Unfolding the multi-length scale domain structure of silk fibroin protein

Hennady Shulha [a], Cheryl Wong Po Foo [b,c], David L. Kaplan [b,c,*], Vladimir V. Tsukruk [a,d,*]

[a] *Materials Science and Engineering Department, Iowa State University, 3155 Gilman, Ames, IA 50011, USA*
[b] *Department of Biomedical Engineering, Tufts University, Medford, MA 02155, USA*
[c] *Department of Chemistry, Tufts University, Medford, MA 02155, USA*
[d] *School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

Abstract

Multi-scale force spectroscopy was applied to measure unfolding properties and the internal domain structure of *Bombyx mori* silk fibroin. We demonstrated that the complex multi-domain sequence and block design in this protein can be directly related to multi-stage unfolding behavior of the specific regions through the use of force extension measurements. These new findings suggest relationships between polymer block chemistry and mechanical features, as origins of the remarkable mechanical properties of native silk fibers in general. We observed multiple consequential unfolding of hydrophilic and hydrophobic domains with characteristics that can be directly related to known molecular dimensions of the protein backbones. Future screening and selection approaches can be envisioned to exploit this approach to optimize specific material functional features for both biopolymers and synthetic polymers in relation to sequence chemistry, a capability not currently available.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Atomic force microscopy; Surface force spectroscopy; Silk fibroin protein

1. Introduction

Silks have been utilized in biomedical materials as sutures for decades, and recent efforts have focused on extending the utility in many biomedical applications as well as general materials needs [1,2]. The scientific and technological interest in this family of proteins derives from the unique mechanical and functional properties of the native fibers harvested from cocoons in the case of silkworm silks and orb webs in the case of spiders. Silks in fiber form are generally considered the strongest natural fibers known and rival even the best synthetic fibers that can be made today [3,4]. Furthermore, the fibers also perform well in compression, unlike most high performance synthetic fibers [5]. The processes by which such remarkable mechanical properties are derived have been the subject of intensive study, since the formation of these fibers occurs under relatively mundane conditions in water at ambient conditions of temperature and pressure.

Fibroin from *Bombyx mori* is an interesting object for study since the entire sequence chemistry is known [6] and the chains are monodisperse due to genetic templating in protein synthesis. This knowledge permits for the first time direct correlation between backbone sequence and block chemistry and the corresponding interpretation of force extension measurements at different spatial scales. When this opportunity is considered alongside the remarkable mechanical properties of the fibers formed from this protein, further impact of these results is expected in guiding future strategies to mimic such remarkable materials from Nature but with synthetic analogs.

The goal of the present study was to reveal the multi-domain hydrophobic–hydrophilic structure of *B. mori* fibroin in the course of its forced unfolding. A multi-length scale surface force spectroscopy (SFS) approach was performed to reveal unfolding scenario at different extension lengths. SFS provides information about persistence length, number of domains, unfolding forces, molecular spring constants, and

* Corresponding authors. Tel.: +1 515 294 6904; fax: +1 515 294 5444.
  E-mail addresses: david.kaplan@tufts.edu (D.L. Kaplan), vladimir@iastate.edu (V.V. Tsukruk).

total length of the protein backbones under different strain and spatial-length conditions [7–10]. The presence of hydrophilic and hydrophobic domains of different sizes and with predictably different molecular spring constants should be revealed by stretching individual protein molecules on different spatial scales. The ultimate goal was to correlate these features with the sequence chemistry of the protein chains to generate the first detailed look into the influence of sequence chemistry on chain interactions and ultimately mechanical properties. This study also establishes new groundwork with which to explore sequence variants and how they impact mechanical behavior at multi-length scales, leading to design templates for synthetic analogs that can be more rationally considered for future high performance fibers and other material needs.

The most intensively studied silk protein is fibroin from the domesticated silkworm, *B. mori*. The structural portion of the fibers spun by the silkworm to form its cocoon consists of two proteins, a heavy chain fibroin ($\sim$390 kDa) and light chain fibroin ($\sim$25 kDa), linked by a single disulfide bond. The heavy chain fibroin is dominated by large stretches of hydrophobic amino acids, particularly glycine–alanine repeats, that form β-sheet crystallites in the spun fibers. The chemical composition of the silk protein and possible conformations of this protein and conformational transitions have been recently discussed [11,12]. In recent studies we have reported on the unique triblock design of this protein that provides a model for how this hydrophobic protein can be processed in an all aqueous environment, avoiding premature crystallization that would otherwise be catastrophic for the animal [6,13]. In this triblock design, the fibroin heavy chain consists of three types of blocks: (1) large hydrophilic chain ends – at both the N- (151 amino acids) and C-termini (50 amino acids), responsible in part for the intriguing surfactancy of the protein; (2) large hydrophobic blocks, a total of 11 or 12 if the comparatively shorter hydrophobic block (45 amino acids long) upstream of the C-terminus is included, that range in size from 159 to 607 amino acids, with all but two of these blocks above 300 amino acids in size, that comprise the majority of the protein and are dominated by the glycine–alanine repeats; and (3) small hydrophilic spacers, usually 32 amino acids in length, a total of 11 spacers, that separate the hydrophobic blocks but are smaller than the large hydrophilic chain ends, responsible for interactions within to prevent crystallization of the hydrophobic domains until spinning (Table 1). Estimates of geometrical dimensions from molecular models are also included.

The question of internal domain structure of *B. mori* silk proteins relative to their contributions to mechanical displacement and the unique properties of silk materials remains an open issue. Although some results for various silk proteins suggest the possibility for a multi-stage backbone unfolding upon stretching and cooperative motion of network elements, no comprehensive experimental evidence of such events exists for silk proteins despite the critical role of block sequence designs in the assembly and function of these proteins into robust and remarkable fibers [13]. Furthermore, insight into these features should permit direct links between sequence chemistry (backbone blocks) and mechanical features (materials level) to

be established and then mimicked; insights not currently attainable due to limitations in the control of chain sequence and length in the synthesis of synthetic polymers; and insufficient mechanical details discerned from single chain force extension curves to date.

2. Experimental

2.1. Materials

2.1.1. Sample preparation of B. mori silk fibroin solution for tapping mode AFM

B. mori silk fibroin solution was prepared by standard protocols [2] and maintained at 4 °C to reduce premature aggregation [14]. A chemically reduced solution was also prepared to disrupt the disulfide bond between the light and heavy chains with 2 mM DTT. Atomically smooth silicon wafers were used as substrates treated in an ultrasonic bath for 10 min followed by a "piranha" solution (30% concentrated hydrogen peroxide, 70% concentrated sulfuric acid) bath for 1 h, rinsing several times with Nanopure water (18 MΩ cm) and drying under nitrogen. Microroughness of wafers was less than 0.1 nm.

