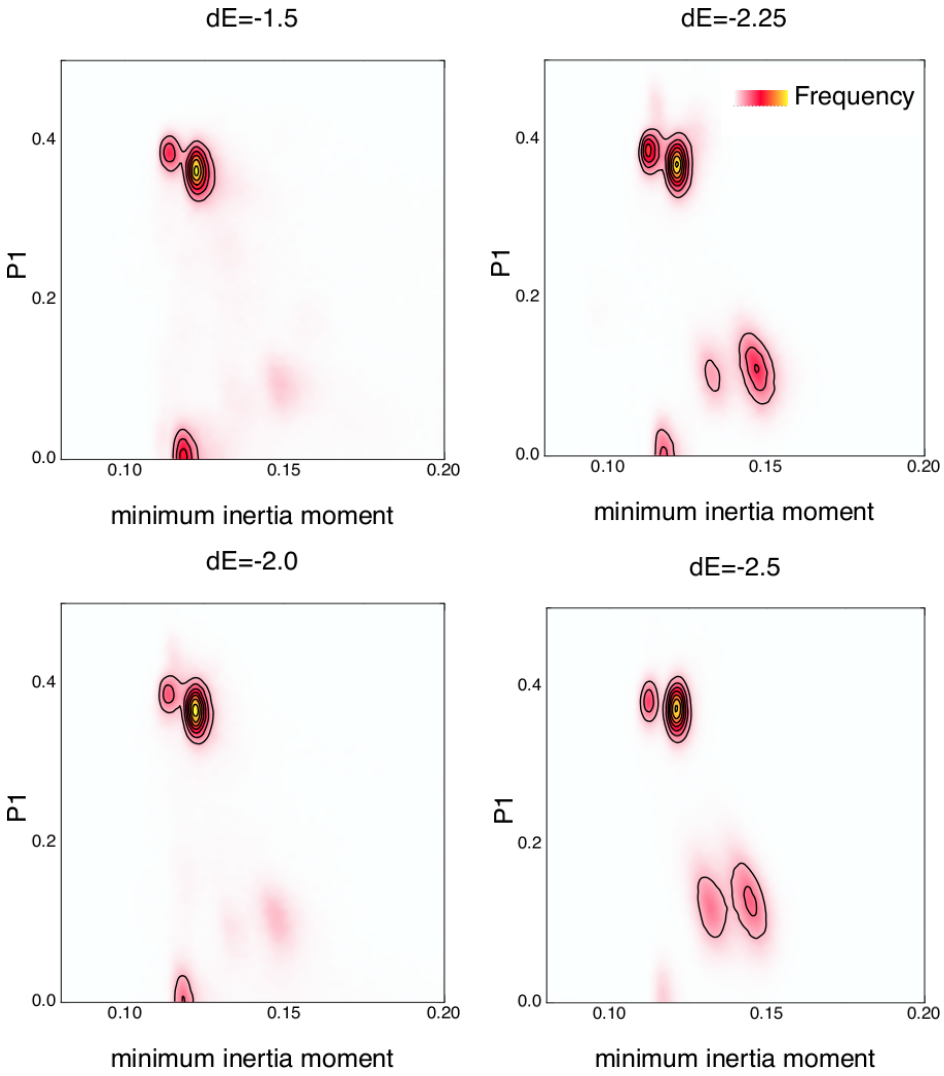


Fig. S3. **2D histograms of P_1 and I_{min}.** Two-dimensional frequency histograms of the order parameter P_1 and the minimum inertia moment calculated at four dE values. Highest frequency is colored with yellow.

