From peptide chains to chains of peptides: multiscale modelling of self-assembling fibril-forming polypeptides
Schor, M.

Citation for published version (APA):
Chapter 5

Self-assembly Mechanism of Fibril-forming Silk-based Block Copolymers

The triblock copolymers described in the previous chapters self-assemble into micrometer long fibrils in response to a trigger. Since the exact mechanism of the fibril formation remains unclear, in this chapter we employ a multiscale modelling approach in combination with rare event simulations to elucidate key processes. Atomistic scale simulations on the silk-based block suggest a mechanism in which a polypeptide prefolded in a β-roll structure docks to the growing end of a fibril through the formation of Glu-Glu sidechain contacts. Subsequently it can slide to the optimal position before water is expelled to form a dry interface between the fibril end and the attaching block copolymer. In addition, we find that the folded state of the silk-based block is further stabilised through interactions with its neighboring block. Templated folding may also play a role in case a partially folded polypeptide attaches. The coarse-grained simulations indicate that the attachment and subsequent sliding is mediated by the hydrophilic flanks in a size dependent manner. The hydrophilic blocks prevent random aggregation and allow growth only at the end of the fibril. Our multiscale approach may be used for other fibril-forming peptides.

5.1 Introduction

One of the main challenges in the field of nanotechnology is the construction of functional molecular devices by self-assembly processes. Biological self-assembling polymers or polypeptides have been a source of inspiration in many approaches to develop such new nanomaterials [2]. In particular, the self-assembling properties of β-sheet rich protein elements have received a lot of attention as promising building blocks in the design of functional supramolecular materials [4, 5].

Among the various β-sheet forming peptides that have been studied is the glycine-alanine (GA)_n repeat which is found in the B. mori silk fibroin. Although the peptide block can in

principle be chemically synthesized, genetic engineering allows for more control as it facilitates the production of monodisperse, high molecular weight polypeptides with a predetermined amino acid sequence. Various triblock copolymers with sequence \( ((GA)_n)(GX)_m \) and hydrophilic outer blocks have been reported \([70, 71, 161]\), where \( n \) denotes the number of repeats. Residue \( X \) can in principle be any bulky hydrophilic residue that disrupts the tight fibroin-like packing promoted by the GA repeats. Smeenk et al. have shown that triblock copolymers consisting of the silk-based repeat (with glutamic acid (E) at the X position) flanked by poly(ethylene glycol) (PEG) self-assemble into well-defined fibrils from methanol upon neutralising the glutamic acid sidechain by lowering the pH \([70]\). More recently Martens et al. have engineered a triblock copolymer with the same silk-based core block with sequence \( ((GA)_3GX)_{48} \), flanked by large, hydrophilic peptides \([71]\). This triblock copolymer also forms fibrils from aqueous solution when the pH is lowered. In both cases the silk-based block forms the core of the fibril and the hydrophilic outer blocks form a corona around this core. Although the fibrils seem similar, the structure of the silk-based block is solvent-dependent \([75]\). In aqueous solution, the silk-based block folds into a \( \beta \)-roll, while in methanol it forms a flat \( \beta \)-sheet. In both cases fibril formation is thought to occur through a nucleation-growth type mechanism.

In this paper we focus on the yet poorly understood mechanism of the fibril self-assembly in water. It is likely that, once formed, fibrils grow in a one-by-one fashion in which diffusing polypeptide molecules attach to the end of fibril. However, it is unknown whether these attaching polymers are prefolded into a \( \beta \)-roll or are in a random coil conformation. Furthermore, due to the non-specificity of the polymer sequence, it is highly unlikely that the polymer will attach at the correct position, even when it is preformed in a \( \beta \)-roll. Therefore, the roll will have to be able to realign to form the more or less perfect fibrils seen in experiment. These steps are indicated schematically in Fig. 5.1. Further complicating the mechanism, the hydrophilic blocks are required to prevent aggregation of the block copolymers as without them only random aggregates form.