Tapping mode AFM imaging was conducted with a Multimode Nanoscope IIIa microscope (Digital Instruments) in ambient air ($\sim$50% humidity) according to our standard procedures [15]. The light tapping mode was used to control normal forces to be below 0.1 nN to avoid deformation and damage of adsorbed protein. The tip radius was measured independently using a gold nanoparticle standard with nanoparticle diameter within 3–6 nm according to the known procedure [16]. The radius of curvature of the tips was calculated from high-resolution cross-sections in spherical approximation. To calculate actual molecular dimensions, the known tip radius was used to solve a linear equation and evaluate true lateral dimensions of globular and cylindrical objects [17].

Protein solutions were placed on freshly prepared silicon wafers and excess solution was removed with a stream of dry air after 2 min of immobilization to freeze initial adsorbed structure of collapsed and compacted protein chains. The surface layer formed was rinsed with Nanopure water and imaged in air in light tapping mode after drying. Spin-casting was conducted from 0.01% protein solution. For time-dependent studies, the silicon wafers containing adsorbed protein were stored in closed vials under normal lab conditions. Several randomly selected locations were imaged at different times with scan sizes from 13 to 1 μm.

2.2. Methods

2.2.1. Estimated silk domain lengths

Estimates of geometrical dimensions from molecular models (end-to-end distances) for all major scenarios were calculated as follows: folded or coiled state (calculated taken the persistent length, number of amino acids, and Gaussian statistics [18]), extended state (α-helix for hydrophobic domains and all-trans zigzag conformation for hydrophilic domains, all oriented in one direction), and fully extended

Table 1
Block domains of fibroin heavy chain of the silkworm *Bombyx mori* and their estimated lengths

		Amino acids			Molecular lengths (nm)		
		Total #	# Positive charges	# Negative charges	Extended	α-Helical	Coiled
Domains (N- to C-terminal)		151	12	20	52.85		3.6
Hydrophobic domain 1		521	0	0	182.35	78.15	6.3
Hydrophilic domain 1		33	1	3	11.55		1.7
Hydrophobic domain 2		521	0	0	182.35	78.15	6.3
Hydrophilic domain 2		32	0	3	11.2		1.7
Hydrophobic domain 3		371	0	0	129.85	55.65	5.9
Hydrophilic domain 3		32	0	3	11.2		1.7
Hydrophobic domain 4		159	0	0	55.65	23.85	3.5
Hydrophilic domain 4		32	0	3	11.2		1.7
Hydrophobic domain 5		439	0	0	153.65	65.85	5.8
Hydrophilic domain 5		32	0	3	11.2		1.7
Hydrophobic domain 6		287	0	0	100.45	43.05	4.7
Hydrophilic domain 6		32	0	3	11.2		1.7
Hydrophobic domain 7		607	0	0	212.45	91.05	6.8
Hydrophilic domain 7		32	0	3	11.2		1.7
Hydrophobic domain 8		507	0	0	177.45	76.05	6.2
Hydrophilic domain 8		32	0	3	11.2		1.7
Hydrophobic domain 9		360	0	0	126.	54.0	5.2
Hydrophilic domain 9		31	0	3	10.85		1.6
Hydrophobic domain 10		315	0	0	110.25	47.25	4.9
Hydrophilic domain 10		32	0	3	11.2		1.7
Hydrophobic domain 11		575	0	0	201.25	86.25	6.6
Hydrophilic domain 11		35	3	4	12.25		1.75
Hydrophobic domain 12		45	1	0	15.75	6.75	1.9
C-terminal domain		50	9	1	17.5		2.1
All domains		5263	26	55	L_1=1842	L_2=730 L_3=900	L_4=89 L_5=258

L_1: fully extended conformation, both types of blocks; L_2: oriented α-helical hydrophobic + coiled hydrophilic; L_3: oriented α-helical hydrophobic + fully extended hydrophilic; L_4: both α-helical hydrophobic and hydrophilic coiled; and L_5: coiled α-helical hydrophobic + extended hydrophilic.

state (all-trans zigzag conformation for both hydrophobic and hydrophilic domains, all oriented in one direction).

2.2.2. Surface force spectroscopy measurements

SFS measurements were conducted with a Multimode™ microscope equipped with Picoforce™ mode and Dimension 3000™ (Digital Instruments) with Ultrasharp CSC11 silicon cantilevers as well as gold-coated cantilevers from Mikro-Masch. Spring constants were in the range 0.06–0.3 N/m determined by thermal resonance [19] or resonance frequency calibration plots [20]. Before the experiment, 10–15 μL of fibroin solution was deposited on the substrate and dried under dry nitrogen. SFS experiments were conducted immediately by placing a droplet of Nanopure water and submerging the AFM tip. Before the measurements, the AFM tip submerged into fluid was brought into close proximity to the protein layer for 30 min. SFS experiments were conducted in a fluid cell at several randomly selected locations for each specimen. SFS data were collected as an array of force–distance curves either as a series of individual measurements (Picoforce™) or in a force–volume mode (Dimension™). An array of curves collected from a selected surface area (size ranging from 1 to 10 μm) typically contained 256 curves (16 × 16). Z-velocity was varied from 0.4 to 6 nm/ms with the z-range varying within 200–300 nm, 1–2 μm, and 2–3.5 μm for three separate experimental setups. These different stretching gaps were selected on the basis of characteristic lengths of protein backbones in different conformations suggested by the theoretical model. SFS experiments were repeated for different solution concentrations with new tips for each measurement with a total number of force curves collected for all samples approaching 10,000 and with about 150 of them selected for further analysis using usual criteria for separation of multi-chain and single chain events [21,22].

To test the consistency of the data collected under different conditions, additional SFS measurements were done with different setups by using silicon, gold-coated, and amine-functionalized AFM tips, different solution concentrations (0.001, 0.01, 0.1%), different substrates (bare and amine-terminated silicon), and by conducting measurements either in water or in the protein solution. The data obtained with different tips, under different protein adsorption conditions, and at different locations, showed consistent results and all have been used in the statistical analysis.