6 Interconversion between 4PF1+ and 4PF1-.

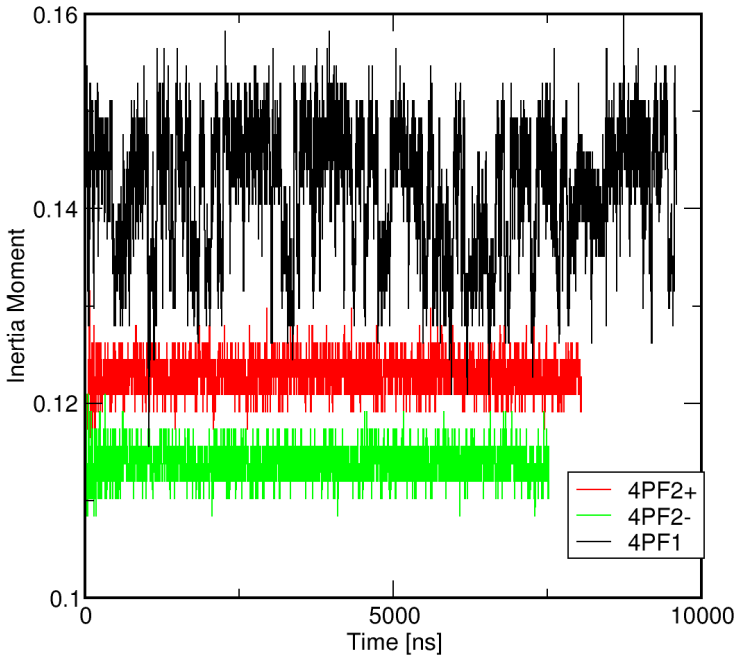


Fig. S4. **Time series of the minimum inertia moment for three different morphologies.** The time series of the minimum inertia moment of mature fibrils of different morphology (4PF1, 4PF2+ and 4PF2-) are compared. The value for 4PF1 fluctuates between two values (corresponding to 4PF1+ and 4PF1- subpopulations), while I_{min} for 4PF2+ and 4PF2- are stable around the 0.125 and 0.115 values, respectively.

7 cFEPs for 4PF2+ and 4PF2- morphologies.

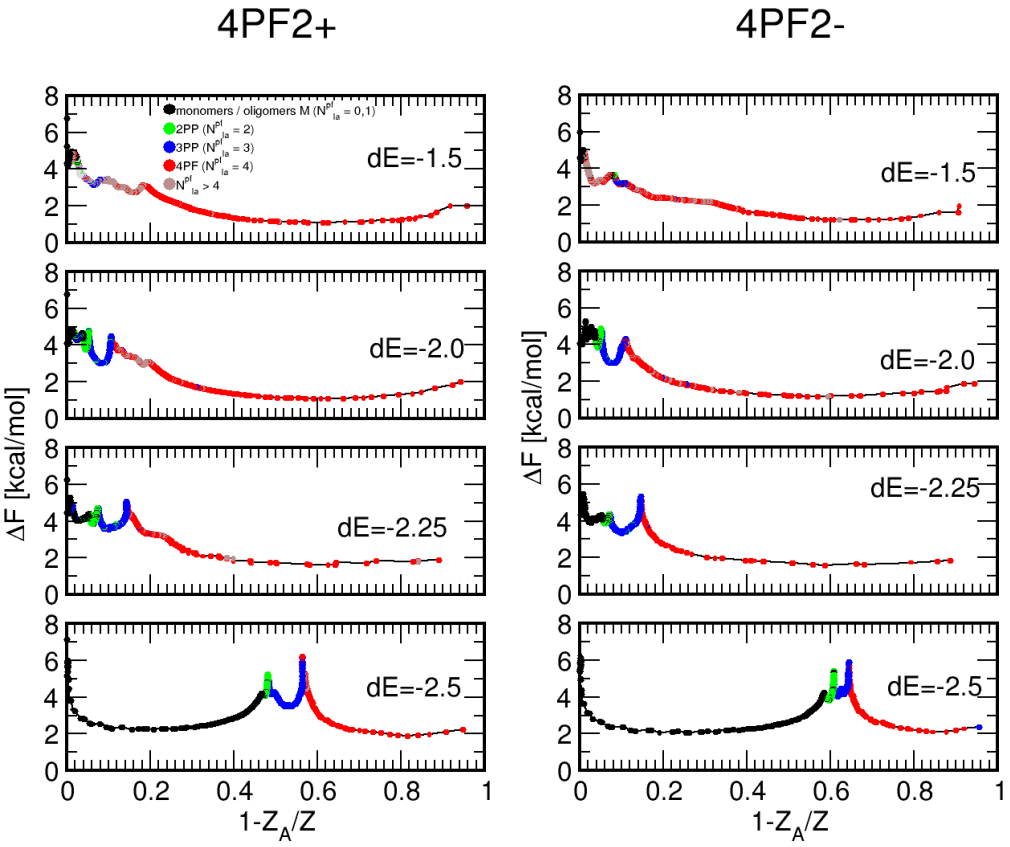


Fig. S5. cFEP calculated for 4PF2+ and 4PF2- morphology at four different dE values. The reference state is the most populated 4PF state. Every state is represented as a circle, whose color reflects the number of protofilaments. Note that the x axis is inverted with respect to the conventional representation of cFEP, having the reference state on the right of the plot.