Knowledge of the self-assembly mechanism allows a rational design and improvement of these type of novel materials. However, experimentally it is very difficult to investigate the process of folding and self-assembly of such block copolymers on a molecular scale. On the other hand, molecular dynamics (MD) simulations, in principle, allow for the elucidation of mechanisms at the molecular level. With the currently available computer power it is possible to reach up to microseconds for system sizes on the order of \( 10^5 \) particles. Our system spans a wide range of time scales: bond vibrations and dihedral angle rotations occur on picosecond and nanosecond timescales whereas folding into the \( \beta \)-roll and self-assembly of the fibril takes milliseconds to hours in the experimental setup. The full experimental polypeptide consists of approximately 800 amino acids. A system with tens of solvated polymers will easily contain millions of atoms. Hence studying the fibril formation of the full copolymer in solution at atomistic resolution with straight-forward MD simulations is clearly not feasible. In contrast, a multiscale modelling approach is able to capture the self-assembly processes at different length and time scales. All-atom simulations can give detailed insight into the interactions governing self-assembly of the silk-based blocks. Coarse-grained descriptions allow reaching the longer time and length scales over which the association of two or more block copolymers takes place.

In a coarse-grained force field the number of degrees of freedom is reduced in comparison to standard all-atom force fields. Reducing the number of degrees of freedom will increase
Figure 5.1: Network of possible self-assembly reactions of the silk-block. The β-roll structures are depicted in cartoon representation (green and purple), the Ala sidechains as red spheres. In the two rightmost structures, Glu residues are represented in licorice (red). Fibril formation can happen either by attachment of an unfolded polymer (left-most path) or by attachment of a prefolded roll (right pathways). If the polymer is unfolded, it can use the fibril surface as a template for folding. In the next step, the folded β-roll has to slide along the fibril surface to minimise exposed hydrophobic surface. If the a prefolded β-roll attaches, this can happen either through direct hydrophobic collapse (center) or through docking through Glu-Glu sidechain interactions followed by hydrophobic collapse. In the last case there are two possibilities to minimise exposed hydrophobic surface. If the hydrophobic collapse is immediate the roll will have to slide along the fibril surface. Alternatively, the roll can slide when in the Glu docked state and expel water in the last step. Note that only fragments of the full β-rolls are shown, and that for clarity, the hydrophilic C-blocks have been omitted.
Thirumalai model [105]. The version we use was optimised to describe the silk-based block as it behaves in water [181] and was shown to successfully reproduce fibril properties. Here we extend it to include a description of the hydrophobic block.

Even though the low resolution coarse-grained force field allows simulating larger systems sizes and time scales, it is still not sufficient to reach the folding and subsequent fibril formation time scales. The long timescales of these processes are governed by the presence of high free energy barriers. The self-assembly processes can thus be considered as rare events. Special simulation techniques have been developed to study such rare event processes, e.g. umbrella sampling [126], metadynamics [128], conformational flooding [129], local elevation [178], steered MD [141] and many others.

In this chapter we aim to elucidate the self-assembly mechanism of silk-based block copolymers into fibrils through a multiscale modelling approach combined with rare event simulations such as steered MD and umbrella sampling. The rare event methods, in this case steered MD and umbrella sampling, are necessary as self-assembly into fibrils involves high barriers.

The possible self-assembling mechanisms in Fig. 5.1 are compared for a small but representative part of the silk-based block which is described atomistically with explicit solvent. Subsequently, we simulate the whole block copolymer with a coarse-grained force field to study the effect of the hydrophilic blocks on self-assembly. The results obtained for the different scales can be combined to get an overall picture of the mechanism of self-assembly of these block copolymers into fibrils.

The full-atom simulations indicate that the most probable mechanism is the rightmost schematic pathway in Fig. 5.1: from pre-folded $\beta$-roll, via docking of Glu-Glu interface and sliding into place to minimise mismatch, followed by a hydrophobic collapse into the correct position. At first glance, a direct hydrophobic collapse of the $\beta$-roll onto the fibril end, skipping the Glu-Glu docked state and subsequent sliding process seems also possible. However, the subsequent hydrophobic mismatch is very unlikely to anneal.

Our coarse-grained simulations suggest that the main role of the hydrophilic blocks is indeed in limiting random aggregation of the silk-based block. This protection is more efficient for a larger hydrophilic block. The second role of the hydrophilic block is to pre-align the silk-based blocks such that they minimise their mismatch, resulting in more regular fibrils. This latter role might work hand in hand with the Glu-Glu docking and sliding, making this step more efficient.