3. Results and discussion

3.1. Fibroin adsorption on hydrophilic surface

Adsorbed silk proteins formed ultrathin uniform layers on bare silicon with occasionally occurring patches and uncovered surface areas. Surface morphology displayed both individual globular shapes and aggregated nanofibrils composed of several dozens of densely packed near-circular domains (Fig. 1). The surface morphology immediately after adsorption was dominated by individual circular features with fibrillar aggregates observed mainly along the boundaries between completely covered and bare surface areas. The height of these surface nanostructures was 3.1 ± 0.9 nm for nanofibrils and 1.6 ± 0.5 nm for individual circular features. The average diameter of the individual features of 5.4 ± 2 nm was estimated from AFM images accounting for the tip dilation and using semi-spherical approximation for the tip with the known radius of curvature as described in Section 2. Here, we have to mention that the estimation of molecular dimensions from the tapping AFM images in ambient air should be done very carefully considering several common artifacts. First, partial dehydration of protein molecules in humid air could result in change in overall shape of adsorbed protein molecules. However, applying fast drying by a stream of dry nitrogen should "freeze" the original global shape even if individual

domains will be compacted. Second, such dense compact packing of molecular segments should affect their molecular dimensions as compared to initial state in solution. Thus, the experimental molecular dimensions and volume (after correction for the tip dilation) should be compared with theoretical molecular volume calculated from molecular weight and packing density typical for loosely packed bulk material. Third, tapping AFM imaging in air could result in squashing of soft protein adsorbed onto a solid substrate. To this end, we applied very gentle, so-called "light" tapping with the ratio of damped to free amplitude higher than 0.95 [23]. Under these scanning conditions, normal forces applied to soft surfaces do not exceed 0.1 nN which could result in their very modest (below 10%) deformation thus, introducing insignificant error.

Thus, a true volume of an individual circular feature estimated with all possible precautions was close to 40 nm^3. This value is many-fold smaller than the estimated molecular volume of a single protein molecule with molecular weight of 390 kDa with a predominantly α-helical backbone compacted into a globular shape based on prior models (180 nm^3) [24]. It is also much smaller than the more realistic estimation of a volume occupied by a loosely packed molecule in a dense disordered state with the packing coefficient of 0.6 (typical for amorphous-like macromolecular materials) (450 nm^3) or a swollen molecule with a high concentration of bound water as determined from light scattering (3×10^4 nm^3) [25].
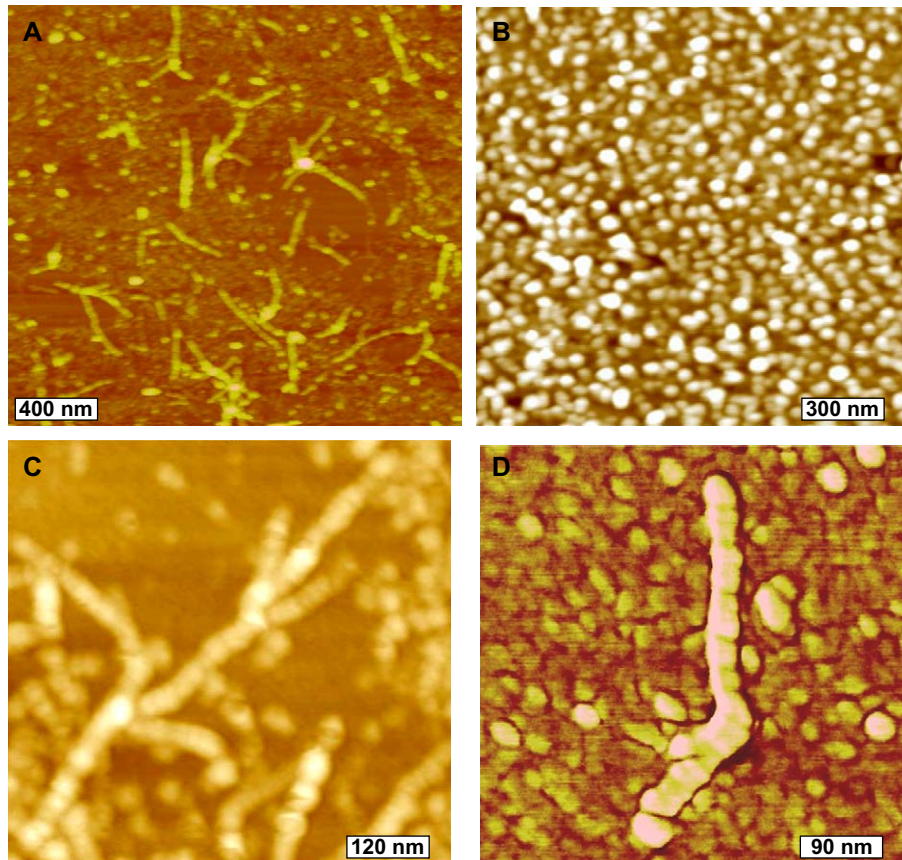

Fig. 1. Topographical AFM images of silk fibroin proteins: (A) near the edge of the protein layer showing globular and fibrillar aggregates; (B) high-resolution images of densely packed globular aggregates; (C) fibrillar structures; and (D) an individual fibril.

These significant differences exclude the possibility of forming an individual globule by a single protein molecule and suggest that a single molecule should extend over several globular structures, effectively bridging several circular features. On the other hand, comparison of the experimental value of 40 nm^3 with the volume occupied by a single hydrophobic domain (estimated from chemical composition to be within 10–40 nm^3 for different domains) points out their similarity. Thus we can suggest that each globular structure is composed of a single or several (fewer than 3–4) hydrophobic domains which can belong to either the same or different protein backbones. Smaller hydrophilic segments which connect the bulkier hydrophobic domains (Table 1) should spread beneath and in-between these globular aggregates maximizing their interfacial contact and minimizing the contact area of the hydrophobic domains with highly hydrophilic silicon oxide surface similarly to that suggested for synthetic hydrophobic–hydrophilic block copolymers [26]. Considering that the silk protein molecules are composed of 12 bulky hydrophobic blocks, this result also indicates that a single flexible protein molecule should propagate through at least 4–10 globular aggregates, effectively chaining them with hydrophilic segments. This is an important conclusion which establishes an understanding of the evolution of surface morphology.

Individual fibrillar aggregates occasionally observed immediately after adsorption and prevailing at longer times occasionally form straight and curved nanofibrils with lengths within 150–400 nm and heights of 3.1 nm. These nanofibrils are composed of 15 nearly round aggregates with spacing 15–25 nm (Fig. 1). Considering the diameter of these fibrils to be close to 11 nm, the volume of a single "bead" composing these fibrils is within 450–650 nm^3, much higher than the volume of the individual globular features discussed above. The average length of the fibrils is remarkably close to that expected for the molecular length with a coiled conformation of α-helical hydrophobic domains and extended hydrophilic domains (Table 1). Comparing the volume of these "beads" with the molecular volume of the hydrophobic domains suggests an aggregation number of 20–30. These two results indicate that the fibrils are composed of 20–30 aggregated protein molecules in coiled hydrophobic–hydrophilic domains. This estimate is close to 20–35 segments per domain for similar beaded fibrils formed by spider dragline silk molecules [21]. Overall surface morphology remained similar for various fibroin solution concentrations with slightly higher packing density observed for the more concentrated solutions. General surface morphology observed was consistent with other reports with similar silk fibroin proteins where individual molecules, individual nanofibrils, and branched fibrillar aggregates were observed on different surfaces [24,27].

