

# Secondary Structure Transition and Critical Stress for a Model of Spider Silk Assembly

Tristan Giesa<sup>1</sup>, Carole C. Perry<sup>2,\*</sup>, and Markus J. Buehler<sup>1,\*</sup>

<sup>1</sup> *Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA 02139, USA*

<sup>2</sup> *Biomolecular and Materials Interface Research Group, Interdisciplinary Biomedical Research Centre, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom*

Keywords: molecular dynamics, simulation, secondary structure, protein, synthetic silk, silk fiber

## **Abstract**

Spiders spin their silk from an aqueous solution to a solid fiber in ambient conditions. However, to date the assembly mechanism in the spider silk gland has not been satisfactorily explained. In this paper, we use molecular dynamics simulations to model *N. clavipes* MaSp1 dragline silk formation under shear flow and determine the secondary structure transitions leading to the experimentally observed fiber structures. While no experiments are performed on the silk fiber

itself, insights from this polypeptide model can be transferred to the fiber scale. The novelty of this study lies in the calculation of the shear stress (300-700 MPa) required for fiber formation and identification of the amino acid residues involved in the transition. This is the first time that the shear stress has been quantified in connection with a secondary structure transition. By study of molecules containing varying numbers of contiguous MaSp1 repeats we identified the smallest molecule size that gives rise to a 'silk-like' structure contains six poly-alanine repeats. Through a probability analysis of the secondary structure we identify specific amino acids that transition from  $\alpha$ -helix to  $\beta$ -sheet. In addition to portions of the poly-alanine section these amino acids include glycine, leucine and glutamine. Stability of  $\beta$ -sheet structures appears to arise from a close proximity in space of helices in the initial spidroin state. Our results are in agreement with the forces exerted by spiders in the silking process and the experimentally determined global secondary structure of spidroin and pulled MaSp1 silk. Our study emphasizes the role of shear in the assembly process of silk and can guide the design of microfluidic devices that attempt to mimic the natural spinning process and predict molecular requirements for the next generation of silk-based functional materials.

## **Introduction**

Spider silk is one of the strongest and toughest natural fibers known and offers many promising technical applications.<sup>1, 2</sup> While first attempts to synthesize silk fibers have been successful, and industry-scale production of recombinant silk is possible, the mechanical properties of the spun fibers either do not currently compare with natural fibers or the assembly is performed under non-ambient conditions.<sup>3, 4</sup> Furthermore, many synthesized fibers contain a substantial amount of regenerated silk, i.e. the spinning dope originates from natural spider silk.<sup>4</sup> Using genetically engineered sequences presents a way to reduce limitations on fiber supply from spiders, but yields weaker and more brittle fibers.<sup>5</sup> The inferior quality of the engineered fibers mainly arises from the selection of assembly process parameters in industrial spinning devices, different rheological behavior and molecular weights of the material.<sup>5</sup>

### **Fiber formation- spinning from solution**

Silks have a modular structure that can be identified from their primary amino acid sequence, Figure 1a. The sequence is highly repetitive, where the number of repetitions  $a$  is in the order of 100's.<sup>6</sup> The dragline silk spidroin, i.e. silk before spinning, which is stored in the glands of the spider, consists mainly of two proteins, MaSp1 (studied here) and MaSp2.<sup>7</sup> Although major ampullate silk fibers contain up to 17 distinct proteins that are spun into fibers,<sup>8</sup> our study on MaSp1 is useful for synthetic silk fiber production protocols, which are largely focused on spinning materials exclusively from block repeats derived from MaSp1 recombinant proteins.

Spider dragline silk after spinning comprises crystalline  $\beta$ -sheet crystals embedded in an amorphous or poorly structured matrix. Recent studies on spidroins present in the glands of *N. clavipes* using HR-MAS-NMR spectroscopy and Raman microspectrophotometry have shown that spidroins vary in structure from being either totally disordered or natively folded.<sup>9-11</sup> None

of the spinning dopes were found to possess silk in  $\beta$ -sheet conformation. The presence of  $\alpha$ -helices and left handed  $3_{10}$ -helices were identified by Raman microspectrophotometry,<sup>10</sup> with NMR studies in contrast suggesting the silk is stored in a random coil state.<sup>11</sup> The MaSp sequences were shown to have no natural propensity for aggregation to give  $\beta$ -amyloid structures, a property of importance for silks in biomedical applications.<sup>6,9</sup>

The process for silk assembly and spinning into fibers is complex and there are a variety of theories for the work-flow of the natural assembly process.<sup>12</sup> Shear stresses at the wall together with removal of water from the protein lead to the formation of a nano-composite with an aligned  $\beta$ -sheet rich crystalline phase. Spinning conditions including humidity, temperature, reeling speed and ionic conditions are believed to control the assembly of silks.<sup>12</sup> The protein needs to be protected from premature assembly, which would clog the spider's spinneret or the spinning device.

Two theories have been proposed for the transition of silk spidroin to fibers.<sup>13</sup> First, the micelle theory states that the silk dope is stored in the spider's abdomen in micelle form, where the hydrophobic terminal regions of the sequence shield the interior  $\alpha$ -helical structure from water.<sup>14</sup> This theory assumes a low concentration of the spinning dope. During the spinning process, the applied shear induces the formation of  $\beta$ -sheets, while a change in  $pK_A$  and ionic conditions mainly affects the structure of the terminal region.<sup>15</sup> Simulation studies have confirmed that the terminal domains are critical in the initiation of multimerization.<sup>16</sup> It has been suggested that the C- and N-terminal of MaSp are required to delay premature assembly and transitions from monomer to homodimer upon the shift from neutral to acidic conditions in the spinning duct.<sup>16,17</sup>

The second theory, called the liquid crystal theory, states that spider silk proteins are stored in a nematic liquid-crystalline phase.<sup>18</sup> This theory assumes a high concentration of the spinning dope (~50%).<sup>18</sup> During the assembly in the spinning duct, the proteins organize in bilayered

disks that elongate along the fiber axis under shear flow. Under these conditions, random-coil and polyproline-II helix-like conformations transition to  $\beta$ -sheet-rich structures.

It is evident that shear stress facilitates secondary structure transitions during the assembly of silks.<sup>19</sup> Without shear stress, silk fibers do not form.<sup>20</sup> Experimental studies at mid-range concentrations to mimic the spinning process in microfluidic devices have shown that shear flow is essential for fiber formation.<sup>21</sup> High elongational flow rates of about  $1000 \text{ s}^{-1}$  in an *in vitro* microfluidic device were shown to be essential, though the shear forces were not measured in the spider. An increase in orientation of the crystallites in the fiber, shear thinning and an increase in force along the spinning duct has been observed.<sup>18</sup> Higher draw rates also lead to smaller crystals which increase stiffness and fracture resistance.<sup>22, 23</sup> Up to a limit of 50mm/s, XRD results show that the higher reeling speed also helps the orientation and formation of the silk network, while no influence on the secondary structure composition could be determined.<sup>24-26</sup> Premature *in vitro* self-assembly poses one of the issues involved with the synthetic production of silk fibers. Furthermore, in the current design of the microfluidic device the silk solution flows in an ionic solution that induces very little shear stresses in the silk solution.<sup>27</sup>

### **Silk structure after spinning**

Once the MaSp silk has been spun there is experimental evidence that all residual native-like structures are absent.<sup>28</sup> As might be expected, there is more known about the crystalline regions than the disordered regions owing to a paucity of methods to study amorphous or poorly ordered materials. In *N. clavipes* drag-line silk the crystalline phase arises from the organization of poly-alanine segments of the silk protein sequence into  $\beta$ -sheet structures. However, the precise  $\beta$ -sheet content reported varies according to the measurement technique used. Analysis by solid state NMR spectroscopy or Raman microspectroscopy suggests levels of 30 - 40%,<sup>29-31</sup> as

opposed to about 20% identified by XRD.<sup>32</sup> Computer simulations of the crystalline components of silk suggest that the poly-alanine sections align anti-parallel in the H-bonding direction with parallel stacking in the side-chain direction to stabilize the  $\beta$ -sheets.<sup>33, 34</sup> The high strength of silk materials is mainly attributed to the nanoconfined crystals (4-6 nm), while the breaking elongation is mainly attributed to the unfolding amorphous region.<sup>35-37</sup> Recently, the Brillouin spectroscopy technique has been applied to the non-invasive determination of a range of elastic moduli in spider silks.<sup>38</sup> Whereas this method is capable of determining the elastic tensor and mechanical properties of assembled fibers, here we are focusing on the assembly of the molecular structure.