5.2 Methods

5.2.1 All-atom Simulations

5.2.1.1 MD Simulations

The all-atom MD simulations have been performed with the GROMACS package version 4.0.5 [79]. The pressure was kept at 1 bar using Parrinello-Rahman coupling and the temperature was kept at 298K using a Nose-Hoover thermostat. LINCS was used to constrain the bonds, allowing for a 2 fs timestep. Electrostatics were treated with PME. The VMD package [172] was used to visualise structures.
5.2.1.2 The β-roll System

The all-atom simulations were performed on a β-roll with sequence $E((GA)_3E)^{10}$. This β-roll is large enough to be representative for the full length silk-based block used in experiments while decreasing the system size enough to allow full-atom simulations [75]. The structure of the β-roll is taken from our previous work [75]. All Glu sidechains are protonated and the termini are uncharged, resulting in a neutral system. The OPLS force field [182] to describe the β-roll was used in combination with the TIP4P water model. Two β-rolls are placed in a rectangular box (14x5x5 nm$^3$ solvated by 11307 waters unless noted otherwise) and are aligned in the $xy$-plane and stacked along the $z$-axis in such a way that they form a perfect dry interface with minimal exposed hydrophobic surface area. Upon solvation, the systems are energy minimised using conjugate gradient. Subsequently a 20 ps simulation with the β-rolls restrained was performed to equilibrate the water around the rolls.

5.2.1.3 Steered MD Simulations

Three different sets of all-atom steered MD (SMD) simulations were performed on the above β-roll system:

1. Unfolding two strands of one of the β-rolls by increasing the centre of mass (com) distance between residues Glu1 and Glu49 along the $x$-axis. Refolding is studied by releasing conformations from different com distances and simulating for 5 ns (5 simulations per com distance). The results are compared to those for a single β-roll (same setup; 1 β-roll and 11281 TIP4P waters) and are described in section 5.3.1.1

2. Unstacking the β-rolls by increasing the com distance along the $z$-axis (section 5.3.1.2). Here the box size is 6.5x6.5x8 nm$^3$ (2 β-rolls and 10679 TIP4P water molecules).

3. Sliding of one β-roll over the other one in the direction perpendicular to the strands ($x$-axis) by pulling on the com of the strands in contact with the other polypeptide (section 5.3.1.3).

One of the β-rolls is position restrained during these SMD simulations. All three sets of SMD simulations use a velocity $v = 0.2$ nm/ns and a force constant $f_c = 50000$ kJ$^{-1}$nm$^{-2}$. From SMD simulations it is possible to construct the potential of mean force (PMF) through Jarzynski’s equality [139, 141] (see chapter 2). The PMFs shown in Figs. 5.3 and 5.5 are based on 20 trajectories each.

5.2.2 Coarse-grained Simulations

5.2.2.1 MD Simulations

All MD simulations with the coarse-grained force field were run with the CM3D program [80]. As the program employs a multiple timestep integration algorithm (RESPA), bonds are not constrained. All simulations were run in the NVT ensemble at 300K unless described otherwise. The temperature was kept constant with a Nose-Hoover thermostat. The timestep was set to 4 fs. The VMD package [172] was used to visualise structures.
5.2.2.2 Setup of the Coarse-grained MD Simulations

The experimental block copolymer consists of a silk-based block with sequence \( E((GA)_3E)_{48} \) flanked by 198 amino acid long C-blocks. To keep our simulations tractable we have used smaller block copolymers. The size of the block copolymer should still be large enough to be representative for the full length block copolymer.

A stack of 32 \( \beta \)-rolls with sequence \( E((GA)_3E)_{24} \) flanked by two C-blocks of 96 residues was equilibrated at 100 K for 500 ps and subsequently for 1 ns at 300 K. Subsequently, the silk-based core was position restrained. This system was simulated for 20 ns at 300 K and the C-block density along the fibril axis was extracted to test the effectiveness of the shielding of the silk-core of the fibril by the hydrophilic C-block. (section 5.3.2.1).

5.2.2.3 Umbrella Sampling Simulations

To investigate the effect of the (relative) sequence length of the C-block with respect to the \( \beta \)-roll on self-assembly we simulate two small fibrils consisting of different block copolymers. All have a silk-based block with sequence \( E((GA)_3GE)_{12} \) and the length of the flanking C-blocks is 21, 48 or 95 amino acids, corresponding to \( R_g \) values of 1, 2 and 3 nm respectively. For each sequence, five block copolymers were stacked to form a small, regular fibril. Two such small fibril fragments were equilibrated at a fixed centre of mass (com) distance by first simulating 100 ps at 100 K and subsequently 500 ps at 300 K. Next, umbrella sampling simulations \([85,126]\) were performed with the umbrella biasing potential centered at 0.5 nm intervals, covering the range 5-15 nm (20 nm for the longest C-block to allow the interaction between the fibrils to decay to zero) and the force constant was set to 830 kJ mol\(^{-1}\)nm\(^{-2}\). Simulation time was 10 ns per window. Weighted histogram analysis \([132]\) was employed to obtain the PMFs as a function of the com distance between the fibrils. Note that only the particles of the central silk-based block were taken into account for the centre of mass distance and that the two fibrils are free to rotate during the simulation.