Considering the complex aggregation behavior of the fibroin molecules and variety of surface morphologies, the temporal evolution of the surface morphology after deposition on a silicon substrate was assessed (Fig. 2). Immediately after
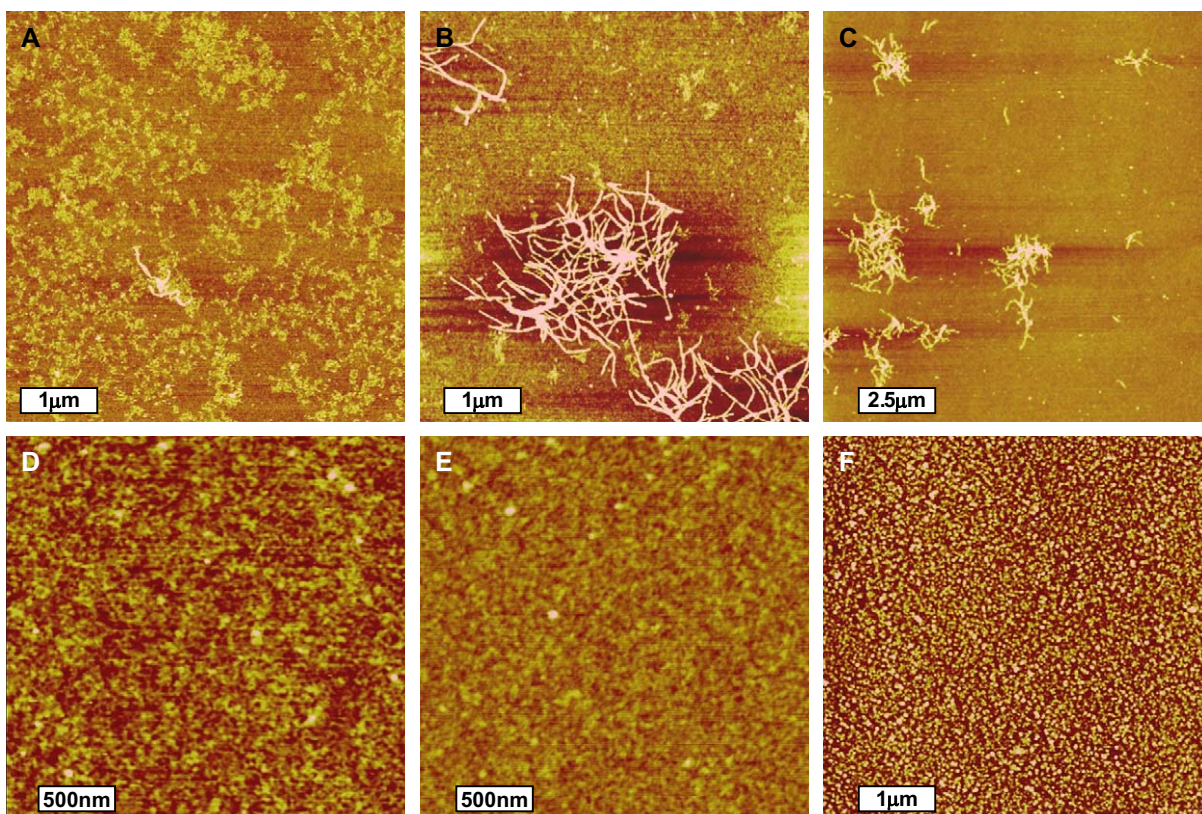

Fig. 2. Topographical AFM images of time-dependent changes in structure of adsorbed silk fibroin proteins: (A–C) native protein and (D–F) dialyzed protein after 1, 11, and 21 days (from left to right).

formation of the protein surface layer, the majority of the protein formed globular aggregates irrespective of initial solution concentration for native protein or after treatment of the protein with the reductant, DTT, to disrupt the disulfide bond. The gradual temporal evolution of initial surface morphology was observed over several days and highly branched well-defined clusters formed through aggregation of initial fibrillar structures (Fig. 2). This process may be due to slow (days–weeks) de-wetting and aggregation of predominantly hydrophobic backbone blocks from the highly hydrophilic silicon oxide surface assisted by the presence of an ultrathin water layer under normal humid conditions. Remarkably, after dialysis (to remove the light chain (~25 kDa) fibroin that had been bound via the disulfide bone to the heavy chain), the silk protein loses its ability to form these fibrillar structures and the initial globular surface morphology persists (Fig. 2).

3.2. Stretching strategies for fibroin molecules

During the stretching experiments, the protein molecules are occasionally picked up by the AFM tip and stretched to a certain limit. The stretching is controlled with a precision of around 0.1 nm that can be conducted with different levels of stretching (large and small strains) and rates while monitoring normal forces [28]. The resulting force–displacement curves with characteristic saw-tooth shapes and consistent spacing were selected and used for further analysis [21,22, 29,30]. The usual criteria of periodicity, force sequence, shape of a single peak, and appropriate spatial distribution are applied for this selection as will be discussed below. Although the unfolding stretching behavior for different strains was observed for some silk proteins tested in the solid state, the internal domain structure of *B. mori* fibroin silk protein has not been revealed [31]. For example, good force statistics were observed for spider dragline silk protein unfolding with a primary unfolding length of 14 nm assigned to specific A_nG_m-type flexible moduli [21]. However, others observed only long-range stretching events covering the range between 800 nm and several micrometers that were assigned to either stretching of a whole molecule or the complex stretching of protein aggregates [28,31]. In all these cases, the interpretation of the SFS data was compromised because of a lack of independent information on exact and complete primary amino acid sequence for the protein backbones.

