Review

Journal of PeptideScience

Received: 8 April 2012 Revised: 18 April 2012 Accepted: 19 April 2012 Published online in Wiley Online Library: 8 May 2012

(wileyonlinelibrary.com) DOI 10.1002/psc.2417

A structural view on spider silk proteins and their role in fiber assembly

Franz Hagn*‡

Spider silk is the toughest known biomaterial and even outrivals modern synthetic high-performance materials. The question of understanding fiber formation is how the spider can prevent premature and fatal aggregation processes inside its own body and how the chemical and mechanical stimuli used to induce the fiber formation process translate into structural changes of the silk material, finally leading to controlled and irreversible aggregation. Here, the focus will be on the structure and function of the highly conserved N-domains and C-terminal domains of spider dragline silk which, unlike the very long repetitive sequence elements, adopt a folded conformation in solution and are therefore able to control intermolecular interactions and aggregation between other spider silk molecules. The structures of these domains add valuable details for the construction of a molecular picture of the complicated and highly optimized silk assembly process that might be beneficial for large-scale *in vitro* fiber formation attempts with recombinant silk material. Copyright © 2012 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: biomaterials; NMR spectroscopy; spider silk; structure

Background

Spider silk is one of the toughest (bio-) materials known so far, with a tensile strength comparable with modern high-performance synthetic materials [1]. The ongoing evolutionary optimization of spider silk over the last 450 Myr has created a material that can be stronger than steel and elastic such as rubber [1–3]. Its outstanding mechanical properties, low weight, and biocompatibility render spider silk a very attractive target for material science. However, the *de novo* design of artificial spider silks with comparable mechanical properties as the natural silk has not been successful so far. For achieving this goal, the relationship between structure and function is a crucial prerequisite. We have fairly good information on the amino acid sequence motifs present in spider fibroins, and our understanding of the molecular architecture of spider silks is growing daily. However, the mechanism of silk assembly is still not completely understood. Only such knowledge can facilitate the design of silk molecules and optimize spinning techniques required for the artificial production of this interesting material.

Orb-web weaving spiders such as *Araneus diadematus* (garden spider) and *Latrodectus hesperus* (black widow) utilize a variety of silk types for different purposes [2] as shown for two silk types in Figure 1a. Each silk type is produced in one of the six different silk glands, each of which synthesizes a unique blend of structural polymers and produces a fiber with a defined set of functional properties [4,5]. The most interesting gland from a material research perspective is the major ampullate gland that produces major ampullate spidroin used for the extremely stable dragline silk and for the frames of the spider's web. The minor ampullate spidroin is utilized for the initial nonsticky scaffold of the web spiral, which is later replaced by other silk types. Being extremely elastic, flagelliform silk is used for the catching spiral to absorb the kinetic energy of the captured prey. For a detailed list of all silk types, please see recent review articles [2,3]. This review is instead focused on the structure and function of major ampullate spidroin proteins.

Sequence and Molecular Architecture of Silk Proteins

Silk proteins show amphiphilic properties (Figure 1b) and consist of three major parts. The fiber core of the tough web frame and dragline silk contains mainly two different major ampullate spidroin proteins that are composed predominantly of glycine, alanine, and proline [2]. Major ampullate spidroins are reminiscent of block copolymers containing stretches of polyalanine and either $(GGX)_n$ or GPGXX (X is typically tyrosine, leucine, or glutamine) forming a highly repetitive core adopting an extended polyproline-type II helix conformation in solution, flanked by highly conserved nonrepetitive (NR) amino-terminal [6,7] and carboxy-terminal [8–10] domains (Figure 2). The N-terminus contains a putative signal sequence responsible for protein export, and the C-terminal domain mediates homodimerization via a disulfide bond. Because of their high sequence conservation, both NR domains are expected to be of functional relevance. The C-terminal NR domain seems to be involved in the control of solubility and fiber formation [8,10], whereas the repetitive sequence elements determine the mechanical properties of the fiber. Molecular weights of dragline silk proteins are estimated to range from 250 to 350 kDa. The extremely low

* *Correspondence to: Franz Hagn, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave, Boston, MA 02115, USA. E-mail: franz_hagn@hms.harvard.edu*