The non-crystalline regions are alternately described as amorphous, poorly oriented or randomly coiled. They are not very well understood, although analysis of silk fibers by NMR has provided significant detail on the likely structures present in the less ordered domains.<sup>39-41</sup> The presence of  $3_{10}$ -helical structures with GLGXQG motifs forming  $\beta$ -turns was proposed in the late 90s.<sup>39, 42</sup> Later studies suggested that these amino acids are 'ordered' on the NMR timescale of microseconds - with poly-alanine and GGX motifs probably located in microcrystalline domains and the remainder of the silk in a  $3_{10}$ -helix arrangement.<sup>40</sup> Other structures including  $\beta$ -turn,  $\beta$ -spiral and helices have been suggested with different views as to the structure adopted by GGX;  $3_{10}$ -helices, less ordered helices/ distorted  $\beta$ -sheets have all been proposed.<sup>10, 30, 40, 41</sup> To further complicate matters, more recent studies have suggested that there are three distinct structure types in silk: crystalline  $\beta$ -sheet, weakly ordered  $\beta$ -sheet, and amorphous or disordered structures.<sup>25, 43-45</sup>

### **Scope of this paper**

In this study, we model silk formation under shear flow and determine the secondary structure transitions leading to the experimentally observed fiber structures, Figure 1a. The structures modeled have two, four and six poly-alanine repeats respectively, Figure 1b. Additionally, intra-chain interactions are studied, Figure 1c. After shearing, the stability of single chains and layers in water and vacuum is determined, Figure 1d.

The novelty of this study lies in the calculation of the shear stress required for fiber formation and identification of the amino acid residues involved in the transition. We quantify the importance of shear in the assembly process and predict the smallest molecule size that gives rise to a ‘silk-like’ structure, providing an important step towards the design of artificial spinning devices. Given the current activity in the field, a detailed understanding of the role of shear stress in the assembly mechanism could lead to the next generation of silk-based functional materials.

## **Materials and Methods**

### **MaSp1 Silk Model**

The 33 amino acid repeat unit of the wildtype *Nephilia clavipes* silk MaSP1 sequence AAAAAAGGAGQGGYGGLGSQGAGRGLGGQGAG (accession number UniProt KB P19837) is used in this study. Structures for simulation were built to contain two, four or six poly-alanine regions connected by the 27 amino acid glycine-rich semi-amorphous region, see schematic in Figure 1b.

The structures differ from those used in the studies by Keten and Buehler<sup>34</sup> and Bratzel and Buehler<sup>46</sup> where bundles of polypeptide chains, each comprising a single poly-alanine region with an amorphous region at each end were studied by REMD to replicate the  $\beta$ -sheet crystal dimensions and explore the length of the poly-alanine region to obtain maximum theoretical strength and toughness. In this study, multiple copies of the complete silk MaSp1 sequence are

present within an individual structure. Structures are initially built using the Biopolymer module in Tripos Sybyl. Structures in random conformation with neutral ends were minimized in vacuum before being exported for replica exchange calculation (REMD).<sup>47</sup>

### **Replica Exchange Molecular Dynamics (REMD)**

The REMD method is used to identify molecular equilibrium structures. 48 replicas of strands with initial configuration, as schematically shown in Figure 1b, are simulated at temperatures between 300 and 800 K to ensure sampling of the entire conformational space. The simulations are carried out using the CHARMM force field<sup>48</sup> implemented in NAMD,<sup>49</sup> using the Generalized Born implicit solvent model ( $\alpha$ -cutoff 12.0, ion concentration 0.3). Solvent friction is added via a Langevin friction term ( $10 \text{ ps}^{-1}$ ). A cutoff of  $15 \text{ \AA}$  is used for long range interactions and a timestep of 2 fs. REMD is run for 20 ns /sample (total of 960 ns) and the temperatures exchanged every 0.2 ps. Statistical analysis is performed using the MMTSB toolbox,<sup>50</sup> with a k-means clustering algorithm (cluster distance  $2 \text{ \AA}$ ). Representative final ensemble structures are chosen from the lowest temperature replica (300 K). The higher temperatures are used for fast conformational search and overcoming kinetic trapping in the REMD scheme.

The center of the largest cluster, meaning the structure with minimum RMSD, is exported for analysis in explicit solvent. To obtain more realistic molecular conformation and secondary and tertiary protein structure, the structure is equilibrated for 20 ns using Gromacs (CHARMM27 forcefield) in a charge-neutralizing periodic waterbox of TIP3P containing 150 mM sodium chloride. To prevent image interactions, the waterbox pads the protein by at least  $10 \text{ \AA}$ . Equilibration is performed with Langevin dynamics in an isothermal-isobaric ensemble (300K, 1bar) using the Nose-Hoover and the Parrinello-Rahman algorithm. Particle Mesh Ewald (PME)



electrostatics with 1.4 nm cut-off is used to more accurately capture solvent interactions.<sup>51</sup> It is possible to use a 2fs timestep because of holonomic constraints (LINCS algorithm). The secondary structure of each structure is determined using DSSP toolset and VMD.<sup>52, 53</sup>

### **Shear Flow and Steered Molecular Dynamics**

In order to simulate shear boundary conditions, Steered Molecular Dynamics (SMD) simulations are performed.<sup>54</sup> Soft springs (stiffness  $k$ ) are fixed to the symmetry points of the equilibrated structures, as shown in Figure 1c. The springs are set to provide stiffness only in the pulling direction so that the structure is able to relax in all other directions and hence no unnecessary constraints are enforced. An isothermal ensemble in an explicitly solvated periodic waterbox is set up. Isobaric constraints are enforced in the two directions that are perpendicular to the pulling direction. The box size in pulling direction is constant and periodic, but significantly larger than the molecule. A set of pulling speeds (0.5 – 5 m/s) and spring stiffness's (100-2000 kcal/mol/Å ) is used to probe the robustness of the results.

Two shear boundary conditions, Figure 2a (i) and (ii), are tested to investigate the evolution of secondary structure and shear stress. The part of the sequence colored in red is the glycine rich region and the part of the sequence colored in blue is the alanine rich region. The symmetry positions of the sequence are held by soft springs that act only in the pulling direction to ensure that the transition states can be captured during the simulation. The 'x' in Figure 2a(ii) marks a fix (in the symmetry positions of the chain) at the alpha-carbon of the terminal amino acid. In order to obtain an ensemble of different structures and associated force-displacement results, rather than changing the attachment points, several initial conditions are chosen from the equilibrium trajectory of the molecules. The results and error bars in Figure 4a and 4b comprise these initial conditions.

The pulling force is determined and the diameter calculated at all points of the trajectory by the in-plane radius of gyration of the structure. Figure 2b shows the trajectory of secondary structure of the equilibrated structures during the pulling experiment. The plot shown is one out of four tests for  $a = 3$  (with boundary condition (i)). All following plots show averages over the results of the two shear boundary conditions.