5.3 Results and Discussion

5.3.1 Self-assembly Behaviour of Silk-based Blocks

Fibrils are thought to self-assemble \textit{vis-à-vis} a nucleation and growth process in which the silk blocks attach to an existing fibril nucleus of properly stacked \( \beta \)-rolls. Once a \( \beta \)-roll is properly attached to a fibril, the surface between two polypeptides chains is completely dewetted, and all water molecules are expelled. In this \( \beta \)-roll stack the Ala sidechains on the outside of the rolls form a tightly interdigitated Ala interface analogous to the steric zippers observed in amyloids \([35]\). Here, we seek to address two aspects of the self-assembly process. First, it is unknown if the silk-based blocks first fold into a \( \beta \)-roll and subsequently assemble or if they assemble first and then use the fibril surface as a template for folding. These options correspond to the left and right pathways in Fig. 5.1. Furthermore it is unclear how the exposed hydrophobic surface at the interface between two \( \beta \)-rolls is minimised. Sliding along the Ala ‘rails’ is one of possible scenario’s in Fig. 5.1, but it could also be that this optimisation is done in the “Glu-docked” state.
Straightforward MD simulations would be unpractical to compare between various possibilities due to the high free energy barriers involved in these processes resulting in very long simulation times. To overcome this time scale issue, we employ SMD simulations to enforce the $\beta$-roll to perform according to a certain mechanism in an artificially fast manner. Nevertheless, the SMD results can be translated using Jarzynksi’s equality into an equilibrium PMF that will facilitate comparing various mechanisms.

5.3.1.1 Templated Folding of $\beta$-rolls

Although we have shown previously that one $\beta$-roll is stable in solution [75], this stability is enhanced when two or more polypeptides have stacked to form a fibril. An open question is whether or not the presence of the exposed fibrils Ala surface enhances the folding of a partially folded polypeptide attached to such a surface. We investigate this by unfolding a few strands of a $\beta$-roll in solution and of a $\beta$-roll stacked with another $\beta$-roll and subsequently monitoring the refolding behaviour of the two systems.

When the strands are fully unfolded (the C$\alpha$-C$\alpha$ distance between glutamates 17 and 33 (d17-33) is larger than 0.6 nm) no proper refolding was observed within the simulation time for either of the two systems. Instead the unfolded strands collapse into a molten-globule. However, when the simulations are initiated with a partially unfolded $\beta$-roll such that the distance d17-33 and the distance between residues 15 and 31 (d15-31) is around 0.5 nm spontaneous refolding of the second strand (residues 9 to 17) is observed within 5 ns when the $\beta$-rolls are stacked (Fig. 5.2) in 2 out of 5 simulation runs. Folding is enhanced by the formation of a dry alanine interface between this strand and the underlying polypeptide surface. Refolding of the first unfolded strand (residues 1 to 9) is not enhanced by the presence of the second polypeptide as it would be positioned at the solvent exposed side of the $\beta$-roll. Indeed refolding of the first strand only seems to occur when most contacts are already properly formed and does not depend on the presence or absence of the second $\beta$-roll.

These results indicate that the presence of the $\beta$-roll enhances folding. However, combining these results with the fact that a single $\beta$-roll in solution is very stable suggests that although templating will enhance folding, the presence of a template is not really necessary and stable $\beta$-rolls will also form in solution [75].

