MRS Communications (2013), *3*, 185–190
© Materials Research Society, 2013
doi:10.1557/mrc.2013.33

MRS Communications

Research Letters

Viscoelastic relaxation time and structural evolution during length contraction of spider silk protein nanostructures

Graham Bratzel, Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue Room 1-235A&B, Cambridge, Massachusetts; Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts
Zhao Qin and **Markus J. Buehler,** Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue Room 1-235A&B, Cambridge, Massachusetts

Address all correspondence to Markus J. Buehler at **mbuehler@mit.edu**

(Received 12 June 2013; accepted 21 August 2013)

Abstract

Spider dragline silk is a self-assembling protein that rivals many engineering fibers in strength, extensibility, and toughness, making it a versatile biocompatible material. Here, atomistic-level structures of wildtype MaSp1 protein from the *Nephila clavipes* spider dragline silk sequences, obtained using an in silico approach based on replica exchange molecular dynamics and explicit water, are subjected to nanomechanical testing and released preceding failure. We approximate the relaxation time from an exponential decay function, and identify permanent changes in secondary structure. Our work provides fundamental insights into the time-dependent properties of silk and possibly other protein materials.

Spider silk achieves remarkable properties due to specific secondary and tertiary structures of its repeating amino acid units, which self-assemble into a hierarchical fibrillar structure.[1–5] The webs of higher spiders, including the Golden Orb-Weaver *Nephila clavipes*, are composed of different kinds of silk, each with distinct mechanical properties that are adapted for the purpose of the respective part of the web.[6] Major ampullate silk, the strongest of *N. clavipes* silk, is used for the dragline as well as the spokes and outer frame of the web.[7,8] Two distinct proteins are typically found in dragline silks: major ampullate spidroin proteins (MaSp) MaSp1 and MaSp2, which have different repeating units and distinct mechanical functions.[6,9,10] MaSp1 contains poly-alanine $(A)_n$ and glycine-rich $(GGX)_n$ repeats, where X typically stands for alanine (A), tyrosine (Y), leucine (L), or glutamine (Q).

Recent investigations have revealed that variations in crystallinity and alignment within the silk thread, for example, due to the reeling speed of the collected sample, have been mapped to macroscopic mechanical properties by affecting the formation of the β-sheet crystals.[11–17] These cross-linking β-sheet crystals employ a dense network of aligned hydrogen bonds and have dimensions of a few nanometers.[11–13] A previous study of dragline silk employing replica exchange molecular dynamics (REMD) has yielded the first atomistic results in comparison with experimental structure identification methods.[11] Building on this work, the nanomechanical behavior of these silk structures has recently been reported using classical molecular dynamics (MD).[18] The work presented in this paper is aimed to reveal the elastic return and pseudo-plasticity effects of the silk structure after large deformation.

Indeed, while experimental studies have shown the elastic return of the silk thread, and molecular-level analyses of the nanostructure of silk assemblies have been stretched to failure, the elastic returns of these nanostructures have not been investigated using molecular simulation. The availability of powerful new methods to computationally design and experimentally synthesize varied protein materials from the bottom up and with full control over the amino acid sequence will open exciting opportunities to engineer materials beyond spider silk for specific mechanical properties and other functional purposes.[19,20] The following sections will describe the simulation and analysis methods, report and discuss the results of the computational experiments, and offer a conclusion as to the insights of the work.