8 Simplified networks of interbasin transitions.

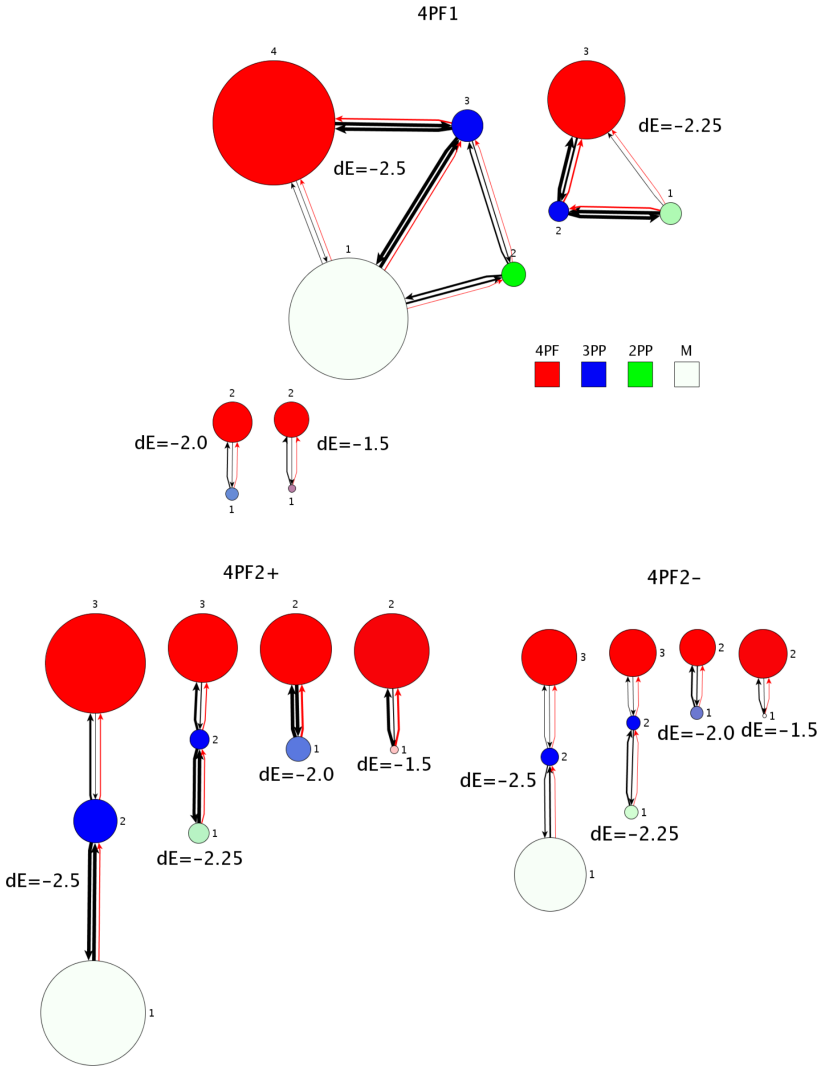


Fig. S6. **Simplified networks of interbasin transitions.** Simplified graphs describing the process of fibrillation for different potentials and different morphologies. The size of the nodes is proportional to the statistical weight of the corresponding cFEP basin, while the thickness of the links is proportional to the number of transitions. The red link is the difference between the forward and backward black links. The color of the nodes is obtained by mixing four colors: red, blue, green and white using the statistical weights of the 4PF, 3PP, 2PP and M states that populate the basin. The node number one is the basin where the simulations starts.

9 3PP protofibrils polymorphism.