5.3.1.2 Stacking of $\beta$-rolls

In the first of the possible pathways on the right in Fig. 5.1 a prefolded $\beta$-roll attaches to the fibril end. We mimic this step by considering the stacking of just two $\beta$-rolls. For two $\beta$-rolls in solution to come together, water has to be expelled from the interface. In principle this process is reversible and calculating the free energy barriers for the process where two rolls unstack and water enters between the two hydrophobic surfaces will yield information on the reverse process of stacking of two $\beta$-rolls. The PMF from 20 SMD trajectories of unstacking two folded $\beta$-rolls and several representative configurations in Figs. 5.3 and 5.4 show that the free energy cost of the first step, where the stack opens up on one side and water penetrates the interface is approximately 400 kJ/mol. (80 kJ/mol per strand corresponding to 1920 kJ/mol for the full experimental silk-based block). The two $\beta$-rolls remain in contact through interactions between the Glu sidechains on one side (Fig. 5.4a3). Once the Ala interface between the two $\beta$-rolls is
Figure 5.2: Templated folding. Starting from a state where the first strand is fully unfolded and the second strand is mostly unfolded (c) refolding of the second strand seems to be enhanced by the presence of the other β-roll resulting in structure (e). The plot in (b) shows that the strand refolds according to a zipper-like mechanism. In the absence of the other β-roll, no refolding is observed (a). Instead, the strands form a collapsed state (d).

solvated by a layer of water, a further increase of the center of mass distance is almost without free energy cost until the hydrogen bonds between the Glu sidechains have to be broken. This hydrogen bond breaking step occurs when moving from point 3 to 4 in Fig. 5.3a) and costs around 150 kJ/mol (30 kJ/mol per strand, 720 kJ/mol for full silk block) in free energy. Only when these h-bonds are broken are the β-rolls fully detached. Remarkably, the unrestrained β-roll remains stable during this forced process.

The reverse process of two rolls spontaneously assembling from solution can be studied by taking starting configurations at various com distances from the SMD simulations (indicated by the arrows in Fig. 5.3a), reinitialising momenta, and running straightforward MD simulations. Below a com distance of 1.4 nm (d4) the β-rolls reassemble within 1.5 ns, and completely expel the water layer from the Ala interface. At distances larger than 1.4 nm the β-rolls do not spontaneously reassemble within the simulation time (5 ns). Nevertheless, the folded polypeptides
Figure 5.3: Unstacking and restacking of β-rolls. (a) PMF of unstacking two β-rolls. Going from point 1 to point 3 in the PMF the two rolls open up allowing water to penetrate the interface. At point 4 the Glu-Glu sidechain interactions are broken and the two β-rolls are fully solvated. From point 3 (also d5) reverse pulling simulations have been run. The resulting PMF (dashed grey line) largely overlaps with the unstacking PMF but remains finite. This hysteresis arises from one or two water molecules that are trapped between the two rolls due to too high pulling velocity. The distance between the two β-rolls (b) as a measure for reattachment taking starting configurations from different points along the PMF (indicated with arrows in (a)). This suggests that the β-rolls quickly reattach provided that there is still enough contact between the Glu sidechains.

...still are in contact with each other at their Glu sidechains. In addition to investigating spontaneous assembly, we have calculated the PMF for pulling the two β-rolls together starting from point 3 (20 trajectories). This PMF is shown as a dashed grey line in Fig. 5.3a. The resulting end structure is not entirely identical to the initial stacked β-roll structure, as one or two water molecules get trapped between the two rolls, and hence the PMF does end at a value higher than zero. Lower pulling velocities may avoid this problem. Still, the slope of the PMF (i.e. the force) agrees well with the PMF for unstacking, indicating that the observed hysteresis for this process is not that large.

The PMF in Fig. 5.3 indicates that assembly of the β-rolls is an almost irreversible step, as overcoming a free energy barrier of 550 kJ/mol ≈ 220 k_BT (corresponding to 1060 k_BT for the full experimental silk block) by thermal motion will take longer than the lifetime of the universe. In addition the results suggest that during the initial contact between two β-rolls, hydrogen bonds forms between the Glu sidechains. Subsequently, the rolls slowly approach each other and water is expelled to form the dry Ala interface, in an almost irreversible step.
Figure 5.4: Representative structures along the unstacking PMF (Fig. 5.3a). The structures on the left show how the distance between the β-rolls increases. The structures on the right indicate how water penetrates the space between the two rolls (for clarity only the fixed β-roll is shown).

5.3.1.3 Sliding of β-rolls

It is possible that two β-rolls approaching for assembly are not correctly aligned, resulting in a mismatch of their hydrophobic Ala-surfaces. To minimise their unfavourable exposed hydrophobic surface areas the β-rolls may, in principle, slide along each other in the direction perpendicular to the strands. Again, whether or not this sliding mechanism is kinetically accessible, can be tested by computing the PMF from SMD simulations.