Materials and methods

Spider silk filaments are composed of cross-linked extended proteins with distinct typical repeating sequence units: the poly-Ala $(A)_{n1}$ repeat unit and the Gly-rich $(G_mX)_{n2}$ repeat unit, where X is a peptide composed of one to three amino acids aside from glycine, m is the number of repeats of glycine in a repeat unit with a value between one and two, and $n1$ and $n2$ denote the numbers of repeat units with the values of an arbitrary integer. Experimental studies of recombinant silk suggest that mechanical shear within a narrowing elongational flow extends and aligns MaSp oligomers during spinning into a

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press

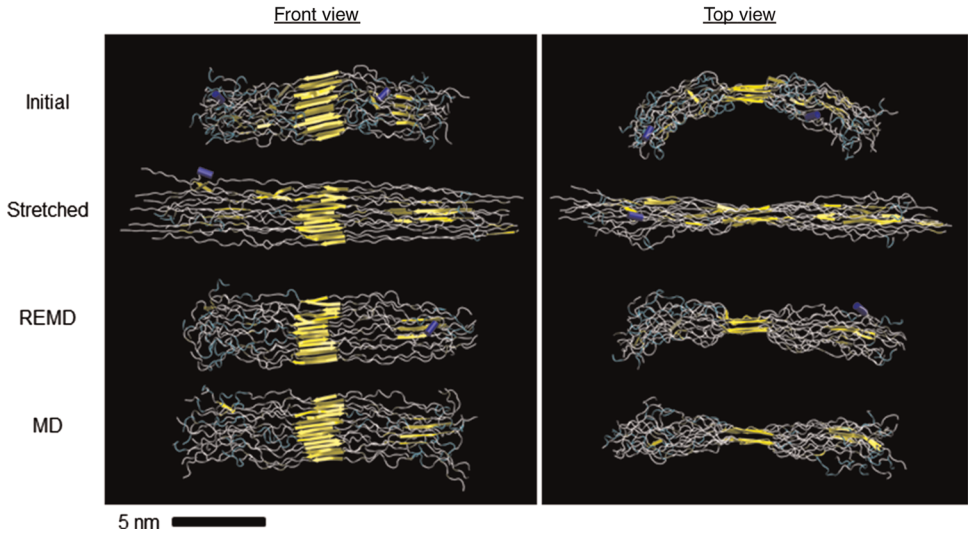


Figure 1. Snapshots of the MD trajectory before stretching, deformed by application of force to just before failure by multi-lap shearing, and after relaxation using REMD and classical MD. While the initial curvature is lost, REMD and classical MD predict very similar structures after convergence. The central β-sheet crystal remains intact, while the secondary structures of the semi-amorphous regions change before and after stretching.

cross-linked semi-amorphous filament network. It is noted that the cross-links for stabilizing the network are composed of hydrogen bonding, van der Waals interactions, and electrostatic interactions, which belong to physical cross-links that can be reformed relatively easily. By mimicking this process, the test case structure (Fig. 1) used in this study represents a β-sheet crystal and surrounding semi-amorphous network. Previous computational studies have shown that the extended and aligned poly-Ala units in the MaSp1 sequence amyloidize into β-sheet nanocrystals. The extended partial polypeptide chains used in this study are composed of a single poly-Ala repeat with one instance of the Gly-rich semi-amorphous repeat unit on each side. The wildtype *N. clavipes* MaSp1 partial peptide sequence used in this study, in one-letter amino acid code, is: (GAGQGGYGGLGSQGAGRGGLGGQGAG)(AAAAAA)(GGAGQGGYGGLGSQGAGRGGLGGQGAG), with repeated units marked by parentheses. Fifteen such chains are arranged in parallel and anti-parallel in the test structure, totaling 870 residues and forming both parallel and anti-parallel β-sheets in the crystal region.

The test structure for equilibration is exported from the MD trajectory of a previous study involving nanomechanical testing,[18] in which the protein was stretched to failure by loading alternate ends. In that study, the stretching force is applied to each loaded alpha-carbon terminus and increased by a step of 20 pN for every 20 ps of equilibration. The starting conformation in the current study is taken from the trajectory right before the rupture under stretching.[18] Equilibration is performed using classical MD at 298 K as well as REMD with $N = 96$ replicas evenly spaced between $T_{min} = 298$ K and $T_{max} = 480$ K. Each case is simulated on the NICS Kraken Cray XT5 system using the NAMD (Not Another Molecular Dynamics) molecular modeling program and CHARMM-22 force field with explicit water solvent at 1 atm pressure in periodic conditions. The temperature of each replica i is assigned as $T_i = T_{min}(T_{max}/T_{min})^{(i-1)/(N-1)}$. The timestep is set to 2 fs with the assumption of rigid bonds that involve hydrogen atoms during MD run between two exchange events. The total energy is computed every 1000 steps and the exchange event is decided thereafter. This algorithm is designed in order to enable us to reach the sufficiently large acceptance ratios (10–20%) that is necessary for computational efficiency as shown in Fig. 2(a). Particle Mesh Ewald (PME) electrostatics is employed to better capture solvent effects and solvent-mediated changes in secondary structure.