The pulling speed is set to 0.5 m/s, which is within the range of experimental drawing speeds (usually 1 mm/s – 500 mm/s),<sup>19</sup> leading to a simulation time of 10 ns. Higher pulling speeds have no significant influence on the results (see supplementary material). This is in agreement with NMR experimental results that state that the reeling speed has little effect on the secondary structure of the fiber.<sup>55</sup> Similarly, the spring constant used in the SMD experiment has little influence on the results, see supplementary material Figure S6. The average  $\beta$ -sheet content for each structure is determined by averaging the  $\beta$ -sheet content for 1 ns as indicated in Figure 2b. The transition shear stress, an example shown in Figure 2c, is averaged in the same region as that taken to assess the  $\beta$ -sheet content. Notably, after approximately 8 ns the stress exceeds the strength of the material and the simulation results are valid only before this point in the simulation. In all cases, the structure transition happens prior to reaching this limit. There is most likely an influence of the specific pulling residue positions on the behavior of the molecule. It is also highly likely that the precise residue involved in particular structures may be dependent on how the simulation is prepared. However, due to variety of chain lengths and boundary conditions as well as long equilibration times, we are confident that the results provide general insights into the transition mechanism.

### **Post Shear Equilibration and Stability**

After exposing the structures to shear, a representative structure with a high  $\beta$ -sheet content (in the domain, that has not been pulled) is further equilibrated for 50 ns both in vacuum and in explicit (TiP3P) water containing salt at 150 mM to measure the stability of the secondary structures induced by shearing. In vacuum, an isothermal-isochoric ensemble is used with cut-off for Coulombic and van-der-Waals interactions (3 nm), velocity-rescale thermostat and dispersion correction (accounting for the cut-off van-der-Waals scheme). The secondary structure is tracked using the DSSP algorithm.

### **Transition Probabilities**

In order to obtain a better understanding of the secondary structure transitions from helix/turn to sheet, the probability of being either in a helical/turn (spidroin) or in a sheet structure (after pulling and after further equilibration) is calculated residue by residue. Three probabilities are calculated whose product forms the joint probability. The first probability is for the initial equilibration, and is defined as the percentage of frames each residue is part of a  $\alpha$ -helix or turn/ $3_{10}$ -helix divided by the total number of frames, is calculated from the last 10 ns (out of 30 ns) of the equilibration trajectory. The second probability, during the pulling simulation, is calculated from the 1 ns of the simulation that defined the  $\beta$ -sheet content, see above. The third probability, from the equilibration after pulling, is calculated from the last 10 ns (out of 50 ns) of the simulation. The joint probability for each residue is calculated as the product of the three probabilities. This allows us to track which amino acid residues and which geometric relations are key to the transition. The relative location of the residues appearing in all three structures is compared using the Sybyl Tripos Biopolymer module and VMD. For comparison, we have performed this analysis for all structures ( $\alpha = 1,3,5$  as well as inter-chain and layered structures) and compiled the results in the supplementary material, Figures S8-S12 and Tables S1-S6.

## **Results and Discussion**

### **Size Dependence of Structural Transition and Shear Stress**

The spider silk dope is stored in the abdomen of the spider in globular form to prevent premature assembly and clogging of the spinning duct. In our simulation, the general form of the simulated protein structure is largely independent of the size of the molecule (left hand column, Figure 3). The spidroin remains unassembled and predominantly contains helical shapes ( $\alpha$ -helices, turns and  $3_{10}$ -helices), as observed in the deconvolution of Raman spectra.<sup>10, 56</sup> Our simulations show that helices towards the beginning of the sequence are typically arranged at a ca. 60 degree angle to one another.

During the shear induced assembly (second column, Figure 3) all helices and other structures transition into  $\beta$ -sheets or are destroyed, i.e., they transition into random coil. This is in agreement with experimental observations of the processes in the spinning duct.<sup>56</sup> Figure 2b and 2c show the simulation results from the shear simulation for a chain with three structural repeats,  $a = 3$ . Secondary structure transitions and associated shear stress for other chain lengths can be found in the supplementary material. We find the amount of  $\beta$ -sheet after the transition to be insensitive to pulling speed, in agreement with NMR measurements.<sup>55</sup> In addition, at higher reeling speeds, smaller crystals are formed which positively affects the toughness and ultimate strength of the fibers.<sup>24, 57</sup>

Figure 2b and 2c provide further valuable insights into the assembly mechanism of silk structures. The initial unfolding of the spidroin solution takes place under very low shear stress. This is expected from an  $\alpha - \beta$ -transition of relatively short chains.<sup>58</sup> The stress drop (at  $\sim 7$  ns in Figure 2c, see also supplementary material) is a phenomenon also observed when measuring forces during *in vivo* silking.<sup>59</sup> Here, we are able to link the drop in stress to the point where a

high  $\beta$ -sheet content in the structural transition is reached. This indicates that the creation of the crystal brings the molecule into a low energy state with the force providing the energy to overcome the barrier for state transition.

The secondary structure transition from  $\alpha$ -helical agglomerations to  $\beta$ -sheet structures is independent of the number of chain repeats  $a$  and also very robust with respect to the simulation setup. Figure 4a shows the average and maximum  $\beta$ -sheet content attained after the shear flow experiment (determined by the methodology shown in Figure 2b). The attained structure content is well within the range of experimental observation (20-25%,<sup>60</sup>; 36%,<sup>29</sup>). Figure 4b shows the associated transition shear stresses (determined by the methodology described in Figure 2c). For all chain lengths, the stress is above the elastic limit of assembled silk fibers (150-300 MPa). This agrees with experimental observation that reorganization of silk requires significant shear stresses and the average force during silking is higher than the conventional yielding force.<sup>61</sup>

We quantify the transition shear stress for silk assembly to be between 300 and 700 MPa. We note that initial secondary structure transitions takes place at lower stresses already,  $\sim$ 20% of the breaking stress. However, in order to reach the highest  $\beta$ -sheet content, the stress needs to be increased to up to 50% of the breaking stress. Experimental observations made by measuring the *in vivo* silking force directly on the spun fiber, identify the transition stress at 20-60% of the breaking stress (i.e., 300-850 MPa for *N. clavipes* silk).<sup>59</sup> Interestingly, predictions from flow simulations showed that a much smaller pressure gradient is needed to push the solution through the gland.<sup>62, 63</sup> However, such studies neglect the need for secondary structure transition and investigate the process solely from a fluid-mechanical and rheological point of view.

### **Structural Stability of the Assembled Molecules**

We determine the structural stability of the pulled protein, i.e., stability after the force-induced state transition, by equilibrating it for 50 nanoseconds in explicit water/ ions and also in vacuum, see third and fourth column in Figure 3. The initial structure of this ‘post-pulling’ equilibration is a  $\beta$ -sheet rich structure after the transition. The structures are considered stable if the  $\beta$ -sheet content does not change significantly during the simulation either in vacuum or in water.

In vacuum, irrespective of the chain size  $a$ , all structures are stable and retain their  $\beta$ -sheets, though the amorphous ends of all of chains show some compaction during the equilibration stage. In solvent (water with ions present at 150mM), the smaller silk structures (two and four poly-alanine stretches;  $a = 1, a = 3$ ) essentially return to their original largely disordered ‘spidroin-like’ state and only the larger structure (six poly-alanine stretches;  $a = 5$ ) remain relatively stable, Figure 4c. The natural condition will be an intermediate between fully hydrated and vacuum, due to the removal of water and ions from the silk as it is spun out of the spider and into air.<sup>64</sup> From a simulation point of view, such an intermediate state is not straightforward to model.