The PMF for forcing one β-roll to slide over the other one in the direction perpendicular to the strand is shown in Fig. 5.5. The results indicate that sliding involves overcoming a free energy barrier of approximately 300 kJ/mol (60 kJ/mol per strand, 1440 kJ/mol for the full silk
The slope of the PMF is confirmed by performing umbrella sampling simulations at three points along the curve. Note that the PMF is expected to be symmetric, as the final structure is virtually identical to the initial structure. The observed asymmetry is possibly due to hysteresis, or inaccuracies in the PMF computation. Note however, that thermal motion alone will not be able to overcome such a large barrier in a single collective step in any case.

Inspection of the configurations at the points indicated by 2 and 3 in Fig. 5.5a) reveal that the \( \beta \)-rolls do not slide along each other in a rigid body fashion. Instead, the structure is deformed: first one horizontal Ala row moves one strand to the left, then the second one follows and in a third step the last row moves as well. The representation of the \( \beta \)-roll stack in Fig. 5.5c) shows clearly that due to the interdigitation there is indeed very little room for Ala sidechains to wiggle their way through the line of Ala of the other \( \beta \)-roll to their next optimal position.

![Figure 5.5: Sliding of \( \beta \)-rolls.](image)

As the sliding mechanism involves overcoming a rather high free energy barrier in the order of 1440 kJ/mol it is unlikely that two \( \beta \)-rolls will minimize their mismatch in this way. Alterna-
Figure 5.6: Sliding in the Glu docked state (b). The resulting PMF (a) indicates that the free energy barrier for this process is much lower compared to the process of sliding from a fully stacked state (Fig. 5.5a).

*Figure 5.6: Sliding in the Glu docked state (b). The resulting PMF (a) indicates that the free energy barrier for this process is much lower compared to the process of sliding from a fully stacked state (Fig. 5.5a).*

...tively, the rolls could slide along each other in the Glu docked state suggested in the previous section (see Fig. 5.4a3 or Fig. 5.6b). Forcing one β-roll to move over the width of one strand with respect to the other β-roll results in the PMF shown in Fig. 5.6a. The free energy barrier for sliding in this docked state is only 50 kJ/mol (10 kJ/mol per strand or 240 kJ/mol for the full silk block). Moreover, in the docked position, it is possible that the sliding is not a single step collective process but a diffusive process in which hydrogen Glu-Glu bonds are being formed and broken on a per strand basis. This would involve only around 10 kJ/mol $\approx 4k_B T$, which should be accessible by thermal motion alone on a ns timescale.

To summarize, the results from the atomistic simulations suggest a mechanism where a β-roll attaches to the growing end of a fibril by docking through the formation of Glu-Glu hydrogen bonds. From this docked state the roll can slide to a more optimal position prior to the formation of a dry interface. Sliding may be mediated in part by the hydrophilic blocks. The flexible hydrophilic C-block flanking the silk β-roll is the subject of the next section.

### 5.3.2 Effect of the Hydrophilic Blocks on Self-assembly

Experiments show that the presence of the hydrophilic flanking blocks is needed for fibril formation as without them the silk-based blocks aggregate randomly. Including the hydrophilic blocks increases the system size such that all-atom simulations would be unfeasible. Therefore we use a coarse-grained force field to study the effect of the hydrophilic blocks on self-assembly. Development and testing of the coarse-grained force field is discussed in the previous chapter.
5.3.2.1 Shielding of the Silk-based Fibril Core by the C-block

Fibril growth can occur only through attachment of a block copolymer at the hydrophobic growing ends of the fibril and not at one of the hydrophilic Glu sides. This is facilitated by efficient screening of the Glu sides and the strand sides by the corona formed by the hydrophilic C-blocks. To investigate this screening effect we simulated a fibril of 32 block copolymers consisting of a silk-based block with the sequence $E((GA)_3E)_{24}$ flanked by two 96 residue C-blocks. The fibril is aligned along the $z$-axis and the C-block density in the $xy$-plane is extracted for several intervals along the fibril axis (see Fig. 5.7). Indeed, the C-block forms a corona of about 1 nm around the silk-based core. At the ends of the core (0 and 10), the C-blocks occupy the entire $xy$-range of the hydrophobic surface. The C-block density in these intervals is significantly lower compared to the other intervals.