We scale the simulation time during the REMD run according to the protein length as a function of simulation time in the classical MD simulation. This is performed because the REMD combines Monte-Carlo exchanges during the dynamic runs, and thus its definition of simulation time is lost. The reason to conduct both classical MD simulation and REMD is to understand the recovery time for silk as well as ensure that the protein structure has been fully equilibrated by comparing the results with the two different methods. REMD is able to increase the sampling space and lead to the configuration of minimum free energy, which is beyond the capability of other classical MD simulations such as a simple elevated temperature. To scale the simulation time of the REMD, we first fit the length of the protein in the classical MD simulation as a function of simulation time t by using an exponential decay function as $\lambda_{MD}(t) = (\lambda_0 - 1)\exp(-t/t_0) + \lambda_\infty$, where $\lambda = l/l_0$ is the stretch of the protein structure with an equilibrated length of l_0 and stretched length l, λ_0 is the stretch before relaxation and t_0 is the time constant for relaxation. We perform a similar

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press

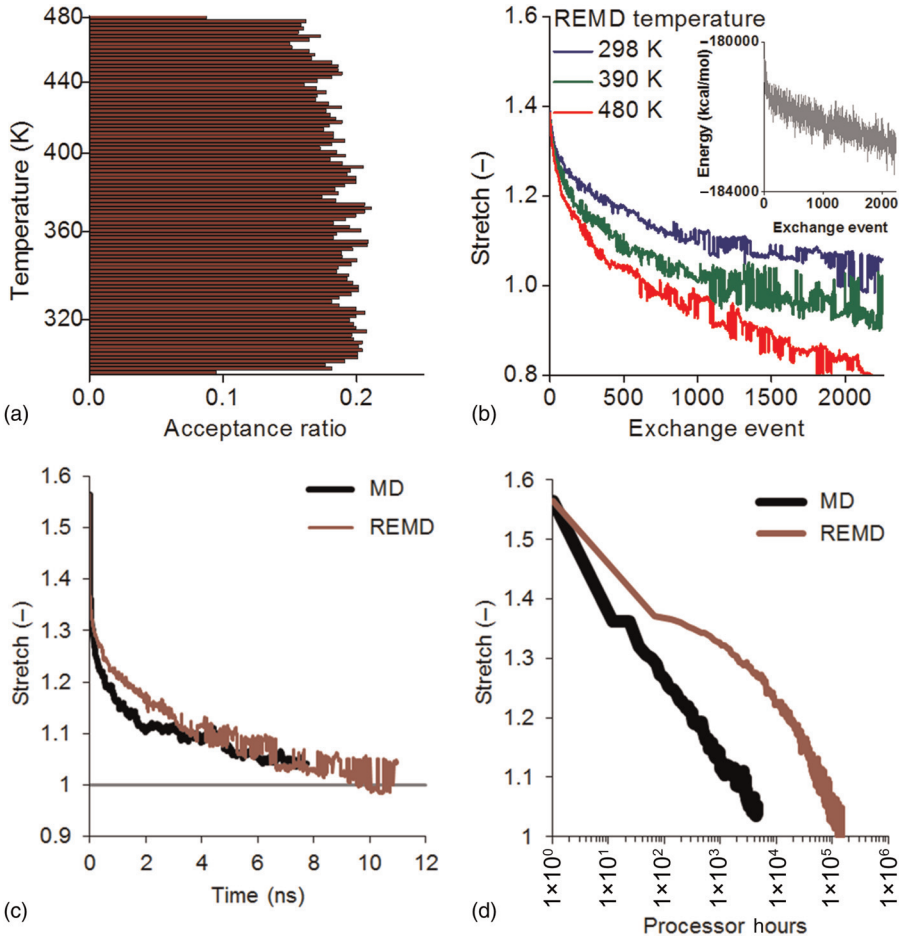


Figure 2. Length contraction of the protein in classical MD and REMD simulations at 298 K after being stretched almost to predicted failure. (a) The acceptance ratio of each replica during REMD run as a function of its temperature. (b) The contraction of the total length of the protein for replicas with different temperatures as a function of the number of exchange events in the REMD simulation. The inserted plot gives the total energy evolution of protein conformation in the replica of the lowest temperature (of 298 K). (c) Total length contraction occurs within 10 ns from a stretch of 1.56. Time for the REMD simulation is scaled so as to have the same fitted exponential decay time constant as the classical MD simulation. (d) Stretch versus computational expense. REMD converges to a structure with the initial length by overcoming local conformational energy wells, but simulating multiple replicas simultaneously is much more computationally expensive than typical classical MD.