‡ *Friedrich-Weygand Award 2011: The young investigator award in peptide and protein chemistry was assigned to Dr. Franz Hagn on occasion of the annual meeting of the Max Bergmann Society in Kloster Irsee, October 9–12, 2011. This review covers his lecture on that occasion.*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Ave, Boston, MA 02115, USA

Abbreviation used: *ADF3, Araneus diadematus fibroin 3; CD, circular dichroism; HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; NR, nonrepetitive.*

Biography

Franz Hagn grew up near Munich and studied Biochemistry at the University of Bayreuth. In his undergraduate work in the group of Jochen Balbach, he used real-time NMR spectroscopy for monitoring protein folding. After a short stay at Aventis Pharma in Frankfurt, he continued as a graduate student in the lab of Horst Kessler at the Technische Universität München, Germany. His work was focused on using NMR spectroscopy for structure determination, probing molecular interactions, and dynamics of various target proteins ranging from the tumor suppressor p53 to molecular chaperones such as Hsp90 and spider silk proteins. Currently, he is a postdoctoral research fellow at Harvard Medical School in Boston, USA and is focused on new biochemical and NMR methods for the production and structure determination of membrane proteins.

content of charged amino acids and the extremely high abundance of glutamine differentiate spider silk proteins further from other extracellular and structural proteins, such as collagen.

Fiber Protein Storage and Assembly

For controlled extracellular fiber formation, a tightly controlled aggregation and assembly process is mandatory. The identified major ampullate proteins from the garden spider are ADF3 and ADF4 (*Araneus diadematus* fibroin), which are stored as highly concentrated solutions [11] in the lumen of the major ampullate glands in the presence of sodium chloride [12], without the onset of undesirable aggregation inside the spider. During passage along the spinning duct, sodium and chloride ions are exchanged for the relatively kosmotropic potassium and phosphate ions [12,13]. This ion exchange triggers β-sheet assembly in engineered and recombinantly produced variants of ADF3 and ADF4 [14–18]. In spiders, after changing the ionic composition, the protein solutions are further concentrated via epithelial cells extracting water [12],

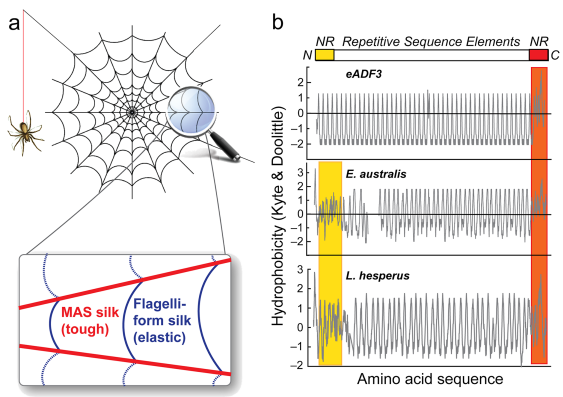

Figure 1. Sequence architecture of spider silk proteins.(a) Cartoon of an orb web. The very tough web frame and the dragline are made of major ampullate spidroin (MAS, red); the elastic catching spiral of flagelliform silk (blue).(b) Schematic domain organization of spider silk proteins and hydropathicity plots calculated according to Kyte and Doolittle [34] for eADF3, an engineered MAS silk from *Araneus diadematus* [16], and the major ampullate silk proteins from *Euprosthenops australis* and *Latrodectus hesperus*.