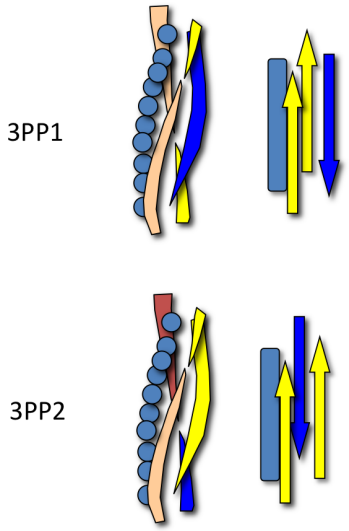


Fig. S7. **3PP protofibrils polymorphism.** By visual inspection of the trajectories performed at dE=-2.5, two different 3PP intermediates are observed. 3PP1 and 3PP2 are competent to 4PF1 and 4PF2(+,-) morphologies, respectively. They differ in the arrangement of the up and down protofilaments with respect to the deposit of π monomers (represented by blue beads), thus they are indistinguishable by using the parameter P_1. No apparent difference is observed between the protofibrils that generated 4PF2+ and those that produced 4PF2-.

10 Chemical master equation of the fibril elongation.

To investigate the behavior of the system beyond the size and time limit of the Langevin dynamics simulations, a fibril elongation master equation was developed, based on the nucleated polymerization equation introduced by Goldstein and Stryer [1]. The dynamic variables of the equations are the concentration of dissolved monomers m and the concentration of polymers $A_n^{(I)}$ with size $n > 2$ and morphology index I=4PF1, 4PF2+, 4PF2-. The model describes the growing phase of the fibrils with dE=-2.5 after the nucleation phase and incorporate the competitive growth effects. The model is based on the following assumptions:

(1) nucleation of fibrils is neglected.

(2) polymers with subnuclear size $n < 30$ are unstable, i.e. they cannot grow (this threshold value is chosen according to the measure of stability of fibrils under membrane-assisted degradation [2]).

(3) the growth occurs only by monomer addition, thus the self-assembly and breakage of fibrils is neglected.

(4) fibril intermediates (protofibrils, oligomers) are not taken into account.

(5) the rate of addition of a monomer on a polymer is independent of its size and morphology.

The equation for the concentration of monomers can be written as:

$$\frac{dm(t)}{dt} = -k_{on} m(t) \sum_I F^{(I)}(t) + \sum_I k_{off}^{(I)} P^{(I)}(t)$$

Where the k_{on} is the rate of addition of monomers on the aggregates, $F^{(I)}(t) = \sum_{n\geq 30} A_n^{(I)}(t)$ is the total concentration of polymers with morphology (I) and size $n \geq 30$, and $P^{(I)}(t) = \sum_{n\geq 2} A_n^{(I)}(t)$ is the total concentration of polymers with morphology (I) and size $n \geq 2$.

The equations for the concentration of aggregates with $n > 2$ and morphology I are

$$\frac{dA_n^{(I)}(t)}{dt} = k_{off}^{(I)} \left[-dA_n^{(I)}(t) + dA_{n+1}^{(I)}(t) \right] \quad 2 \leq n \leq 30$$

[1] R.F. Goldstein, L. Stryer, (1986) *Biophys. J.* **50** p. 583
[2] R. Friedman, R. Pellarin, A. Caflisch, (2010) *J. Phys. Chem. Lett.* **1** p. 471

$$\frac{dA_{30}^{(I)}(t)}{dt} = k_{off}^{(I)}\left[-dA_{30}^{(I)}(t) + dA_{31}^{(I)}(t)\right] + k_{on}m(t)A_{30}^{(I)}(t) \quad n = 30$$

$$\frac{dA_{n}^{(I)}(t)}{dt} = k_{off}^{(I)}\left[-dA_{n}^{(I)}(t) + dA_{n+1}^{(I)}(t)\right] + k_{on}m(t)\left[A_{n-1}^{(I)}(t) - A_{n}^{(I)}(t)\right] \quad n > 30$$

The rate constants of monomer addition k_{on} and unbinding $k_{off}^{(I)}$ are evaluated from the simulations. For dE=-2.5, the rate of monomer addition k_{fibril} is 0.012 ns^{-1} ([3]), which divided by the average concentration of monomers in equilibrium with the fibrils $C_F^r = 2.5mM$ ([4]) yields the rate constant $k_{on} = 4.8 \cdot 10^{-3} ns^{-1} mM^{-1}$. The morphology dependent rates of monomer unbinding are obtained from the dissociation constants $K_d^{(I)}$ (reported as concentration C_r in Tab. S1), and $k_{off}^{(I)}$ can be obtained as $k_{on} \cdot K_d^{(I)}$, yielding for 4PF1, 4PF2- and 4PF2+ the values 0.012, 0.0098 and 0.0075 ns^{-1}, respectively.