analysis for the 298 K trajectory of the REMD except that the protein stretch is a function of the number (N) of the configuration after every exchange event as $\lambda_{REMD}(N) = (\lambda_0 - 1)\exp(-N/N_0) + \lambda_\infty$, where N_0 is the constant of the number of the configuration for relaxation in REMD [Fig. 2(b)]. The scaled time of REMD is thereby given as $t_{REMD} = Nt_0/N_0$, yielding the same time constant of exponential decay as the classical MD simulation [Fig. 2(c)] by direct comparison. It is also noted that the contour length of the protein is calculated by averaging the distance from the central crystal to each of the 15 ends for each side and summing these two half-lengths. This is used to take into account the slight total curvature of the initial structure.

The instantaneous secondary structure content of each equilibrated structure is determined using the STRIDE algorithm built into the Visual Molecular Dynamics (VMD) Molecular Graphics Viewer[21] and customized Tool Command Language (TCL) scripts. The STRIDE algorithm holds an advantage over Define Secondary Structure of Protein (DSSP) and other secondary structural algorithms by employing pattern recognition of statistically-derived backbone dihedral angle information. For the purpose of counting the number of the hydrogen bonds, they are defined by using the cut-off length of 3.5 Å and 35° between the donor and acceptor.

Results and discussion

The MaSp1 protein structure considered here returns from a stretch of 1.56 to its initial contour length within 10 ns [Fig. 2(c)]. Although fitting the length over time with an exponential decay function is only an approximation, most of the length contraction (from 1.56 to 1.2 stretch) occurs within the first 2 ns. The time constant of the exponential decay of the classical MD simulation is found to be 4 ns, and the timescale of the REMD simulation is scaled to have this same