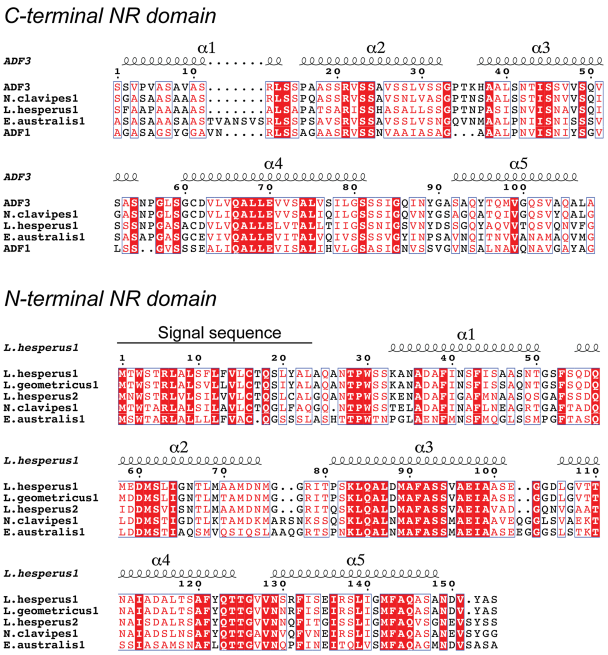

Figure 2. Multiple sequence alignment of the N-terminal and C-terminal nonrepetitive domains of different MAS proteins.Multiple sequence alignments of the C-terminal (upper panel) and N-terminal (lower panel) NR domains of MAS silks of various spider species calculated with ClustralX [35] and visualized with ESPript [36]. The degree of conservation is expressed by the intensity of the red color, and the secondary structure of both domains is depicted above each alignment.

which both support hydrophobic interactions and increase the effect of mechanical stimuli such as shear force and elongational flow [18]), known to align the proteins and promote β-sheet formation [2,11,19–21].

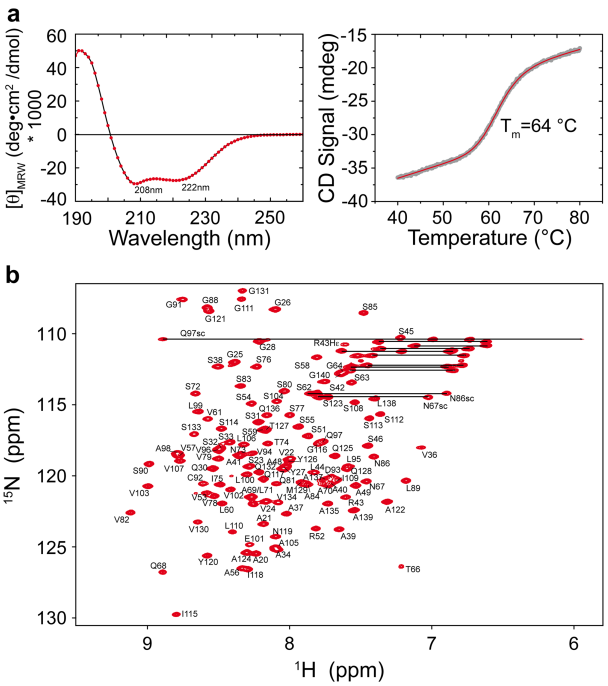

Figure 3. CD and NMR analysis of the C-terminal NR domain of dragline silk (NR3).(a) CD spectrum of the NR domain indicates α-helical secondary structure. The thermal stability of NR3 was measured with CD at 222 nm. (b) 2D [^{1}H,^{15}N]-HSQC of ^{15}N-labeled NR3 with the assignment labeled for each resonance. Side chain amides are marked with lines.

Journal of PeptideScience

NMR Structure Determination and Functional Studies with the C-terminal Domain

Previous studies showed that the nonrepetitive C-terminal NR domain of ADF3 (NR3, nonrepetitive domain of ADF3) contains α-helical secondary structure [16]. This domain is highly conserved among spider silk proteins (Figure 2), suggesting an important role in storage and/or the fiber assembly process. However, the three-dimensional structure of this homodimeric protein was not known until recently, and modeling attempts [22] were unsuccessful because of the general lack of structural information for this protein class. Because of the high aggregation tendency of NR3, which might have hampered crystallization attempts, NMR spectroscopy is a perfect tool for obtaining structural information in solution. In addition, NMR allows monitoring of structural changes induced by alterations in the chemical environment present during storage and silk assembly in the spider's duct, which is essential for understanding the functional role of these domains.