Thus, it seems most likely that fibril growth occurs by attachment at the hydrophobic surface of the fibril as the C-block density is lowest here and the effective interaction between a monomer and the fibril is strongest. Shielding will be less effective for shorter C-blocks.

Occasionally, fibrils are observed to align in experiments [71]. Alignment is most likely to occur along the Glu-surfaces as the C-block density is lower here. This lower shielding also provides a way for an attaching β-roll to be ‘guided’ to the correct position for Glu-docking (see section 5.3.1). In this sense, the lower densities of the hydrophilic C-block at the fibrils ends naturally induces the self-assembly properties.

5.3.2.2 Self-assembly of Small Silk-based Fibrils

It has been suggested that the main role of the hydrophilic outer blocks is to limit random aggregation of the hydrophobic silk-based polymers. Indeed, experiments of silk-based polymers without flanking hydrophilic blocks show rapid formation of amorphous aggregates. It seems intuitive that the efficiency by which the hydrophilic blocks limit random aggregation would depend on the size ratio between the radius of gyration ($R_g$) of hydrophilic block and the linear size of the folded hydrophobic block. Three scenarios can be distinguished: hydrophilic blocks too small for effective screening, hydrophilic blocks that are too large to allow any interaction between the hydrophobic blocks and the in-between case where efficient screening results in the formation of regular fibrils.

We compute the effective interaction free energy (identical to the PMF) between the centers of mass of two short fibril fragments for three different C-block lengths (21, 48 and 95 residues, with a $R_g$ of 1, 2 and 3 nm, respectively) through umbrella sampling. The resulting free energy curves plotted in figure fig. 5.8 are repulsive at most distances, indicating a steric (excluded volume) stabilisation that is common in polymer and colloid science [183]. Clearly, the range of the repulsive interaction as well as the strength of the repulsive interaction increases for longer C-blocks. For center of mass distances corresponding to near contact, the C-blocks become depleted from the interface between the silk-based fibril fragments. At these small distances, the fibril fragments will snap into contact, as soon as a sufficiently large hydrophobic surface becomes available. This will sometimes result in a mismatch of the β-roll stacks. While subsequent minimisation of the exposed hydrophobic surface, for instance by sliding of block copolymers along each other, should in principle be possible, it would require overcoming such large barrier, that it has not been observed in the simulations. In fact, as argued in section 5.3.1, this sliding
Figure 5.7: Two-dimensional plots of the residue density of the C-block along the length of a fibril. For clarity a cartoon of the fibril (red/green) surrounded by its corona (blue) is shown. The long fibril axis is aligned with the z-axis, the width and height are orientated along the x and y axis, respectively. The fibril’s center of mass is placed in the origin. The second panel on the top row shows the normalised C-blocks’ residue density profile projected on the z-axis for a fibril consisting of $32 \text{C}^{96}_E((\text{GA})_3\text{GE})_{24}\text{C}^{96}$ block copolymers. The ends of the fibril core formed by the silk-based blocks are indicated by black dotted lines. The C-block density naturally is highest around the center of the fibril and decreases rapidly beyond the core. The remaining three panels show contour plots of the density of the hydrophilic residues in the xy-plane perpendicular to the fibril direction for specific z-value, indicated by the grey dashed lines in the overall density plot. We show the density for regions 0, 5 and 9 representative of the outsides and the centre of the fibril. For regions 0 and 9, the contours indicate densities of approximately $1.9 \text{ particles/nm}^3$, $3.8 \text{ particles/nm}^3$ and $5.7 \text{ particles/nm}^3$. In region 5 the density of the C-block is highest and surrounds the fibril. For this region the contours correspond to densities of approximately $5.5, 11, 16.5$ and $22 \text{ particles/nm}^3$.

will not happen on a reasonable timescale, making the attachment virtually irreversible. The amount of mismatch seems to be dependent on the size of the C-block. In Fig. 5.8c we show a snapshot with a clear mismatch in case of 48 residue C-block. In contrast, for the longer 95 C-block (Fig. 5.8d), the mismatch has almost disappeared. It is hence conceivable that the C-block length not only plays the role of avoiding random aggregation, but also helps to minimise the mismatch between the polypeptide surfaces. Note that in the coarse-grained simulation, the Glu-Glu docking mechanism does not seem to occur. In the full-atom simulations, the Glu docked state is relatively long lived because water has to be expelled from the interface between the rolls. Water is not explicitly taken into account in the coarse-grained simulation.
Figure 5.8: (a) PMFs of self-assembly of CSC stacks with different lengths of C-block. (b) Starting structure of $C_{48}S_{97}C_{48}$. (c,d) Self-assembled structures of $C_{48}S_{97}C_{48}$ and $C_{95}S_{97}C_{95}$ respectively. While the first structure shows a significant mismatch in terms of exposed hydrophobic surface, this is much reduced in the second structure as a result of the larger C-blocks.