Figure 4d summarizes the secondary structure content of the spidroin and the simulated structures after equilibration in explicit solvent (data taken from 40-50 nanoseconds of the simulation). While the shorter chains start to form  $3_{10}$ -helices again and retain only a low percentage of  $\beta$ -sheets, the larger structure is stable and shows the secondary structure composition expected from assembled dragline silk fibers.<sup>56</sup> We anticipate that the assembly mechanism proposed here may work similarly for much larger structures. To test this hypothesis, we compare the stability of larger structures, i.e., stacks generated from two or three  $\beta$ -sheet layers ( $a = 3$ ). Irrespective of their orientation one to the other, or the number of such sheets, these structures are stable, see supplementary figure S6.

In summary, we have quantified the stress required to convert a largely disordered structure into a  $\beta$ -sheet structure by simulation and measured the minimum structural size required to stabilize the formed  $\beta$ -sheets. We have observed that the structures contain many more helices than those transitioning to  $\beta$ -sheets. Furthermore, we have shown that the minimum size for a stable structure is  $a = 5$  for single chain and  $a = 3$ , for multi-chain (see Supplementary Material). This result is supported by studies of genetically engineered silk repeat units without terminals, where the minimum number of repeats was  $a = 2$  to form indistinguishable nanofibrils, and with higher  $a$  the  $\beta$ -sheet content increased and fibril growth was facilitated.<sup>65</sup> How this transition is affected, i.e., whether there is a rationale as to which residues undergo the transition and why structures remain, is an unanswered question.

### **Molecular details of the transition mechanism**

To gain a more detailed understanding of the transition mechanism, we compare representative structures before pulling, after pulling, and after relaxation in vacuum and in water/ ions to investigate how a largely helical structure might transform into a stable  $\beta$ -sheet containing structure. We compute the probability of these structures containing helical or sheet content for each individual residue, see methods section, and plot the joint probability (blue line) for the residue to transition. Figure 5a shows the probability for the structure with  $a = 5$  to transition from  $\alpha$ -helix to  $\beta$ -sheet. Figure 5b shows the probability for the structure with  $a = 5$  to transition from  $3_{10}$ -helix/turn to  $\beta$ -sheet. This approach allows us to identify key residues of the transition, summarized in Table 1. From the transition probabilities ( $\alpha$ -helices - bold,  $3_{10}$ -helices/turns - italicized) we identify the residue numbers whose joint probability to transition is higher than 10%. These residue groups are the potential key players in the silk assembly mechanism of the core structure. For comparison, we have performed this analysis for all structures ( $a = 1,3,5$  as

well as inter-chain and layered structures) and compiled the results in the supplementary material.

Figure 5c depicts structure snapshots ( $\alpha = 5$ ) with highlighted residues identified from Figure 5a of the four transition stages, depicted in Figure 3. In spidroin state, Figure 5c(i), the structure consists principally of  $\alpha$ -helices (in this example, 93/206 residues), turns and a small number of residues in  $3_{10}$ -helices. This ‘exemplar’ structure contains four long helices (10-17 amino acids in length) and nine shorter helical sections. During the shear simulation, the structure comprises three principal sections of  $\beta$ -sheet structure (55/206 residues, 13-16 amino acids in length) plus two smaller (four amino acids) regions as well as three extensive coiled/ open structured sections, Figure 5c(ii). The  $\beta$ -sheet sections all align antiparallel to one another within a layer and also between layers.

In the vacuum equilibration after pulling, the structure comprises four sections of  $\beta$ -sheet (56/206 residues, six to seven amino acids in length), as well as four sections three to four amino acids in length, and seven sections two amino acids in length, Figure 5c(iii). Additionally, the structure has a helix (six amino acids) and turn structures, particularly towards the beginning of the sequence. The structure now has two coiled/ open structured sections. The  $\beta$ -sheet sections all align antiparallel to one another within a layer and also between layers.

In the fully solvated structure after equilibration (post-pulling) there are numerous short sections of  $\beta$ -sheet (48/206 residues), though with some residues in turns/  $3_{10}$  structures, Figure 5c(iv). A helix has formed from residues towards the start of the sequence and the ‘loops’ between the ‘pulled’ regions have compressed, leading to additional  $\beta$ -sheet structuring. The structure now comprises  $\beta$ -sheet that is largely formed by the poly-alanine regions (although usually not the whole length of the (Ala)<sub>6</sub> regions) and also has several extensive regions of ‘new/ induced’  $\beta$ -sheet that form after relaxation involving Glycine, Alanine and Glutamine



residues. There are still unfolded regions that sit on the outside of the structure. The  $\beta$ -sheet region is not as symmetrical in terms of the surrounding amorphous structure as in the vacuum relaxed structures though the  $\beta$ -sheet sections still all align anti-parallel to one another within a layer and also between layers. Some H-bonding can be observed between poly-alanine  $\beta$ -sheet regions for those sections that lie virtually parallel to one another. Note, whereas Tyrosine can be found in either helical or turn structures, it shows no propensity for involvement in the generation of  $\beta$ -sheet under conditions where the resultant structure is able to relax as in 'post-equilibration' in the solvated case. This seems to be the sole domain of the Alanine, Glycine, Leucine and Glutamine with only a single Arginine involved.

Analysis of the probability data, Figure 5a/b, shows that residues constantly in 'helical' or ' $\beta$ -sheet' state in the structures after initial equilibration, 'post-pulling', and equilibration in solvent (joint probability, blue line) largely arise from the more compact part of the initial structure. The sequences (four amino acids or longer) represented in the probability data, i.e., parts of the sequence that are most likely to transition, are the second half of the first large helix, the second half of the second large helix and the first half of the third large helix. The additional 'smaller common' segments are either close in sequence position to the longer helices/  $\beta$ -sheet sections or are physically close when the 3-d structure of the protein is considered. An exception to this is Arginine, which appears to sit 'between' several helices in the initial structure but is then close in space to the  $\beta$ -sheet structured elements after pulling and relaxation. This is the point in the sequence from which  $\beta$ -sheet regions develop in the relaxed solvated structure that were not present in the post-pulled structure.

In conclusion, stability of the  $\beta$ -sheet structure seems to arise from an initial close proximity of the helices. Relaxation in solvent, far from destroying the structures completely, leads to an apparent stabilization of shorter strands of  $\beta$ -sheet that can be mostly poly-alanine but may also

arise from other parts of the sequence involving a range of amino acids as has recently been shown from solid state NMR studies.<sup>41</sup>

## **Conclusion**

We have investigated the shear induced secondary structure transition process of the core MaSp1 sequence during the spider dragline silk assembly using molecular dynamics. We found robust results where a shear stress of the order of 20-50% of the failure stress induced an  $\alpha - \beta$ -transition in the poly-alanine region. Our model is validated through agreement to experimentally determined secondary structure and pulling forces of spider dragline silk. We were able to connect the reduction in stress observed in experiment and simulation to the point where a high  $\beta$ -sheet content, and a low energy state, in the structural transition is reached. While the transition stress is independent of the chain length, the minimum size for a stable structure was shown to be six poly-alanine regions for a single chain and four poly-alanine regions for multiple chains. This marks the smallest molecule size that gives rise to a 'silk-like' structure. While the poly-alanine region plays a key role in the transition from helix to sheet, other parts of the sequence (Glutamine, Arginine) may also be involved in the stabilization of the molecules. In general, the stability of the  $\beta$ -sheet structure seems to arise from a close proximity in space of the helices in the spidroin state. Our study emphasizes the role of shear in the assembly process of silk. The molecular level insights are valid for a broader range of biopolymers that self-assemble under stress. The determination of the critical shear stress can guide the design of microfluidic devices that attempt to mimic the natural spinning process, while the determined minimum molecule size and molecular configurations necessary for transition can provide design principles for the synthesis of bioinspired protein materials.