An initial protein construct with an N-terminal T7-tag (T7-NR3) used in previous studies [16] could be produced in milligram quantities. As a first step toward the structural characterization of NR3, the secondary structure and thermal stability were determined using CD spectroscopy (Figure 3a). There, the far-UV CD spectrum clearly shows a very high degree of α-helical content in the protein indicated by the two pronounced minima at 208 and 222 nm. The CD-detected thermal unfolding curve indicates a thermal stability of 64 °C. In general, proteins with high thermal stability should be very well suited for NMR spectroscopy because the NMR signals become sharper at higher temperatures. For NMR backbone resonance assignment, a 1 mM concentrated sample was used for three-dimensional triple resonance NMR experiments [23] (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CO)CACB, and HNCACB that resulted in an almost complete backbone resonance assignment (97%) of all nonproline amino acids in the protein as shown in the two-dimensional (2D) [^{15}N, ^{1}H]-HSQC spectrum (Figure 3b). As expected for an α-helical protein, the signal dispersion of the backbone amide resonances in the proton dimension is low (9.2–7 ppm). However, the linewidths are small, and signal overlap is minimal, which is beneficial for NMR work.

An analysis of secondary chemical shifts, i.e. the deviation from random coil values, can be utilized to map the secondary structure of proteins [24]. In agreement with the CD data, such

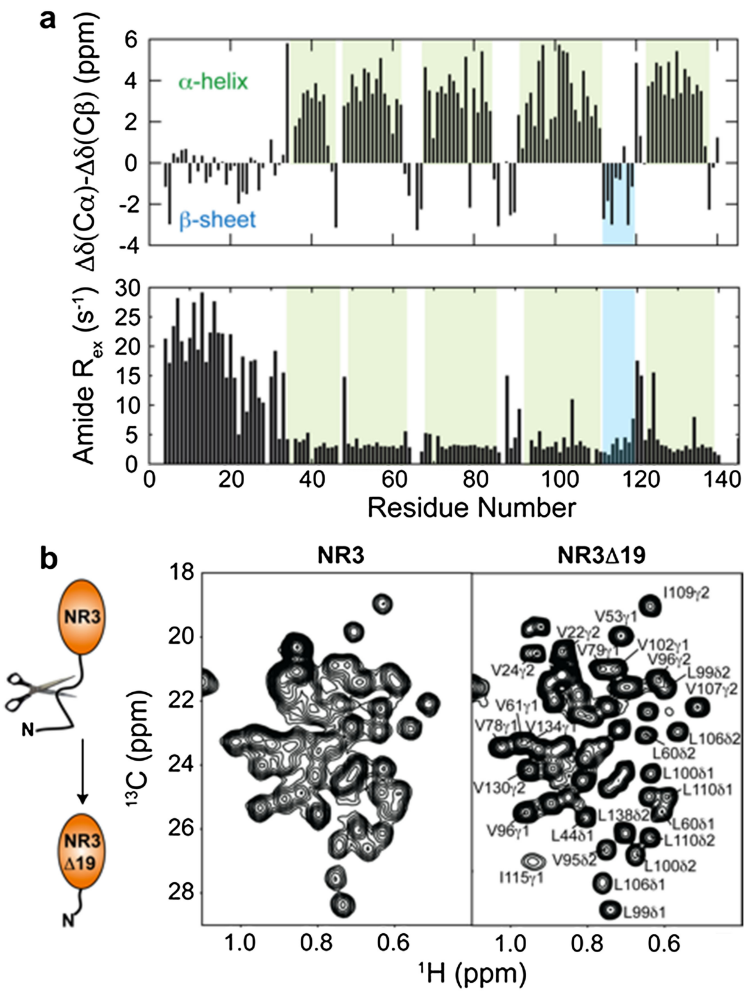

Figure 4. Secondary structure of the C-terminal NR domain.(a) Secondary chemical shifts (ΔδCα-ΔδCβ, upper panel) and amide proton exchange rates (lower panel) show the location of five α-helices and a short stretch of β-sheet conformation and the presence of stable hydrogen bonds, respectively. (b) Comparison between the methyl region of a 2D [^{1}H,^{13}C]-HSQC of NR3 and a truncated version missing 19 residues at the unfolded N-terminus (NR3Δ19).