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press

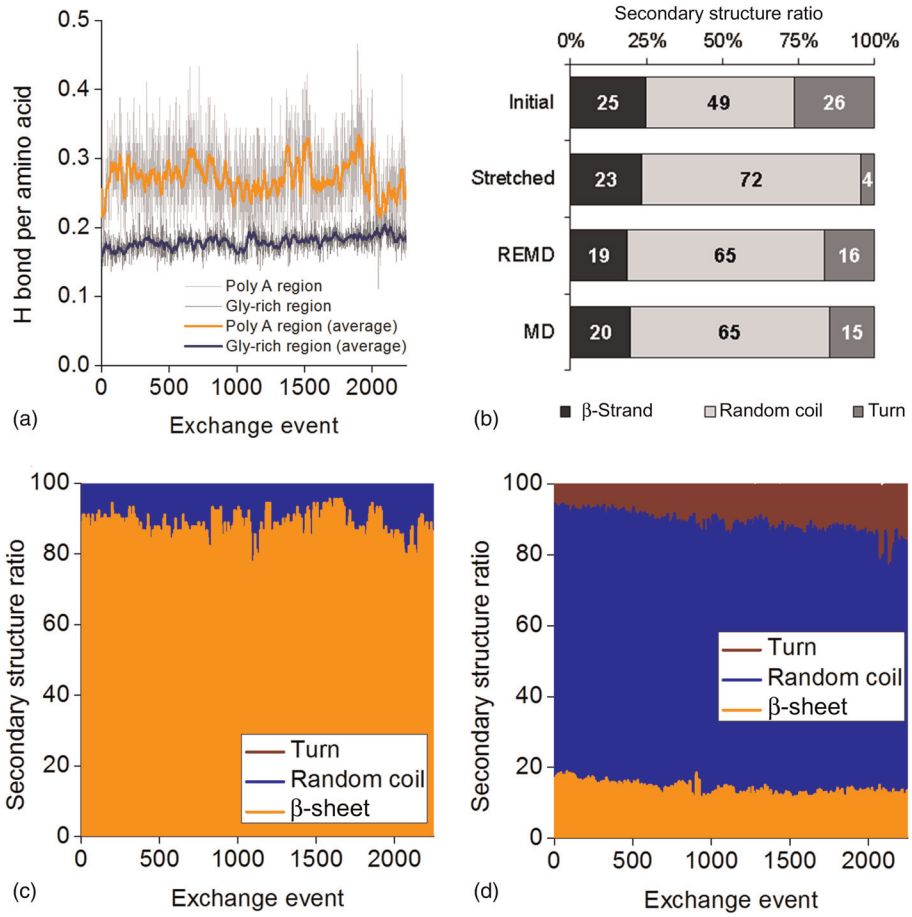


Figure 3. Number of hydrogen bonds and secondary structure evolution before and after stretching as well as after relaxation in REMD and classical MD simulations. (a) The number of hydrogen bonds normalized by the number of amino acids in each of the poly-A and Gly-rich regions as a function of the exchange event during the REMD simulation. (b) Secondary structure evolution before and after stretching as well as after relaxation in REMD and MD simulations as determined by the STRIDE algorithm. During relaxation, both REMD and classical MD simulations predict some return from random coil to turns, but with no new formation of β-sheets. Alpha-helix content is negligible throughout and is hence not shown. (c) Secondary structure evolution during relaxation in REMD for the crystal region (poly-A). (d) Secondary structure evolution during relaxation in REMD for the semi-amorphous region (Gly-rich). During relaxation β-sheets and random coils disrupt and form back to turn structures in the Gly-rich motifs.

time constant of decay. Although the classical MD simulation only reaches a stretch of 1.04, the REMD acts as a forecast and converges to the initial length, i.e., stretch = 1. The final stretch for the classical MD simulation is different from that of the REMD result because there are regions of random coils which can adopt many possible conformations during relaxation, which may prevent the protein from fully recovering to its initial length before stretching in the classical MD simulation. The conformation of the protein structure at the end of the classical MD simulation may correspond to a local energy minimum that is different from its initial state. This limit is overcome by the REMD because of the increased sampling space. It is also noted that the protein length sampled at higher temperatures [of 390 and 480 K, as shown in Fig. 2(b)] in REMD more quickly recovers its initial length, showing that the higher temperature replica actually accelerates the relaxation rate. However, the protein structure in those replicas shrinks to even smaller length after reaching stretch = 1 because the equilibrated conformations of the disordered region of the protein are different at different temperatures. That explains why the exchange steps are required apart from the elevated temperature for REMD simulation of the silk relaxation at 298 K.

The additional computational expense of the REMD over classical MD is evident in Fig. 2(d). The classical MD simulation used 96 processors over 48 hours for a total of nearly 5×10^3 CPU-hours to achieve 8 ns of simulation time. The 96 replicas of the REMD simulations used 768 processors over 8 days for a total of 10^5 CPU-hours to achieve the equivalent of 11 ns simulation time.