Also, in the experimental system, large mismatches can lead to branching of the fibrils. Such branching is relatively rare, but is sometimes observed [71].

5.4 Conclusions

By combining extensive all-atoms and coarse-grained simulations with rare event techniques, we have investigated the self-assembly of fibril forming polypeptides consisting of a central silk
block flanked by two hydrophilic blocks. We employed SMD and umbrella sampling simulations to test several keys steps for self-assembly. The multiscale nature of the modelling attempt is crucial, due to large size of the fully formed fibrils, and the slow time scales of polymer relaxation.

Based on our all-atom and coarse-grained rare event simulations we can propose a putative mechanism for the self-assembly of silk-based block copolymers into fibrils. Upon changing the pH, the polymer first folds into a $\beta$-roll. While we have not studied the folding rate of the polypeptide, we can assume it is faster than the growing fibril. Experiments indicate that fibril growth is diffusion limited (L. Beun, F. de Wolf, M.A. Cohen Stuart, private communication), which indicates that the folding into the $\beta$-roll is not the rate limiting step. Another reason is the presence of hydrophilic C-block flanking the silk-based block, which probably allows the silk-based blocks to fold properly without random aggregation of multiple polymers [184].

Once folded, a $\beta$-roll, protected by the hydrophilic, C-block corona, diffuses around in solution, until it docks to a existing fibril end through the formation of Glu-Glu hydrogen bonds. Our coarse-grained simulations of fibril fragments shows that the corona induces a repulsion between polypeptide surfaces, but one which is not so large that it is impossible to overcome. The newly attached $\beta$-roll may subsequently slide along the underlying $\beta$-roll to optimise the number of Glu-Glu contacts. Next, intervening water is expelled in an almost irreversible step and the $\beta$-roll is stacked onto the fibril growing end through the formation of a dry interface in which the Ala sidechains form a steric zipper similar to those observed for many amyloid fibrils [35].

A mechanism in which an unfolded polypeptide attaches to the fibril, and folds to a $\beta$-roll through templated folding is unlikely, due to the high cost of correcting for mismatched rolls. Note however, that it might be possible for a partly folded $\beta$-roll to dock to the Glu interface, followed by enhanced templated folding, followed by sliding and hydrophobic collapse. We did not investigate this possibility in our simulations.

Thus, the most probable mechanism is the rightmost schematic pathway in figure 5.1: from pre-folded $\beta$-roll, via docking of Glu-Glu interface and sliding into place to minimise mismatch, followed by a hydrophobic collapse into the correct position. At first glance, a direct hydrophobic collapse of the $\beta$-roll onto the fibril end, skipping the Glu-Glu docked state and subsequent sliding process seems also possible. However, the subsequent hydrophobic mismatch is very unlikely to anneal.

Our coarse-grained simulations suggest that the role of the hydrophilic blocks is actually twofold. The main role of the hydrophilic blocks is indeed in limiting random aggregation of the silk-based block. This protection is more efficient for larger hydrophilic blocks. The second role of the hydrophilic block is to pre-align the silk-based blocks such that they minimise their mismatch, resulting in more regular fibrils. This latter role might work hand in hand with the Glu-Glu docking and sliding, making this step more efficient.

Our predicted mechanisms may apply more generally, for instance in stacking of natural amyloid fibrils into bundles as many of the amyloidogenic regions forming the core of the amyloid fibril are also flanked by several hydrophobic residues. Or predictions could ultimately be tested in future experiments. Instead of the pH trigger, the Glu sidechains could be neutralised using another counter ion. In this case no hydrogen bonding between the Glu sidechains can be formed and Glu-Glu docking will be inhibited. Also, the Glu residues could be mutated into
lysines or histidines as this would also result in different hydrogen bonding behaviour. Besides a range of different length hydrophilic flanks could be tested.

Our predictions can in principle be tested in future experiments, for instance using single molecule pulling.

Finally, our multiscale modeling approach could be applied to the self-assembly other fibril forming polypeptides.