## **Acknowledgments**

The research was initiated while CCP was a visiting professor at the Laboratory for Atomistic and Molecular Mechanics (LAMM) at MIT, and the Edward, Frances and Shirley B. Daniels Fellow and Wyss Fellow at the Radcliffe Institute for Advanced Study, Harvard University (2012–2013). CCP is grateful for the extended use of the research computing facilities at Harvard University.

## **Supporting Information**

This material is available free of charge via the Internet at <http://pubs.acs.org>.

Convergence of the REMD algorithm, Secondary Structure and Shear Stress Trajectories, Spring stiffness and pulling (shear) speed, Layered Structures, Probability for  $\alpha - \beta$ -Transition

## **Corresponding Author**

\* Corresponding authors, E-mail: [mbuehler@mit.edu](mailto:mbuehler@mit.edu), [carole.perry@ntu.ac.uk](mailto:carole.perry@ntu.ac.uk)

## **Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## **Funding Sources**

TG and MJB acknowledge support from NIH (U01/EB0149), ONR (N00014-10-1-0562), AFOSR (FA9550-11-1-0199) and ARO (W991NF-09-1-0541). CCP acknowledges support from NIH (RO1-DE017207) and AFOSR (FA9550-1-13-0040).

## **References**

1. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J. S.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003**, 24, (3), 401-416.
2. Omenetto, F. G.; Kaplan, D. L., New Opportunities for an Ancient Material. *Science* **2010**, 329, (5991), 528-531.
3. Xia, X. X.; Qian, Z. G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y., Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, (32), 14059-14063.
4. Heidebrecht, A.; Eisoldt, L.; Diehl, J.; Schmidt, A.; Geffers, M.; Lang, G.; Scheibel, T., Biomimetic Fibers Made of Recombinant Spidroins with the Same Toughness as Natural Spider Silk. *Adv. Funct. Mater.* **2015**, 27, (13), 2189-2194.
5. Elices, M.; Guinea, G. V.; Plaza, G. R.; Karatzas, C.; Riekkel, C.; Agullo-Rueda, F.; Daza, R.; Perez-Rigueiro, J., Bioinspired Fibers Follow the Track of Natural Spider Silk. *Macromolecules* **2011**, 44, (5), 1166-1176.
6. Lewis, R. V., Spider silk: Ancient ideas for new biomaterials. *Chem. Rev.* **2006**, 106, (9), 3762-3774.
7. Sponner, A.; Schlott, B.; Vollrath, F.; Unger, E.; Grosse, F.; Weisshart, K., Characterization of the protein components of *Nephila clavipes* dragline silk. *Biochemistry* **2005**, 44, (12), 4727-4736.
8. Chaw, R. C.; Correa-Garhwal, S. M.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y., Proteomic Evidence for Components of Spider Silk Synthesis from Black Widow Silk Glands and Fibers. *J Proteome Res* **2015**, 14, (10), 4223-4231.
9. Lefevre, T.; Boudreault, S.; Cloutier, C.; Pezolet, M., Diversity of Molecular Transformations Involved in the Formation of Spider Silks. *J. Mol. Biol.* **2011**, 405, (1), 238-253.
10. Lefevre, T.; Leclerc, J.; Rioux-Dube, J. F.; Buffeteau, T.; Paquin, M. C.; Rousseau, M. E.; Cloutier, I.; Auger, M.; Gagne, S. M.; Boudreault, S.; Cloutier, C.; Pezolet, M., Conformation of spider silk proteins in situ in the intact major ampullate gland and in solution. *Biomacromolecules* **2007**, 8, (8), 2342-2344.
11. Jenkins, J. E.; Holland, G. P.; Yarger, J. L., High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft Matter* **2012**, 8, (6), 1947-1954.
12. Heim, M.; Keerl, D.; Scheibel, T., Spider Silk: From Soluble Protein to Extraordinary Fiber. *Angew. Chem. Int. Edit.* **2009**, 48, (20), 3584-3596.
13. Romer, L.; Scheibel, T., The elaborate structure of spider silk Structure and function of a natural high performance fiber. *Prion* **2008**, 2, (4), 154-161.
14. Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A.; Johansson, J.; Knight, S. D., Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **2010**, 465, (7295), 236-U125.

15. Jin, H. J.; Kaplan, D. L., Mechanism of silk processing in insects and spiders. *Nature* **2003**, 424, (6952), 1057-1061.
16. Gronau, G.; Qin, Z.; Buehler, M. J., Effect of sodium chloride on the structure and stability of spider silk's N-terminal protein domain. *Biomater. Sci.* **2013**, 1, (3), 276-284.
17. Gauthier, M.; Leclerc, J.; Lefevre, T.; Gagne, S. M.; Auger, M., Effect of pH on the Structure of the Recombinant C-Terminal Domain of Nephila clavipes Dragline Silk Protein. *Biomacromolecules* **2014**, 15, (12), 4447-4454.
18. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001**, 410, (6828), 541-548.
19. Knight, D. P.; Knight, M. M.; Vollrath, F., Beta transition and stress-induced phase separation in the spinning of spider dragline silk. *Int. J. Biol. Macromol.* **2000**, 27, (3), 205-210.
20. Greving, I.; Cai, M. Z.; Vollrath, F.; Schniepp, H. C., Shear-Induced Self-Assembly of Native Silk Proteins into Fibrils Studied by Atomic Force Microscopy. *Biomacromolecules* **2012**, 13, (3), 676-682.
21. Rammensee, S.; Slotta, U.; Scheibel, T.; Bausch, A. R., Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, (18), 6590-6595.
22. Keten, S.; Xu, Z. P.; Ihle, B.; Buehler, M. J., Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. *Nat. Mater.* **2010**, 9, (4), 359-367.
23. Giesa, T.; Pugno, N. M.; Wong, J. Y.; Kaplan, D. L.; Buehler, M. J., What's Inside the Box? - Length-Scales that Govern Fracture Processes of Polymer Fibers. *Adv. Mater.* **2014**, 26, (3), 412-417.
24. Du, N.; Liu, X. Y.; Narayanan, J.; Li, L. A.; Lim, M. L. M.; Li, D. Q., Design of superior spider silk: From nanostructure to mechanical properties. *Biophys. J.* **2006**, 91, (12), 4528-4535.
25. Rousseau, M. E.; Cruz, D. H.; West, M. M.; Hitchcock, A. P.; Pezolet, M., Nephila clavipes spider dragline silk microstructure studied by scanning transmission X-ray microscopy. *J. Am. Chem. Soc.* **2007**, 129, (13), 3897-3905.
26. An, B.; Hinman, M. B.; Holland, G. P.; Yarger, J. L.; Lewis, R. V., Inducing beta-Sheets Formation in Synthetic Spider Silk Fibers by Aqueous Post-Spin Stretching. *Biomacromolecules* **2011**, 12, (6), 2375-2381.
27. Rising, A.; Johansson, J., Toward spinning artificial spider silk. *Nat. Chem. Biol.* **2015**, 11, (5), 309-315.
28. Lefevre, T.; Boudreault, S.; Cloutier, C.; Pezolet, M., Conformational and orientational transformation of silk proteins in the major ampullate gland of Nephila clavipes spiders. *Biomacromolecules* **2008**, 9, (9), 2399-2407.
29. Lefevre, T.; Rousseau, M. E.; Pezolet, M., Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. *Biophys. J.* **2007**, 92, (8), 2885-2895.
30. Jenkins, J. E.; Creager, M. S.; Butler, E. B.; Lewis, R. V.; Holland, G. P.; Yarger, J. L., Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk. *Chem. Commun.* **2010**, 46, (36), 6714-6716.
31. Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Holland, G. P.; Yarger, J. L., Quantitative Correlation between the Protein Primary Sequences and Secondary Structures in Spider Dragline Silks. *Biomacromolecules* **2010**, 11, (1), 192-200.
32. Sampath, S.; Isdebski, T.; Jenkins, J. E.; Ayon, J. V.; Henning, R. W.; Orgel, J. P. R. O.; Antipoa, O.; Yarger, J. L., X-ray diffraction study of nanocrystalline and amorphous structure within major and minor ampullate dragline spider silks. *Soft Matter* **2012**, 8, (25), 6713-6722.
33. Ma, B. Y.; Nussinov, R., Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Sci.* **2002**, 11, (10), 2335-2350.