an NMR analysis reveals mostly α-helical secondary structure with five α-helices and one short β-strand (Figure 4a). This analysis, together with fast amide proton exchange (MEXICO) experiments [25], show that the first 30 amino acids in the proteins do not adopt a secondary structure nor form hydrogen bonds (Figure 4a). Unfolded protein tails generally slow down the motion of the entire protein because of their apparently high hydrodynamic radius. Therefore, we deleted the first 19 residues in the protein and used this construct for further NMR experiments. The improvement in spectral quality is most obvious when looking at the crowded methyl group region in a 2D [^{13}C, ^{1}H]-HSQC (Figure 4b). The deletion variant shows markedly sharper resonances than the initial construct. This gain in resolution is very important for NR3 because it consists of 17% alanine, 25% serine, 25% valine, leucine and isoleucine, and only two tyrosine residues. The high abundance of these residue types results in a severe signal overlap in most regions of the 2D [^{13}C, ^{1}H]-HSQC. However, the increased spectral quality with the truncated protein construct enabled structure determination without the need for selective methyl group labeling routinely performed with larger proteins [26]. To further stabilize this inherently aggregation-prone protein for prolonged NMR experiments, we found that the addition of 1% (v/v) trifluoroethanol could increase the stability of the NMR sample from only one day to more than one week, at a concentration of 1 mM and temperatures above 25 °C.

After a successful side-chain resonance assignment, a set of three-dimensional ^{13}C and ^{15}N-edited NOESY experiments were recorded to obtain pairwise distance restraints for structure determination. In addition, a 2D [^{12}C,^{14}N]-filtered, ^{13}C-edited NOESY experiment was conducted with a mixed ^{12}C, ^{14}N–^{13}C, ^{15}N sample. Therein, selective inter-monomer NOESY contacts between ^{12}C/^{14}N-bound protons on one side and ^{13}C-bound protons on the other side could be detected. These data were crucial for distinguishing between intra-monomer and inter-monomer NOESY contacts and finally enabled a successful structure calculation.

To this end, around 1000 NOE contacts were quantified and used—together with chemical shifts and coupling constants—for structure calculation. The resulting structure of the homodimeric C-terminal NR domain consists of two disulfide-linked helices in the protein core surrounded by a wreath composed of the remaining helices (Figure 5a) [27]. Helix 5 in each monomer points toward the other monomer and acts like a clamp that further stabilizes the dimeric structure. In addition to the disulfide bridge at the N-terminal end of helix 4, there are two salt bridges in each monomer (R43-D93, R52-E101, Figure 5b) that are crucial for NR3 stability. Single and multiple point mutations in that region lead to reduced stability or to a totally abolished thermal unfolding profile (Figure 5c), highlighting the central role of this part for protein stability.