Considering the complex multi-domain *B. mori* silk protein composed of 12 hydrophilic and 12 hydrophobic domains which comprise a range of possible molecular lengths (Table 1), three independent types of SFS measurements were exploited to include different strategies of stretching with appropriate forces and rates. First, we focused on stretching with relatively high forces and extensions reaching 3 μm which exceeds the completely extended length of the molecules (1.8 μm) (Table 1, Fig. 3). This probing was conducted for the surface areas with "bulk" protein (occasional thicker surface patches). Second, we limited the stretching distance below 1.5 μm with high-resolution in the *Z*-direction to collect
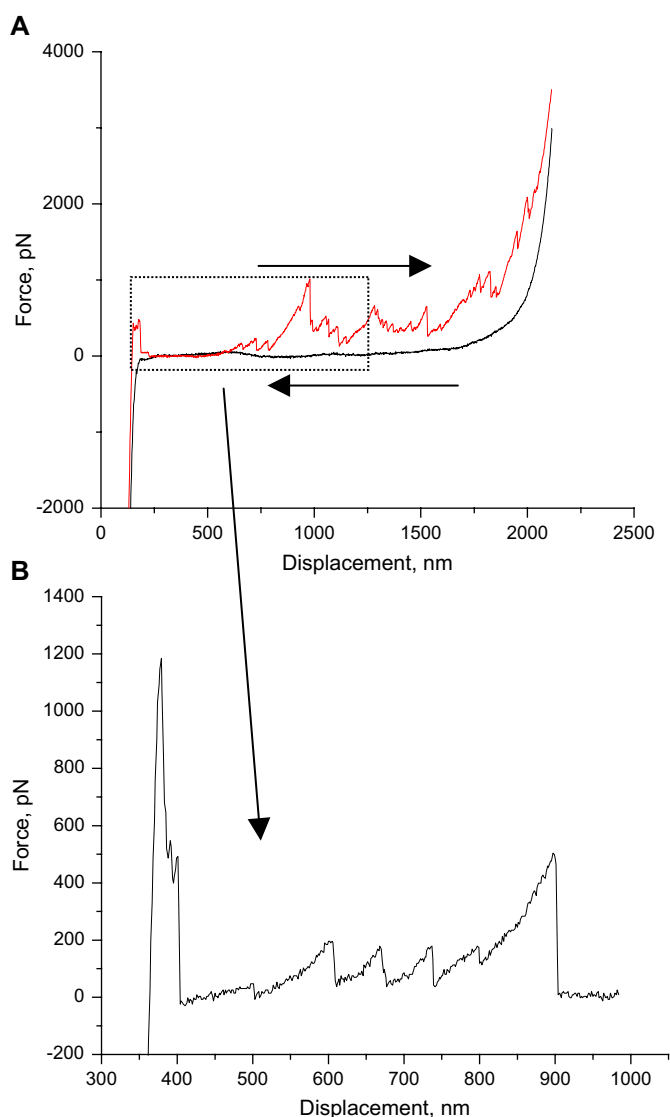

Fig. 3. Force–displacement curves. (A) Largest extension (>2.5 μm, stretching rate is 3000 nm/s) both stretching and relaxation are shown, (B) extension curve for mid-extension (<1 μm, stretching rate is 1000 nm/s).

data on sequential unfolding events for a "normal" surface layer of protein (Fig. 3). In this case, we focused on detailed saw-tooth shape analysis for the extension range within 1000 nm. Finally, we conducted gentle, low-rate, limited stretching with low stretching distances (below 300 nm) to focus on sequences of weak unfolding events related to the hydrophilic domains (Fig. 4). It is worth to note that due to instrumentation limitations (vertical resolution, drifts, temporal limits) it is impossible to obtain high-resolution data which reflect all events in a single experiment. Therefore, split in three different types of stretching ranges was critical for analyzing and collecting data reflecting different stages of stretching. On the other hand, even if we limited our experiments to very small extension, the mode of collection in force–volume regime with lateral displacement for more than 1 μm in most cases provided "fresh" conditions for each tip engagement and eliminated to some extent multi-chain events.
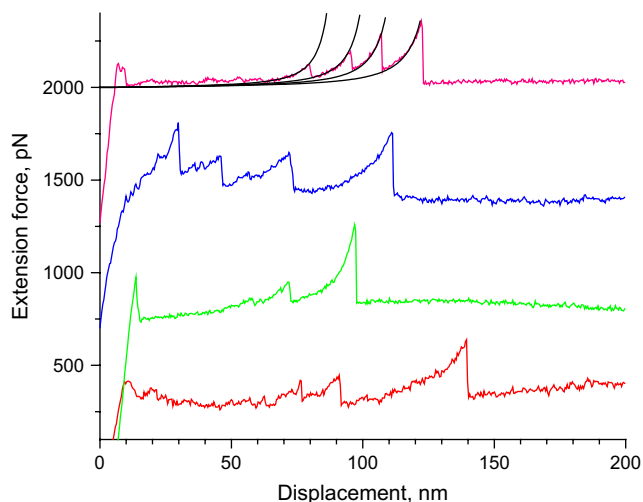
Fig. 4. Examples of several force–displacement curves on a small extension scale (below 300 nm) and an example of corresponding WLC fitting curves used to determine extension domain lengths and Kuhn segment.

Fig. 5. Ultimate stretching force peaks with different lengths at a small extension scale.

The largest and completely reversible stretching consistently collected at thicker protein layers resulted in irregular saw-tooth peaks with variable heights and spacings indicating stretching of multiple and irregularly attached molecular bundles. For the largest extensions a sharp increase in overall force reached 3–5 nN which is close to that sufficient to break protein backbones with covalent bondings. This steep increase in forces acting on the retracing tip occurred at the maximum stretching lengths reaching 2 μm (Fig. 3A).

In contrast, probing thin surface protein layers containing globular aggregates while keeping a limited stretching distances generated data with a regular sequence of saw-tooth peaks, characteristic of single molecule stretching (Fig. 3B). Clear saw-tooth shapes were observed with unfolding forces of 200–400 pN, and peak spacings distributed within 20–100 nm. Relaxation curves showed complete reversibility with forces gradually dropping to the initial level. Typical maximum extensions were ranged from 0.7 to 1.2 μm.

Finally, very gentle stretching to distances below 300 nm with low rates at the surface areas with globular aggregates produced sequences of regular saw-tooth peaks with forces of 50–200 pN and spacings within 10–40 nm (Fig. 4). The shape of curves (about 60 selected from more than 2000 attempts) was reproducible and described nicely by WLC model (see several examples in Fig. 4). The shape and parameters were weakly dependent upon the rate of stretching for the range used here. Finally, the overall forces required to complete unfolding event in this case increased with the extension (although remaining very scattered) (Fig. 5).