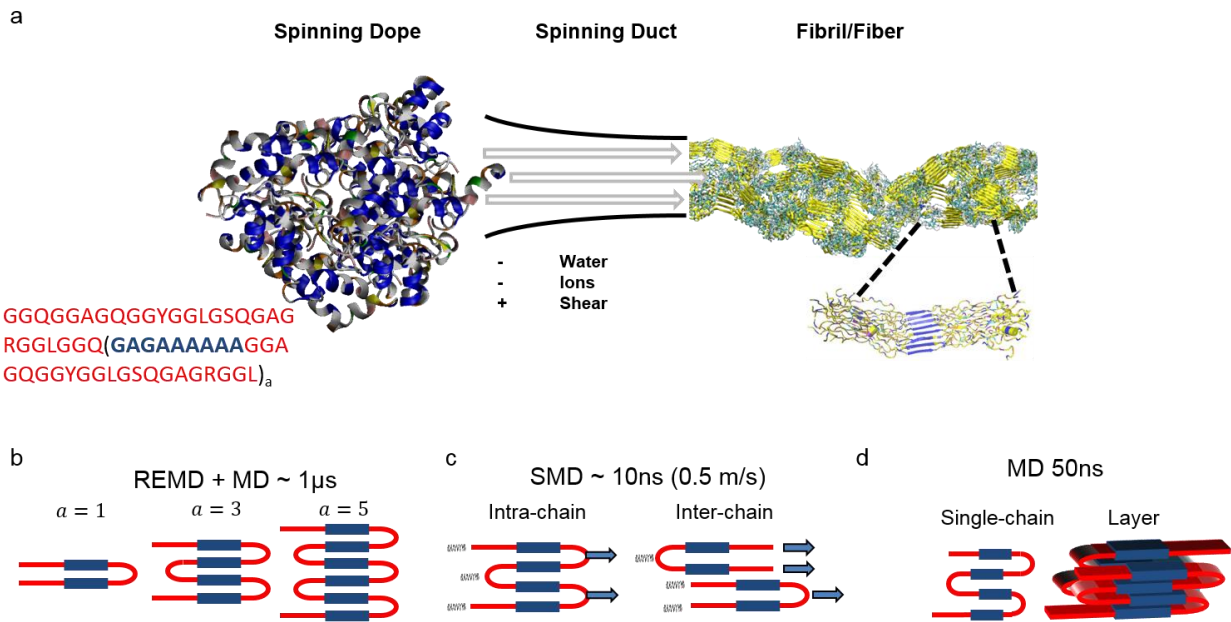
34. Keten, S.; Buehler, M. J., Nanostructure and molecular mechanics of spider dragline silk protein assemblies. *J. R. Soc. Interface* **2010**, 7, (53), 1709-1721.
35. Nova, A.; Keten, S.; Pugno, N. M.; Redaelli, A.; Buehler, M. J., Molecular and Nanostructural Mechanisms of Deformation, Strength and Toughness of Spider Silk Fibrils. *Nano Lett.* **2010**, 10, (7), 2626-2634.
36. Giesa, T.; Arslan, M.; Pugno, N. M.; Buehler, M. J., Nanoconfinement of spider silk fibrils begets superior strength, extensibility and toughness. *Nano Lett.* **2011**, 11, (11), 5038-5046.
37. Giesa, T.; Buehler, M. J., Nanoconfinement and the Strength of Biopolymers. *Annu. Rev. Biophys.* **2013**, 42, 651-673.
38. Koski, K. J.; Akhenblit, P.; McKiernan, K.; Yarger, J. L., Non-invasive determination of the complete elastic moduli of spider silks. *Nat Mater* **2013**, 12, (3), 262-267.
39. Michal, C. A.; Jelinski, L. W., Rotational-echo double-resonance in complex biopolymers: a study of *Nephila clavipes* dragline silk. *J. Biomol. NMR* **1998**, 12, (2), 231-241.
40. van Beek, J. D.; Hess, S.; Vollrath, F.; Meier, B. H., The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *P. Natl. Acad. Sci. U. S. A.* **2002**, 99, (16), 10266-10271.
41. Jenkins, J. E.; Sampath, S.; Butler, E.; Kim, J.; Henning, R. W.; Holland, G. P.; Yarger, J. L., Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction. *Biomacromolecules* **2013**, 14, (10), 3472-3483.
42. Kummerlen, J.; van Beek, J. D.; Vollrath, F.; Meier, B. H., Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules* **1996**, 29, (8), 2920-2928.
43. Plaza, G. R.; Perez-Rigueiro, J.; Riekkel, C.; Perea, G. B.; Agullo-Rueda, F.; Burghammer, M.; Guinea, G. V.; Elices, M., Relationship between microstructure and mechanical properties in spider silk fibers: identification of two regimes in the microstructural changes. *Soft Matter* **2012**, 8, (22), 6015-6026.
44. Paquet-Mercier, F.; Lefevre, T.; Auger, M.; Pezolet, M., Evidence by infrared spectroscopy of the presence of two types of beta-sheets in major ampullate spider silk and silkworm silk. *Soft Matter* **2013**, 9, (1), 208-215.
45. Fossey, S. A.; Tripathy, S., Atomistic modeling of interphases in spider silk fibers. *Int. J. Biol. Macromol.* **1999**, 24, (2-3), 119-125.
46. Bratzel, G.; Buehler, M. J., Sequence-structure correlations and size effects in silk nanostructure: Poly-Ala repeat of *N. clavipes* MaSp1 is naturally optimized. *J. Mech. Behav. Biomed. Mat.* **2012**, 7, 30-40.
47. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, 314, (1-2), 141-151.
48. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoseck, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, 30, (10), 1545-1614.
49. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, 26, (16), 1781-1802.

50. Feig, M.; Karanicolas, J.; Brooks, C. L., MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **2004**, *22*, (5), 377-395.
51. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, (19), 8577-8593.
52. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, (12), 2577-2637.
53. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* **1996**, *14*, (1), 33-38.
54. Sotomayor, M.; Schulten, K., Single-Molecule Experiments in Vitro and in Silico. *Science* **2007**, *316*, (5828), 1144-1148.
55. Cloutier, I.; Leclerc, J.; Lefevre, T.; Auger, M., Solid-state nuclear magnetic resonance (NMR) spectroscopy reveals distinctive protein dynamics in closely related spider silks. *Can. J. Chem.* **2011**, *89*, (9), 1047-1054.
56. Lefevre, T.; Paquet-Mercier, F.; Rioux-Dube, J. F.; Pezolet, M., Structure of silk by raman spectromicroscopy: From the spinning glands to the fibers. *Biopolymers* **2012**, *97*, (6), 322-336.
57. Shao, Z. Z.; Vollrath, F., Materials: Surprising strength of silkworm silk. *Nature* **2002**, *418*, (6899), 741-741.
58. Qin, Z.; Buehler, M. J., Molecular Dynamics Simulation of the alpha-helix to beta-sheet Transition in Coiled Protein Filaments: Evidence for a Critical Filament Length Scale. *Phys. Rev. Lett.* **2010**, *104*, (19), 198304.
59. Ortlepp, C. S.; Gosline, J. M., Consequences of forced silking. *Biomacromolecules* **2004**, *5*, (3), 727-731.
60. Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N., The mechanical design of spider silks: From fibroin sequence to mechanical function. *J. Exp. Biol.* **1999**, *202*, (23), 3295-3303.
61. Perez-Rigueiro, J.; Elices, M.; Plaza, G.; Real, J. I.; Guinea, G. V., The effect of spinning forces on spider silk properties. *J. Exp. Biol.* **2005**, *208*, (14), 2633-2639.
62. Breslauer, D. N.; Lee, L. P.; Muller, S. J., Simulation of Flow in the Silk Gland. *Biomacromolecules* **2009**, *10*, (1), 49-57.
63. Kojic, N.; Bico, J.; Clasen, C.; McKinley, G. H., Ex vivo rheology of spider silk. *J. Exp. Biol.* **2006**, *209*, (21), 4355-4362.
64. Kojic, N.; Kojic, M.; Gudlavalleti, S.; McKinley, G., Solvent removal during synthetic and Nephila fiber spinning. *Biomacromolecules* **2004**, *5*, (5), 1698-1707.
65. Humenik, M.; Magdeburg, M.; Scheibel, T., Influence of repeat numbers on self-assembly rates of repetitive recombinant spider silk proteins. *J. Struct. Biol.* **2014**, *186*, (3), 431-437.