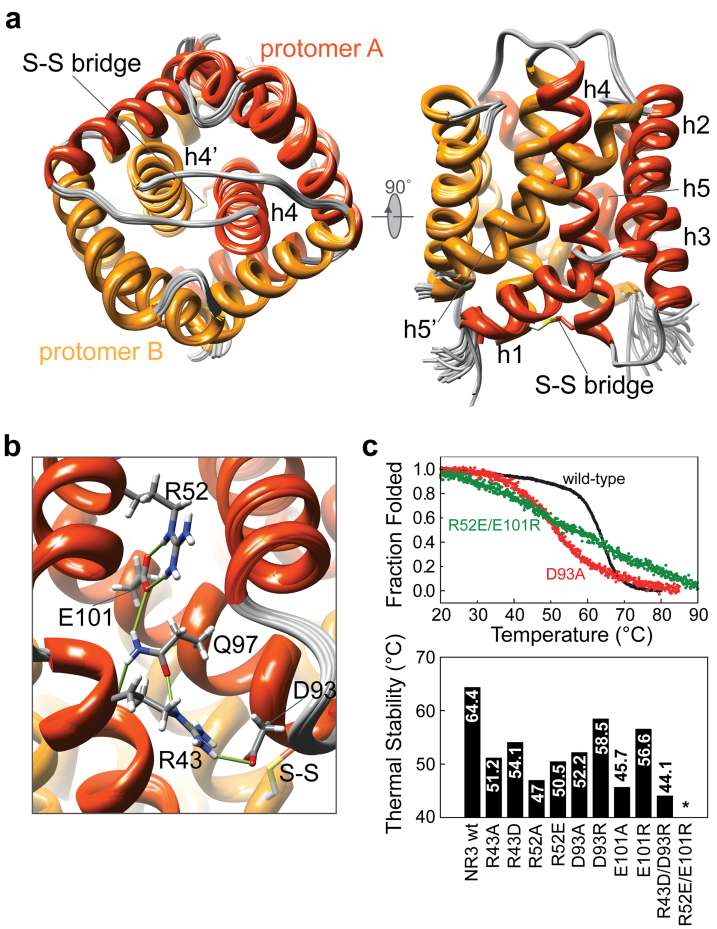

Figure 5. Structure of the C-terminal NR domain.(a) Bundle of the 20 best energy structures of NR3 (2khm.pdb [27]). Each monomer of this new protein fold consists of five α-helices with a long helix serving as main homodimerization site and the other four helices surrounding the central part of the protein. Helix 5 of each monomer points toward the other monomer and acts like a clamp. (b) The two salt bridges (R43-D93, R52-E101) connected to each other by the amide side chain moiety of Q97 in the monomer are crucial for protein stability as probed by CD-detected thermal unfolding curves with NR3 variants. The asterisk in (c) indicates a total loss in the cooperativity of unfolding.

Journal of PeptideScience

The destabilizing effect of such variants can also be observed in protein constructs containing 12 copies of the repetitive sequence modules A (alanine-rich) and Q (glutamine-rich) and the NR domain (AQ_{12}NR3) [16]. Fluorescence dye binding assays indicate a large gain in hydrophobicity upon unfolding of the NR domain (Figure 6a). With a construct of longer repetitive sequence having the same molecular weight (AQ_{24}), no increase in hydrophobicity can be observed. Therefore, the unfolding process of the NR domain can be interpreted as switching between a soluble and an aggregation-promoting position. Upon disrupting the salt bridges in NR3, this switch is prematurely set to an aggregation-prone state. There is strong dye binding already at low temperatures, indicating hydrophobic surface properties. However, high sodium chloride concentration, as present during silk protein storage in the spider's gland, brings back part of the thermal transition profile observed for the wild-type protein and thus seems to stabilize the protein.

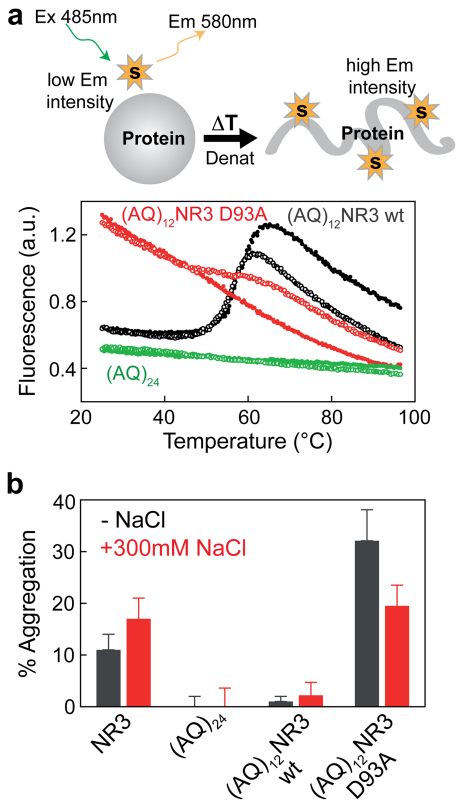

Figure 6. Influence of the C-terminal NR domain on silk protein properties. (a) Top panel: schematic overview of the employed hydrophobic dye binding experiment. Lower panel: unfolding traces of spider silk proteins monitored by the change in fluorescence intensity of the extrinsic hydrophobic dye upon binding to the protein at different sodium chloride concentrations. Open symbols represent high salt buffer (300 mM sodium chloride), filled symbols low salt buffer (50 mM sodium chloride). In contrast to the wild-type $(AQ)_{12}$NR3 domain (black symbols), the salt-bridge variant D93A shows no thermal unfolding transition but a marked increase of dye binding even at low temperatures (red symbols). The unfolding trace can be partly restored for this variant at high salt. As a control, a construct comprised of repetitive sequence elements lacking the NR3 domain shows no thermal unfolding trace (green symbols). (b) Aggregation experiments with different spider silk proteins in the absence of shear forces and at different salt concentrations, as indicated. Because of the more hydrophobic nature of the NR3 variant (D93A), a higher aggregation tendency is observed even in the presence of repetitive sequence elements. $(AQ)_{12}$, $(AQ)_{24}$: repetitive spider silk proteins consisting of 12 or 24 repeats of an alanine (A) and glutamine (Q) sequence element [16]; NR3, C-terminal NR domain.