The initial concentrations $A_n^{(I)}(0)$ and $m(0)$ were derived from the final part of the 100 nucleation simulations (see Fig. S8). The master equations are numerically integrated using the Heun's predictor corrector method.

The time-dependent numerical solutions of the equations are reported in Fig. S8 for dE=-2.5 potential, and show that given the initial mixture of fibril morphologies, obtained from the Langevin simulations, only the 4PF2+ (i.e., the morphology with the lowest K_d) remains at the end. Nevertheless, there are long kinetic phases(40-150 μs) where the three or two different morphologies coexist.

[3] R. Pellarin, E. Guarnera, A. Caflisch, (2007) *J. Mol. Biol.* **374** p. 917
[4] R. Pellarin, A. Caflisch, (2006) *J. Mol. Biol.* **360** p. 882

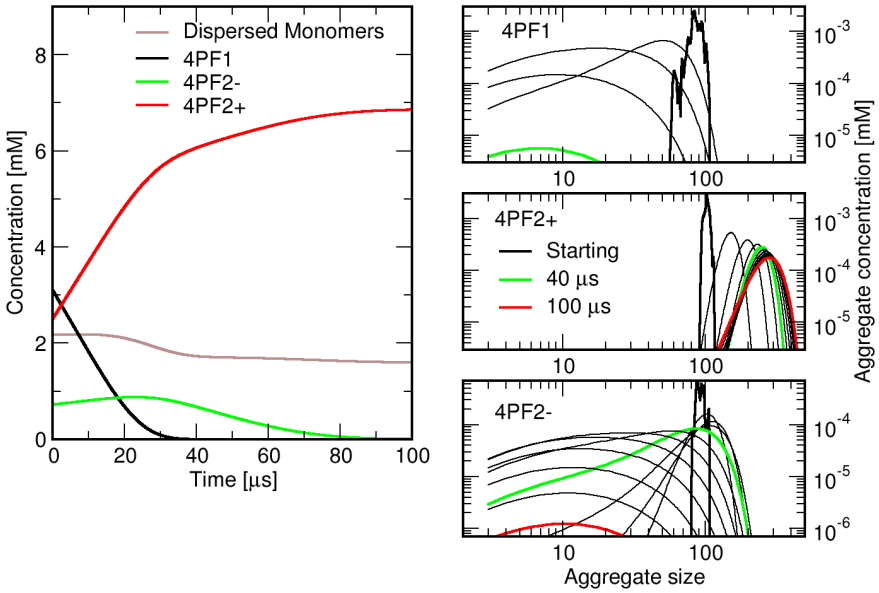


Fig. S8. **Solutions of the master equation of fibril elongation for dE=-2.5.** (Left) Time dependent concentration of monomers either dissociated or associated to fibrils of different morphologies. (Right) Time dependent concentration of aggregates of variable size and distinct morphologies. The curves are calculated every 10 μs and shown by thin black lines except for the 40 and 60 μs curves which are in green and red, respectively. The initial concentrations of aggregates obtained from the Langevin dynamics simulations are displayed as bold black curves.