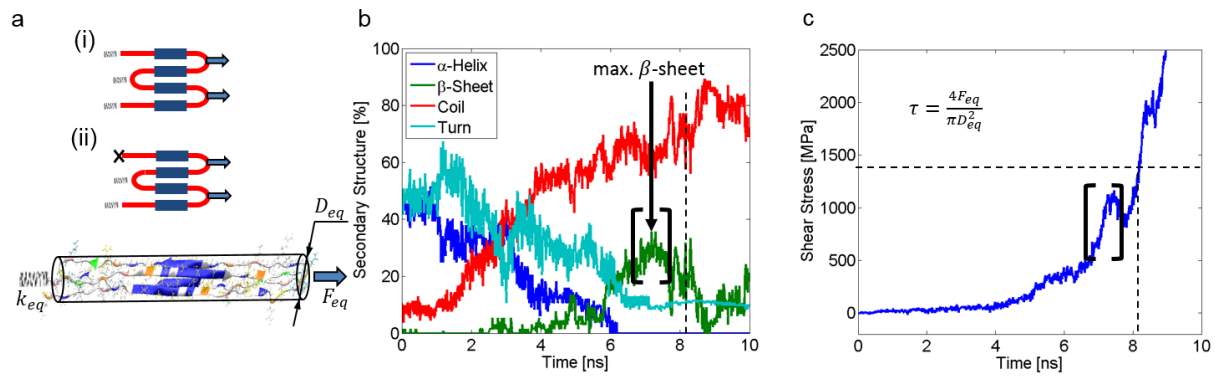
**Table 1.** Residues involved in the structural transition of silk chains with length  $a = 5$  (206 residues).

<b>Residue Number</b>	<b>Residue Name</b>
<b>29-36</b>	AAAAGGA
<i>49</i>	<i>A</i>
<b>54</b>	L
<b>57-65</b>	QGAGAAAA
<i>65</i>	<i>A</i>
<i>89-90</i>	<i>GQ</i>
<b>95-99</b>	AAAAA
<i>103-104</i>	<i>GQ</i>
<b>161-162</b>	AA
<i>173</i>	<i>Y</i>
<b>180</b>	R

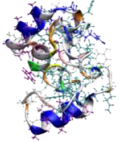
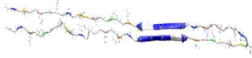

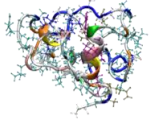
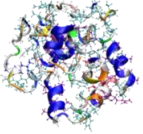
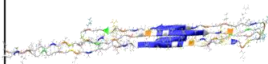
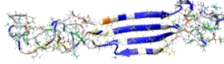
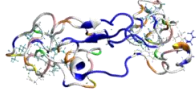
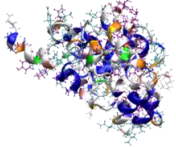
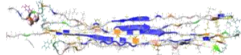

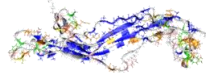




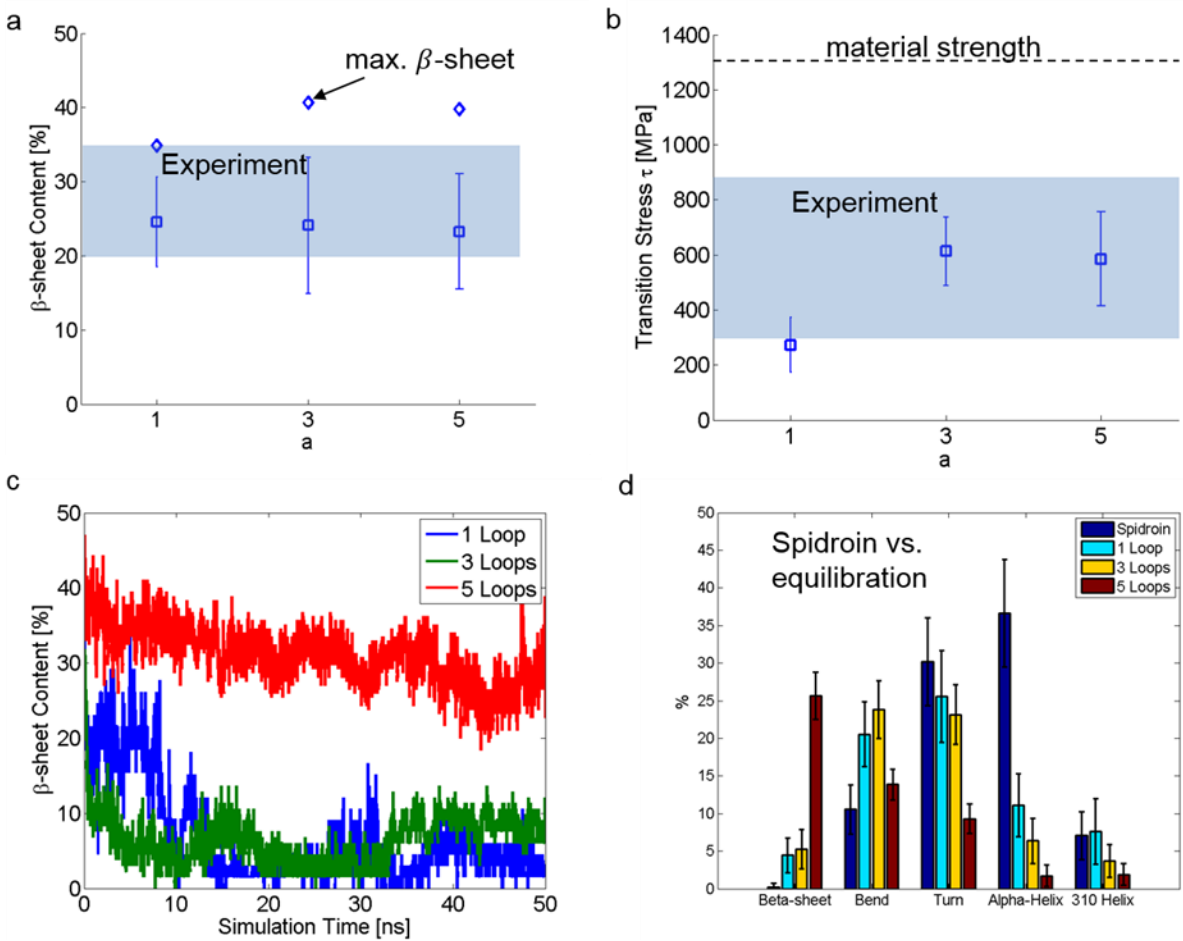
**Figure 1.** Molecular model of the silk fiber assembly process. (a) Silk is processed from spidroin to a solid fiber in the spinning duct under ambient conditions. Shear stresses at the wall together with the removal of water from the protein lead to the formation of a nano-composite having an aligned,  $\beta$ -sheet rich crystalline phase. A change of ionic conditions during the spinning process is believed to lead to a conformational change in the terminal regions of the silk protein. (b) Model of the silk spidroin in equilibrium. Using Replica Exchange Molecular Dynamics for a total of one microsecond the equilibrium structure of the protein was determined for different chain lengths ( $a = 1,3,5$ ), here denoted as poly-alanine regions (blue) and remainder of the intervening sequence (red). (c) Model of the silk assembly process. Using Steered Molecular Dynamics at natural pulling speeds we model a shear flow and determine the secondary structure transitions as well as shear stresses in the silk protein chains. We investigate intra- as well as inter-chain interactions. (d) Model of the final fiber structure. After shearing the assembled structure is simulated in solvent and vacuum to test the stability of the structure.