A similar picture can be observed in aggregation assays with these proteins (Figure 6b). NR3 alone does show some aggregation, but in the presence of the larger repetitive sequence elements, the protein solution is very stable. However, the salt-bridge variant (D93A) leads to significant aggregation even in the presence of the long repetitive elements. The repetitive elements alone do not aggregate spontaneously but only in the presence of shear forces. Shear affects both the repetitive and NR domains, but a well-aligned silk fiber can only be obtained when the NR domain and the repetitive elements are present in a single protein construct [27]. The repetitive elements alone show pronounced precipitation upon shear stress but only form amorphous aggregates instead of a regular fiber. This finding points toward an essential function of the C-terminal NR domain in switching between soluble and aggregation-prone conformations and in promoting the alignment of the repetitive sequence elements for effective fiber formation.

Homology Modeling, Structure Refinement and Function of the N-terminal Domain

Until recently, the role of the N-terminal domain of spider dragline silk was not clear, and it was even thought that this domain might not be part of the final fiber because of cleavage after secretion into the gland of the spider [3]. However, a recent X-ray crystallography study made available the dimeric structure of this domain from the dragline silk of the spider *Euprosthenops australis* [28]. Therein, a link between the pH value and spontaneous fiber formation could be drawn, but a direct connection between solvent conditions and oligomerization of this protein was still missing. Therefore, we used NMR spectroscopy to gain further insight into the underlying mechanism and chose the N-terminal NR domain of major ampullate spidroin 1 (N1, Figure 2 lower panel) of the black widow (*L. hesperus*) [29]. The protein shows high stability in thermal denaturation assays [29] and is, unlike the C-terminal domain, not prone to aggregation. The NMR chemical shift assignment of this domain is shown in Figure 7a. These chemical shift data, together with sparse NOE data and a set of amide H–N residual dipolar couplings, were used for the structure refinement of a homology model of monomeric N1 of the *L. hesperus* spidroin, using the crystal structure of *E. australis* as a template. The best energy structural model obtained by this procedure shows a root mean square deviation to the input structure of 0.08 nm, with the same secondary structure content and protein fold (Figure 7b). Only the last helix is extended by one turn relative to the N-terminal domain of *E. australis*.

The dimerization surface of the monomer is positively charged on one edge and negatively charged on the other side, suggesting a pronounced contribution of electrostatics to dimerization (Figure 7c). We therefore set out to identify the mechanism of dimerization and tried to link this process to the solvent conditions present along the way from the spinning gland to the end of the spinning duct. Size exclusion chromatography combined with multi-angle static light scattering was used to assess the oligomerization state of N1, at a pH present during silk protein storage in the gland (pH7) and present during fiber formation in the spinning duct (~pH6) and at high and low salt (Figure 8a). The data show that the monomer to dimer equilibrium is strongly influenced by the pH, where low pH stabilizes the dimer and neutral pH the monomer. The presence of salt further stabilizes the monomeric state. Only in the presence of