11 Simulation of morphologies coexistence.

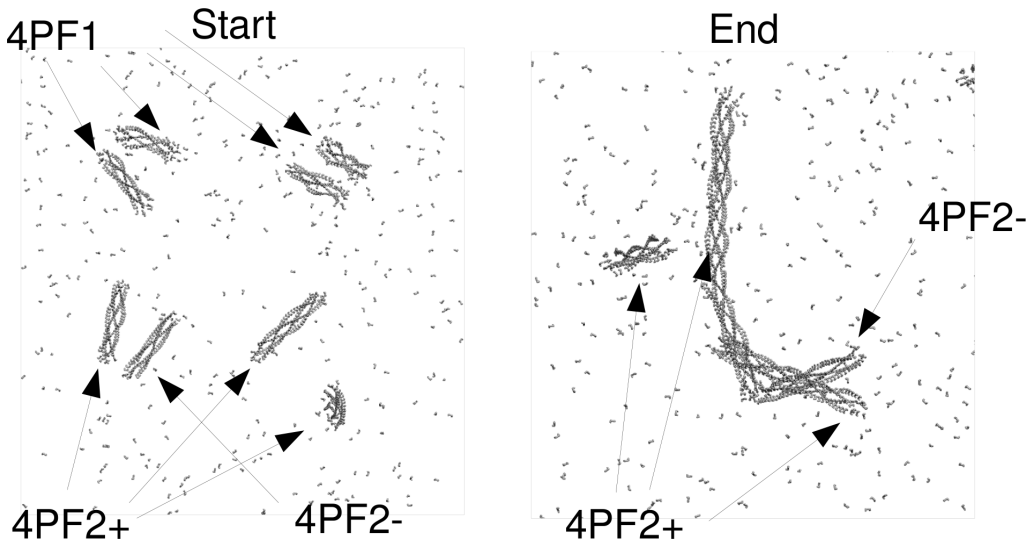


Fig. S9. **Simulation of morphologies coexistence.** (Left) Four 4PF1, three 4PF2+ and one 4PF2- mature fibrils obtained from nucleation simulations at dE=-2.5 are placed together in a box and simulated. The total amount of monomers is 1000. Note that the number of fibrils of different morphologies reflects the nucleation ratio discussed in the main text. During the simulation, the 4PF1 fibrils shrink and disappear. (Right) After 2 μs, the 4PF1 fibrils have disappeared, and 4PF2+ and 4PF2- morphologies merge together into a bundle of fibrils.

12 Oligomer size histogram.

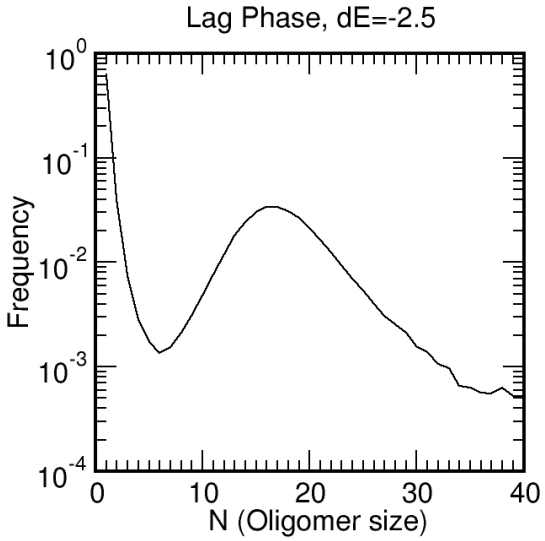


Fig. S10. Oligomer size frequency histogram calculated for dE=-2.5 at the lag phase, when micellar oligomers M are in equilibrium with monomers. The value N=7 separates the small oligomers from the micelles. The hystogram is calculated as the probability that a monomer is in an oligomer of size N.

13 Nucleus size.

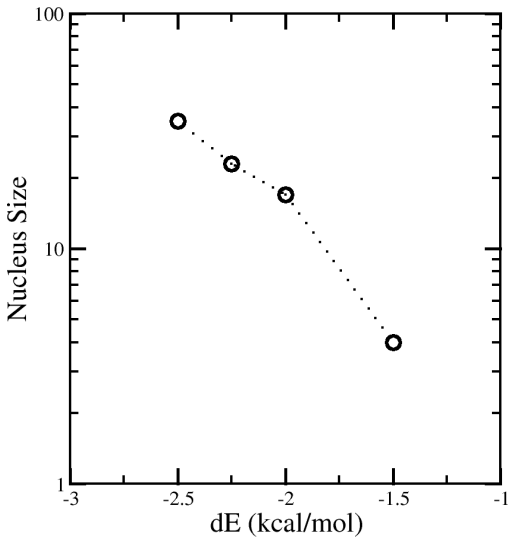


Fig. S11. Nucleus size calculated at four different values of the β-propensity dE. The nucleus is defined as the oligomer that has 50% probability to become a fibril. Details of the nucleus size calculation can be found in R. Pellarin and A. Caflisch, J. Mol. Biol. (2006) vol 360 p. 882.

14 Reference 47

B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, CHARMM: the biomolecular simulation program. *J Comput Chem* (2009) **30** 1545-1614