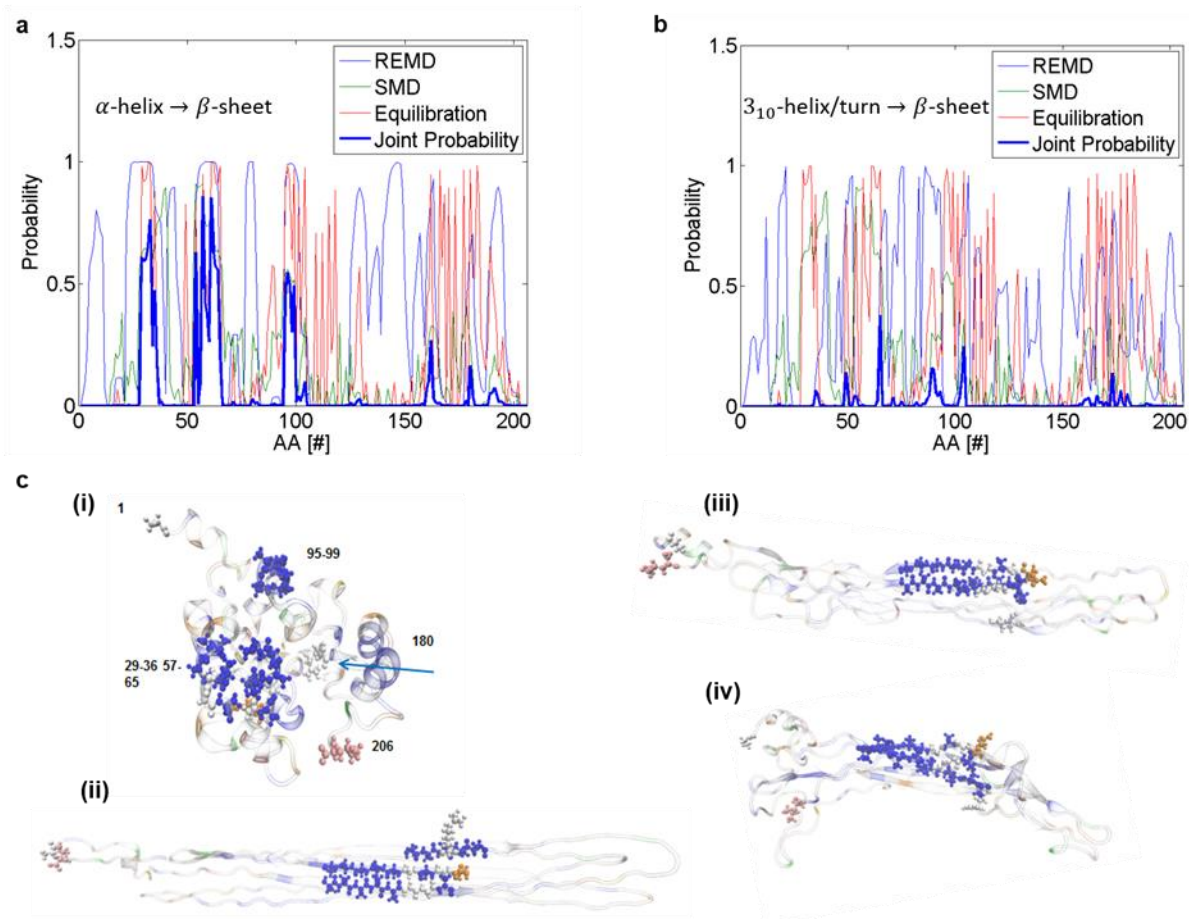
**Figure 2.** Secondary structure transition and shear stress during the assembly of silk. (a) Two different boundary conditions (i) and (ii), are tested to investigate the trajectory of secondary structure and shear stress. The part of the sequence colored in blue is the alanine rich region and the rest of the sequence colored in red. The pulling force is determined with the Steered Molecular Dynamics algorithm. (b) Secondary structure trajectory during the pulling simulation in explicit solvent. The graph shown is one out of four tests for  $\alpha = 3$  (with boundary condition (i)). The  $\beta$ -sheet content after pulling for each structure is determined by averaging the  $\beta$ -sheet content for one nanosecond around the maximum content as indicated in the figure. (c) Shear stress associated with the secondary structure transition. The transition shear stress is determined from the pulling force divided by the effective area, indicated in panel a. The stress is averaged in the same time period as that taken to assess the  $\beta$ -sheet content. In all cases, the structure transition happens prior to reaching the strength limit of the material (indicated by the dashed lines).

	<b>1. Equilibration</b> Replica Exchange + MD Water + Ions 1 $\mu$ s	<b>2. Shear Flow</b> Steered Molecular Dynamics 10ns	<b>3a. Equilibration</b> MD Vacuum 50ns	<b>3b. Equilibration</b> MD Water + Ions 50ns
1 Loop ( $a = 1$ )				
3 Loops ( $a = 3$ )				
5 Loops ( $a = 5$ )				

**Figure 3.** Model structures of the stages during the assembly process. The spider silk dope is stored in the abdomen of the spider in globular form. In the simulation, the general form of the simulated peptide/ protein structure is largely independent of the size of the molecule (left hand column). During the shearing process (second column) a large part of the structure transitions into  $\beta$ -sheet structures. Subsequently, equilibration of the structure in vacuum or explicit water in the presence of ions, the silk relaxes again (third and fourth column respectively). The natural condition can be assumed to be an intermediate between fully hydrated and vacuum due to the removal of water and ions from the silk as it is spun into air.



**Figure 4.** (a) The  $\beta$ -sheet content attained after the shear flow experiment versus chain length and comparison with experimental observation. (b) Transition shear stress versus chain length for the silk assembly (c) Stability of the silk chains ( $\beta$ -sheet content) after shearing and a further equilibration in explicit solvent. (d) Summary of the secondary structure content of the spidroin and the simulated structures after equilibration in explicit solvent.



**Figure 5.** Transition probabilities and structural details. (a) The graph shows the probability for each of the 206 residues of the structure with  $a = 5$  to be in  $\alpha$ -helical state (after REMD), in  $\beta$ -sheet state (after SMD), and in  $\beta$ -sheet state (after Equilibration). The dark blue line is the joint probability defined as the product of the three probabilities and indicates the residues that transition from initial helical to final sheet structure. (b) Probability for each of the 206 residues of the structure with  $a = 5$  to be in  $3_{10}$ -helical/turn state (after REMD) and in  $\beta$ -sheet state (after SMD/after Equilibration). (c) Structure snapshots ( $a = 5$ ) with highlighted residues identified from (a) of the four transition stages (i - iv), in agreement with Figure 3. Only the sections with larger structural segments ( $\geq 4$  residues) are shown. Blue represents Alanine, white represents Glycine, brown represents Glutamine and aqua-color represents Arginine, compare to Table 1.