3.3. Assignment of different stretching events

The assignment of different stretching events observed here was conducted by using several conventional criteria widely discussed in the literature: correspondence of the total
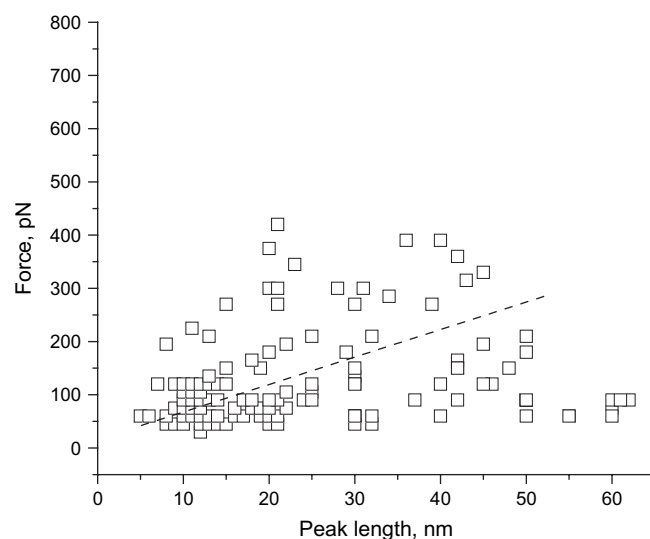
stretching distance to the length of backbones calculated for different conformation; correspondence of the length of a single peak to the extended length of different segments; correspondence of the forces associated with a particular stretching event to the expected forces for unfolding, elastic deformation, or chain scission; consistency in the stretching events (random or concentrated around some mean values); quality of fit for known stretching mechanisms of flexible backbones (e.g., WLC); correspondence of the values of Kuhn segments evaluated from experimental data with expected values for given chains; and consistency of the Kuhn segment values for presumably similar stretching events (to exclude multi-chain contribution) [21,22,32]. According to these approaches we generated statistics of lengths and forces for different events, conducted WLC analysis, and compared these experimental data with geometrical dimension calculated from chemical composition of the molecular chains to conduct final assignment (Figs. 6–8, Table 1).

From this analysis, first we considered large-scale stretching which resulted in a sharp force increase for the length close to 2 μm (Fig. 3A). The total length of this stretching event was comparable with the total, fully extended length of the backbones assuming a completely unfolded structure and an extended conformation of all hydrophilic and hydrophobic domains (1.84 μm, Table 1). This behavior can be assigned to the case of complete stretching of protein backbones with random unfolding events indicating multi-chain attachment and concurrent stretching events under these conditions (high rate and high forces) (Fig. 3A). These events are relatively rare because to complete they require forces of 2–4 nN which are close to the backbone rupture strength [21], thus they can be observed only with very strong tethering of the terminal chains to the AFM tip. The overall shape and force parameters for this type of stretching resemble those observed by Hansma et al. [28] and Gaubb et al. [31].
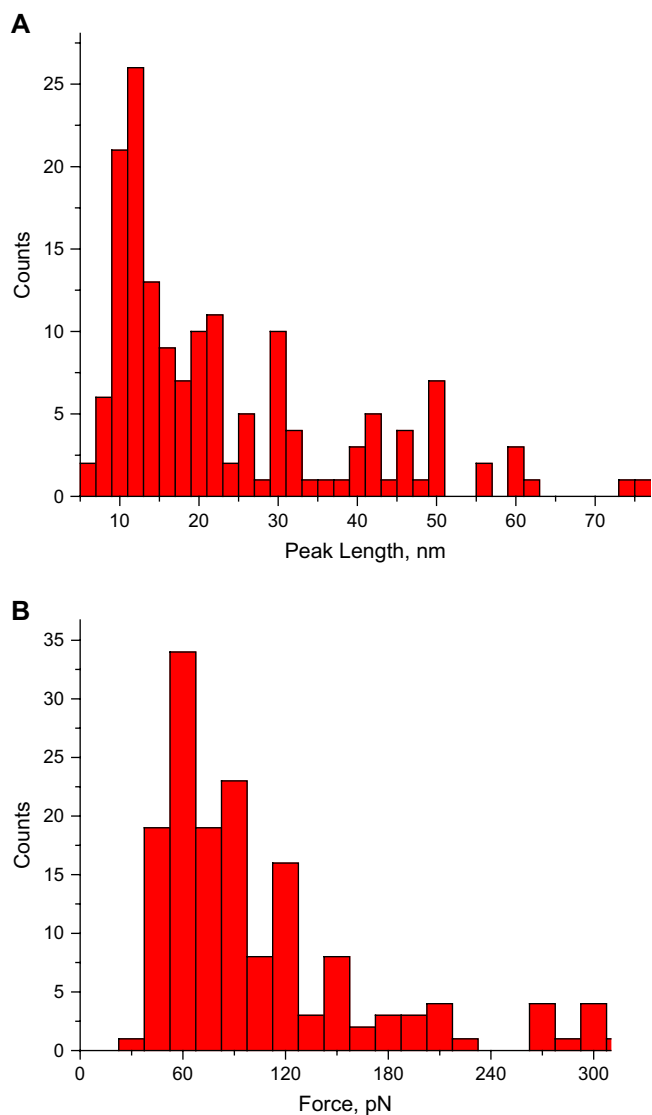
A

B

Fig. 6. Statistics (unfolding peak length (A) and unfolding force (B)) for small extensions.

In contrast, at intermediate and small stretching distances more organized events were observed which can be associated with the regular unfolding in the protein backbone. For low stretching, the well-defined spacing of 11.3 nm dominates the unfolding pattern (Fig. 6A). Overall, 11.3, 22.5, and 32 nm spacings with a ratio close to 1:2:3 were consistently detected (Fig. 6A). The main periodicity coincided precisely with the fully extended length of the hydrophilic domains calculated from the amino acid composition (11.2 nm, Table 1). Forces for a single unfolding event were close to 60 pN, which is common for this type of backbone segment unfolding (Fig. 5) [28]. The analysis of individual domain unfolding in terms of the Hook's spring model suggested a spring constant of about 0.01 N/m, which is expected for stretching of flexible *backbones* governed by an entropic mechanism (Fig. 4) [33]. The persistent length determined from the SFS fitting with the worm-like chain (WLC) model (see examples in Fig. 4) was 0.24 nm, which is characteristic of

very flexible chains [21,32]. Other consistent lengths can be assigned to double and triple unfolding events. Finally, the extended length of the series of events (estimated as the total length of all peaks) spread from 100 to 230 nm and are centered at 160 nm (Fig. 7). These lengths correspond to the length of backbones with folded α-helical hydrophobic and stretched hydrophilic domains (including two terminal hydrophilic domains). The terminal domains can be stretched to the contour length of 53 and 18 nm, respectively (Table 1).

Therefore, based on the analysis of length, forces, statistics, the total length, and Kuhn segments presented above we can assign these events to the unfolding of hydrophilic spacers within the protein backbone. This conclusion is also supported by the literature data as reported by Oroudjev et al. [27] for spider dragline silk protein with A_nG_m segments with long central polyA sequences.