Since the central β-sheet is already an extended conformation, the majority of the stretching and relaxation deformation occurs within the semi-amorphous glycine-rich regions. By analyzing the number of hydrogen bonds

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press

normalized by the number of amino acids in each of the crystal and semi-amorphous regions [Fig. 3(a)],[22,23] it is shown that in the crystal region the hydrogen-bond number stays at a constant level while in the semi-amorphous region the number keeps gradually increasing during relaxation. The final structures of the classical MD and REMD simulations are visually similar to the initial structure before stretching (Fig. 1), but the STRIDE secondary structure prediction algorithm shows some permanent structural changes [Fig. 3(b)]. During stretching, total β-sheet content decreases from 25 to 23% of the total structure. Turn structures also decrease from 26 to 4%, and alpha-helix content remains negligible. Random coil structures become dominant, increasing from 49 to 72%.

During relaxation in the 298 K trajectory of the REMD simulations, the β-sheet content decreases further to 19%, and the turn structures return from some of the random coils (65 and 16%, respectively). The classical MD simulation shows a similar trend within 1%. After relaxation, twice as much random coil structures remain relative to turns. Before stretching, random coils were found at a ratio of 1.9:1 to turns. After relaxation, random coils remain at a 4.0–4.4:1 ratio to turns, yet the overall contour length returned to the initial length. We carry out this analysis for the crystal region (poly-A) and semiamorphous (Gly-rich) individually during the REMD run as shown in Figs. 3(c) and 3(d), from which it is seen that the central β-sheet crystal remains intact, and the decrease in β-sheet content of the entire protein occurs in the secondary content within the semi-amorphous regions. The decreased β-sheet and random coil contents contribute to the increased content of turn structures in the semi-amorphous structure.

Conclusion

This study reported an estimate of the relaxation time of the semi-amorphous regions surrounding the β-sheet crystal nodes of the fibrillar network of spider silk as determined by a combination of classical MD and REMD simulations. This region was found to be fully elastic and to return to its initial length within 10 ns. Investigation of the secondary structure ratios before and after stretching showed that almost all of the turn structures are extended into random coils during stretching. The random coil to turn structure ratio at the end of equilibration remains twice as high as its initial unstretched value, even though the protein has returned to its original unstretched length. Our study also reveals that classical MD and REMD yield similar results for silk relaxation. This finding provides a basis that one may combine classical MD and REMD in one simulation, in an approach to use classical MD for initial relaxation and REMD for further structural refinement, which renders the simulation approach more efficient overall. Further equilibration may show a change in the secondary structure ratios after elastic return, and additional cyclic loading simulations may reveal permanent changes to the protein structure or energy absorption capacity.

Understanding the behavior of the crystal node at the molecular scale and its impact on the larger fibrillar network is critical for potentially bypassing strength limitations and vastly improving silk for medical and textile purposes as well as synthetic elastomers and polymer or aramid fiber composites with a similar molecular structure and noncovalent bonding. Further analysis is required to determine if the stiffness and strength have drastically changed with this new secondary structure composition or further cyclic loading. It is important to note that the structure modeled here represents a very small portion of the total silk thread. The relaxation time of an entire silk fiber is likely to be affected by additional phenomena at larger length scales than is concerned in this study, such as microfluidic flow of water or large microstructural changes within the fibrillar network.

Acknowledgments

This study is supported by XSEDE award number MSS090007, and support from NSF and DOD-PECASE.