Very different statistics observed for the intermediate stretching regime hinted a different cause for the unfolding events observed in this range (Fig. 7). It is worth noting that small-extension events are "hiding" in the initial portion of the curves and are not well-resolved when we optimize mid-range stretching conditions. The absence of a single well-defined spacing found in this case correlates well with the statistics of the expected lengths for hydrophobic domains within the backbone derived from the amino acid composition (Fig. 7, Table 1). This statistic shows a wide distribution without a well-defined single preferential length which favors irregular distribution of unfolding lengths observed in the SFS experiment (Fig. 7A,B). Forces required for these unfoldings are significantly higher in this case and peak at 200 pN (Fig. 7C). This threefold higher value indicates unfolding less compliant segments of the backbone as expected for hydrophobic domains in aqueous environment where in addition to usual entropic resistance, a significant enthalpic contribution caused by unfavorable intermolecular interactions should act against the external forces [34]. The shape of curves is still described by WLC model which generates a twofold higher persistent length (within 0.4–0.6 nm) also consistent with less flexible backbones within hydrophobic domains. Finally, the overall stretching lengths clustered within the 700–1000 nm correlated well with the end-to-end distances of the protein backbones with fully unfolded and oriented hydrophobic domains (with preserved α-helices) combined with extended hydrophilic domains (Table 1).

Thus, considering all these results, we suggest that these events can be assigned to the unfolding of hydrophobic domains. Here we should assume of α-helical structure of the backbones for limited sequences to allow for reasonable correspondence of lengths and forces. It is worth to note that for these series of events the overall statistics of length and forces does not show clear peak values which make the latest assignment somewhat vulnerable and alternative considerations such as the loophole formation and multi-chain stretching can be considered as alternative mechanisms [32,34]. Thus, additional studies with possible use of specially designed proteins and their segments are required to fully justify this assignment.
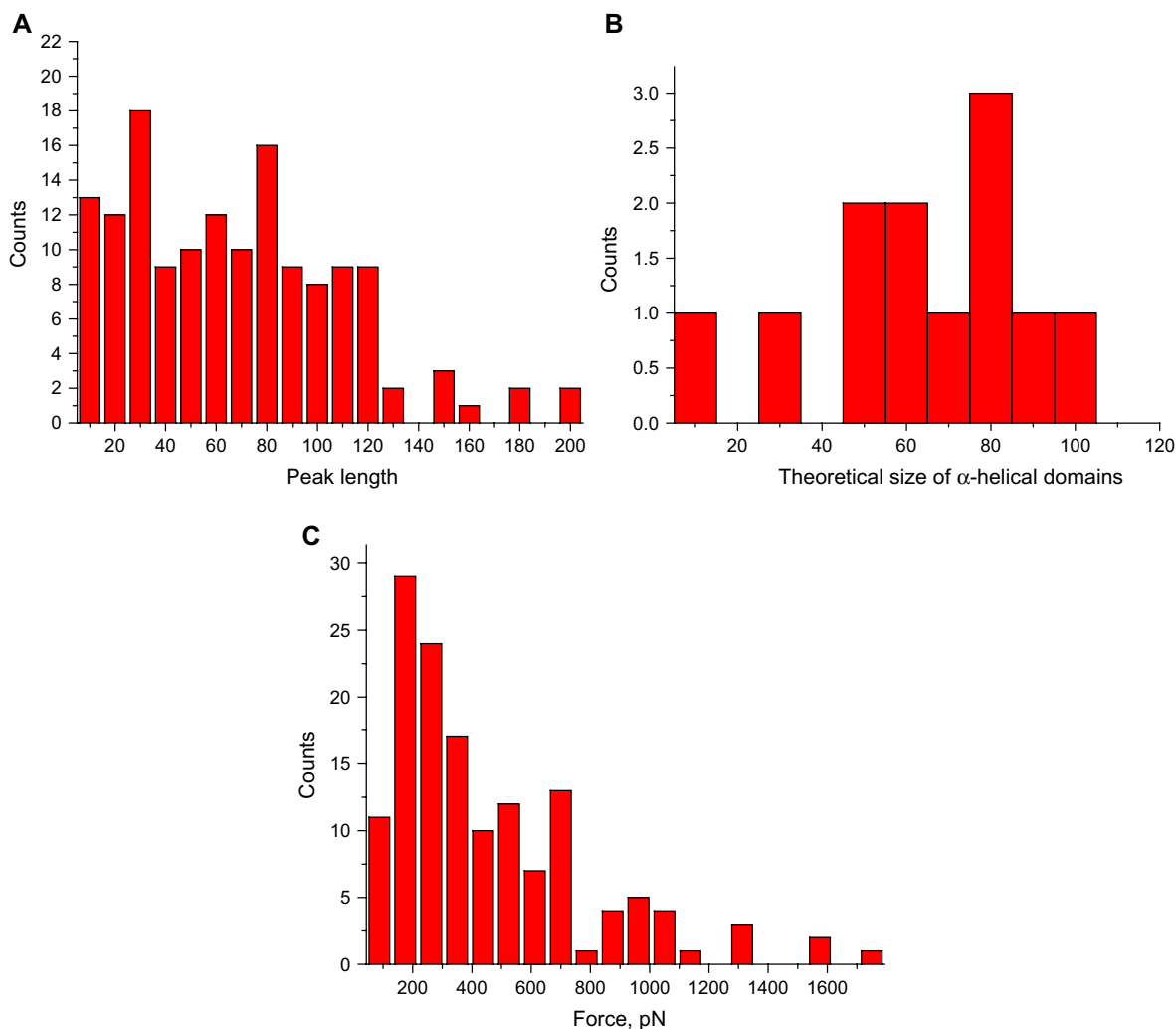
Fig. 7. Statistics (unfolding peak length (A) and unfolding force (C)) for mid-extensions. Statistics for hydrophobic domain lengths derived from Table 1 (B).

4. Conclusions

Overall surface morphology with well-developed fibrillar aggregates and highly branched clusters formed through aggregation of initial fibrillar structures was consistent with other reports with similar silk fibroin proteins where individual molecules, individual nanofibrils, and branched fibrillar aggregates were observed on different surfaces [24,27,35]. The critical role of terminal light fibroin domain attached to the main backbone via disulfide bond was demonstrated in this study. In fact, the silk protein molecules without this domain lose their ability to form fibrillar structures. This morphology is considered to be a critical feature responsible for the interesting mechanical properties of these proteins when formed into fibers by these organisms.