References

1 F. Vollrath and D.P. Knight: Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
2 N. Becker, E. Oroudjev, S. Mutz, J.P. Cleveland, P.K. Hansma, C.Y. Hayashi, D.E. Makarov, and H.G. Hansma: Molecular nanosprings in spider capture-silk threads. *Nat. Mater.* **2**, 278–283 (2003).
3 Z.Z. Shao and F. Vollrath: Materials: surprising strength of silkworm silk. *Nature* **418**, 741–741 (2002).
4 A.H. Simmons, C.A. Michal, and L.W. Jelinski: Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84–87 (1996).
5 Y. Termonia: Molecular modeling of spider silk elasticity. *Macromolecules* **27**, 7378–7381 (1994).
6 C.Y. Hayashi, N.H. Shipley, and R.V. Lewis: Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **24**, 271–275 (1999).
7 J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, and R. Lewis: Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **291**, 2603–2605 (2001).
8 J.M. Gosline, P.A. Guerette, C.S. Ortlepp, and K.N. Savage: The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* **202**, 3295–3303 (1999).
9 A.E. Brooks, H.B. Steinkraus, S.R. Nelson, and R.V. Lewis: An investigation of the divergence of major ampullate silk fibers from Nephila clavipes and Argiope aurantia. *Biomacromolecules* **6**, 3095–3099 (2005).
10 G.P. Holland, M.S. Creager, J.E. Jenkins, R.V. Lewis, and J.L. Yarger: Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **130**, 9871–9877 (2008).
11 G.H. Bratzel and M.J. Buehler: Sequence-structure correlations in silk: Poly-Ala repeat of N. clavipes MaSp1 is naturally optimized at a critical length scale. *J. Mech. Behav. Biomed. Mater* **7**, 30–40 (2011).
12 S. Keten and M.J. Buehler: Nanostructure and molecular mechanics of spider dragline silk protein assemblies. *J. R. Soc. Interface* **7**, 1709–1721 (2010).
13 S. Keten and M.J. Buehler: Atomistic model of the spider silk nanostructure. *Appl. Phy. Lett.* **96**, 153701 (2011).
14 S. Keten, Z. Xu, B. Ihle, and M.J. Buehler: Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. *Nat. Mater.* **9**, 359–367 (2010).
15 T. Lefevre, M.E. Rousseau, and M. Pezolet: Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. *Biophys. J.* **92**, 2885–2895 (2007).
16 B.L. Thiel, K.B. Guess, and C. Viney: Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. *Biopolymers* **41**, 703–719 (1997).

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press

17 J.D. van Beek, S. Hess, F. Vollrath, and B.H. Meier: The molecular structure of spider dragline silk: folding and orientation of the protein backbone. *Proc. Natl. Acad. Sci. USA* **99**, 10266–10271 (2002).
18 G.H. Bratzel and M.J. Buehler: Molecular mechanics of silk nanostructures under varied mechanical loading. *Biopolymers* **97**, 408–417 (2012).
19 M.E. Kinahan, E. Filippidi, S. Ko"ster, X. Hu, H.M. Evans, T. Pfohl, D. L. Kaplan, and J. Wong: Tunable silk: using microfluidics to fabricate silk fibers with controllable properties. *Biomacromolecules* **12**, 1504–1511 (2011).
20 S.T. Krishnaji, G.H. Bratzel, M.E. Kinahan, J.A. Kluge, C. Staii, J.Y. Wong, M.J. Buehler, and D.L. Kaplan: Sequence–structure–property relationships of recombinant spider silk proteins: integration of biopolymer design, processing, and modeling. *Adv. Funct. Mater.* **23**, 241–253 (2013).
21 W. Humphrey, A. Dalke, and K. Schulten: VMD—visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).
22 L. Ruiz, P. VonAchen, T.D. Lazzara, T. Xu, and S. Keten: Persistence length and stochastic fragmentation of supramolecular nanotubes under mechanical force. *Nanotechnology* **24**, 195103 (2013).
23 Z. Qin, A. Gautieri, A.K. Nair, H. Inbar, and M.J. Buehler: Thickness of hydroxyapatite nanocrystal controls mechanical properties of the collagen-hydroxyapatite interface. *Langmuir* **28**, 1982–1992 (2012).

https://doi.org/10.1557/mrc.2013.33 Published online by Cambridge University Press