The overall interpretation of possible stretching scenarios has been made on the basis of analysis of individual statistics for spacings and forces derived from the model of block sizes and distributions in this protein. This unfolding pattern suggested reflects a complex character of fibroin stretching which includes weak—intermediate—strong events of unfolding for the combination of 12 hydrophilic and 12 hydrophobic
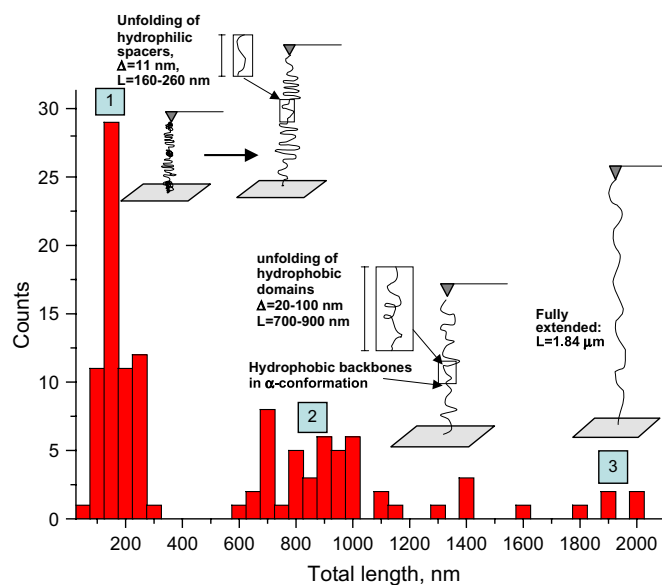

Fig. 8. Statistics for maximum stretched length (3) and the total length of the events (1 and 2) obtained in different extension scenarios and theoretical estimations (boxes) for total extended lengths as estimated from Table 1 for different scenarios (cartoons).

domains and the ultimate stretching of the backbones (Fig. 8). Although different aspects of silk protein unfolding have been previously reported, the established chemical composition of the fibroin backbone combined with the stretching strategy allowed us to suggest the complex character of unfolding of silk protein molecules with the assignments of different events widely supported by the geometrical parameters derived from the backbone composition. The unfolding pattern described here are in good correlation with multi-domain hydrophilic–hydrophobic model suggested before on the basis of amino acid composition of the protein backbone with all major parameters (lengths, spacings, persistence lengths, and forces) consistent with model estimations. This strategy should be transferable to synthetic block copolymer designs. When considered within the context of combinatorial chemistry, important options to exploit sequence variants and relationships to functional features of polymers should emerge.

Acknowledgements

References

[1] Altman GH, Diaz F, Jakuba C, Calabro T, Horan RL, Chen J, et al. Biomaterials 2003;24:401–16.

[2] Sofia S, McCarthy MB, Gronowicz G, Kaplan DL. J Biomed Mater Res 2001;54:139–48.

[3] Cunniff PM, Fossey SA, Auerbach MA, Song JW, Kaplan DL, Adams WW, et al. Polym Adv Technol 1994;5:401–3.

[4] Gosline JM, Demont ME, Denny MW. Endeavour 1986;10(1):37–43.

[5] Mahoney DV, Vezie DL, Eby RK, Adams WW, Kaplan D. Silk Polym 1994;544:196–210.

[6] Jin HJ, Kaplan DL. Nature 2003;424:1057–61.

[7] Zou S, Korczagin I, Hempenius MA, Schöherr H, Vancso GJ. Polymers 2006;47:2483–92.

[8] Shi W, Zhang Y, Liu C, Wang Z, Zhang X, Zhang Y, et al. Polymer 2006; 47:2499–504.

[9] Kühner F, Gaub EH. Polymer 2006;47:2555–63.

[10] Zhang Q, Marszalek PE. Polymer 2006;47:2526–32.

[11] Tsukada M, Gotoh Y, Nagura M, Minoura N, Kasai N, Freddi G. J Polym Sci Part B Polym Phys 1994;32:961–8.

[12] Zhou CZ, Confalonieri F, Medina N, Zivanovic Y, Esnault C, Yang T, et al. Nucleic Acids Res 2000;28:2413–9.

[13] Bini E, Knight DP, Kaplan DL. J Mol Biol 2004;335:27–40.

[14] Kim UJ, Park J, Li C, Jin HJ, Valluzzi R, Kaplan DL. Biomacromolecules 2004;5:786–92.

[15] Tsukruk VV, Reneker DH. Polymer 1995;36:1791–808.

[16] Tsukruk V. Rubber Chem Technol 1997;70:430–67.

[17] Tsukruk VV, Gorbunov VV. Probe Microsc 2001;2:241–7.

[18] Sperling L. Polymeric multicomponent materials: an introduction. New York: Wiley Interscience; 1997.

[19] Butt HJ, Jaschke M. Nanotechnology 1995;6:1–7.

[20] Hazel JL, Tsukruk VV. Thin Solid Films 1999;339:249–57.

[21] Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM. Prog Biophys Mol Biol 2000;74:63–91.

[22] Zhang W, Zhang X. Prog Polym Sci 2003;28:1271–95.

[23] Luzinov I, Julthongpiput D, Tsukruk VV. Macromolecules 2000;33: 7629–38.

[24] Inoue SI, Magoshi J, Tanaka T, Magoshi Y, Becker M. J Polym Sci Part B Polym Phys 2000;38:1436–9.

[25] Hossain KS, Nemoto N, Magoshi J. Langmuir 1999;15:4114–9.

[26] Zhang LF, Yu K, Eisenberg A. Science 1996;272:1777–9.

[27] Oroudjev E, Soares J, Arcdiacono S, Thompson JB, Fossey SA, Hansma HG. Proc Natl Acad Sci U S A 2002;99:6460–5.

[28] Becker N, Oroudjev E, Mutz S, Cleveland JP, Hansma PK, Hayashi CY, et al. Nat Mater 2003;2:278–83.

[29] Meadows PY, Bemis JE, Walker GC. Langmuir 2003;19:9566–72.

[30] Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM. Proc Natl Acad Sci U S A 1999;96:11288–92.

[31] Zhang WK, Xu QB, Zou S, Li HB, Xu WQ, Zhang X, et al. Langmuir 2000;16:4305–8.

[32] Bemis JE, Akhremitchev BB, Walker GC. Langmuir 1999;15:2799–805.

[33] Ikai A, Mitsui K, Tokuoka H, Xu XM. Mater Sci Eng C Biomimetic Mater Sens Syst 1997;4:233–40.

[34] Cui S, Liu C, Zhang W, Wu C. Macromolecules 2003;36:3779–82.

[35] Inoue S, Tsuda H, Tanaka T, Kobayashi M, Magoshi Y, Magoshi J. Nano Lett 2003;3:1329–32.