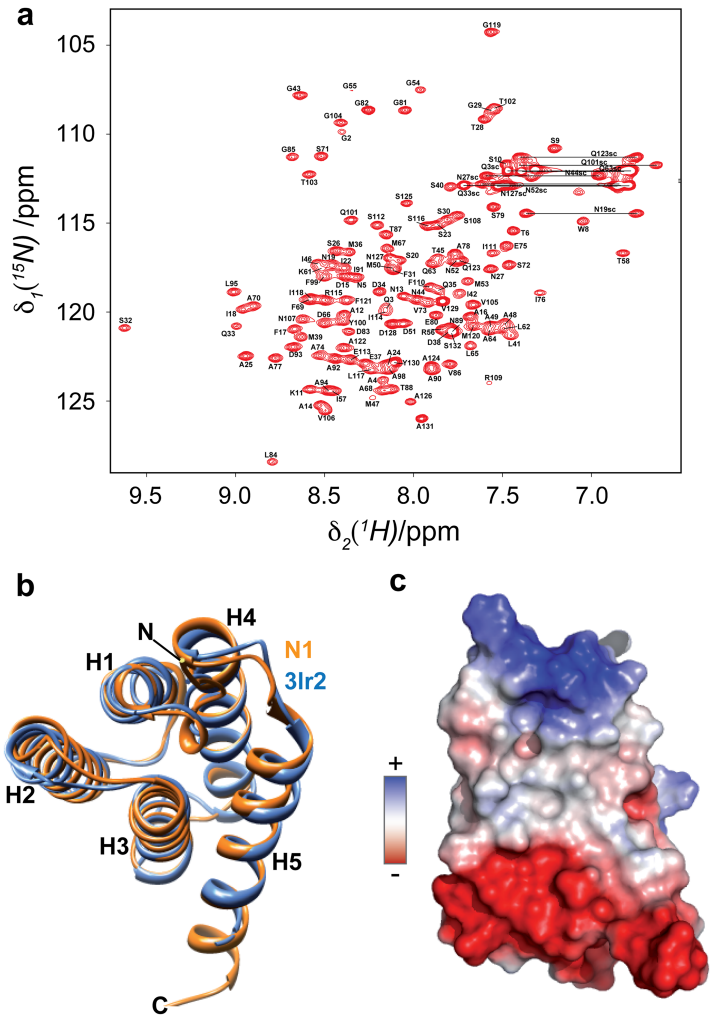

Figure 7. NMR based modeling of the N-terminal NR domain.(a) 2D [^{1}H,^{15}N]-HSQC of the N-terminal NR domain of *Latrodectus hesperus* MAS fibroin 1 (N1) with the NMR assignment of the amide signals labeled. (b) Homology modeling of the N1 monomer structure and subsequent refinement with NMR data (orange ribbon) overlaid onto the crystal structure of the N-terminal NR domain of *Euprosthenops australis* shown in blue (3lr2.pdb) [28]. (c) Electrostatic surface color representation of N1. One part of the dimerization interface is positively charged, and the other is negatively charged.

salt were the actual weights for the monomer and dimer obtained, whereas a slightly higher molecular weight was measured without salt. The effects of salt and change in pH were also mapped onto the structure of N1 by NMR chemical shift perturbations obtained using 2D [^{1}H,^{15}N]-HSQC experiments (Figure 8b). These data show that both pH and salt affect the charged dimerization interface of N1. Salt leads to chemical shift changes (shown red in Figure 8c) and an improvement in the spectral quality, whereas a change in pH severely reduces the quality of the spectra and leads to disappearance of resonances of residues located in the dimerization interface (shown in blue in Figure 8c). The linewidths of the resonances of N1 without salt are broader than in presence of 300 mM sodium chloride, and a lowering of the pH in presence of 300 mM salt again broadens the linewidths because of the formation of a stable dimer (Figure 8b). In summary, the N-terminal NR domain was shown to be a pH sensor that responds to changes in pH and salt concentration by modulation of its monomer to dimer equilibrium, thereby connecting spider silk protein multimerization to the solvent conditions present during fiber assembly.

Model for the Fiber Assembly Process

Taken together, the data herein and previous work [13,18,30] can be used to construct a molecular picture of the fiber assembly process in the spinning gland of the spider (Figure 9). During silk protein storage (Figure 9, top panel) at neutral pH and higher sodium chloride concentration, silk proteins form large micellar assemblies [15] where the hydrophilic and charged NR domains are oriented toward the aqueous phase, whereas the repetitive and amphiphilic sequence elements are located inside the micelles thereby protected from the solvent and from interaction with each other, thus inhibiting the formation of larger aggregates. Micelle formation requires the presence of either of the NR domains [31]. By this type of protein storage, the spider manages to store silk proteins up to a concentration of 40% (w/w). Silk protein storage in large quantities is essential, as the spider immediately requires the material when constructing its web. Furthermore, tight control of aggregation is crucial for the animal because clogging of the spinneret would be fatal. The monomeric state of the N-terminal NR domain couples the neutral pH to inhibition of the aggregation process [28,29],

Journal of PeptideScience

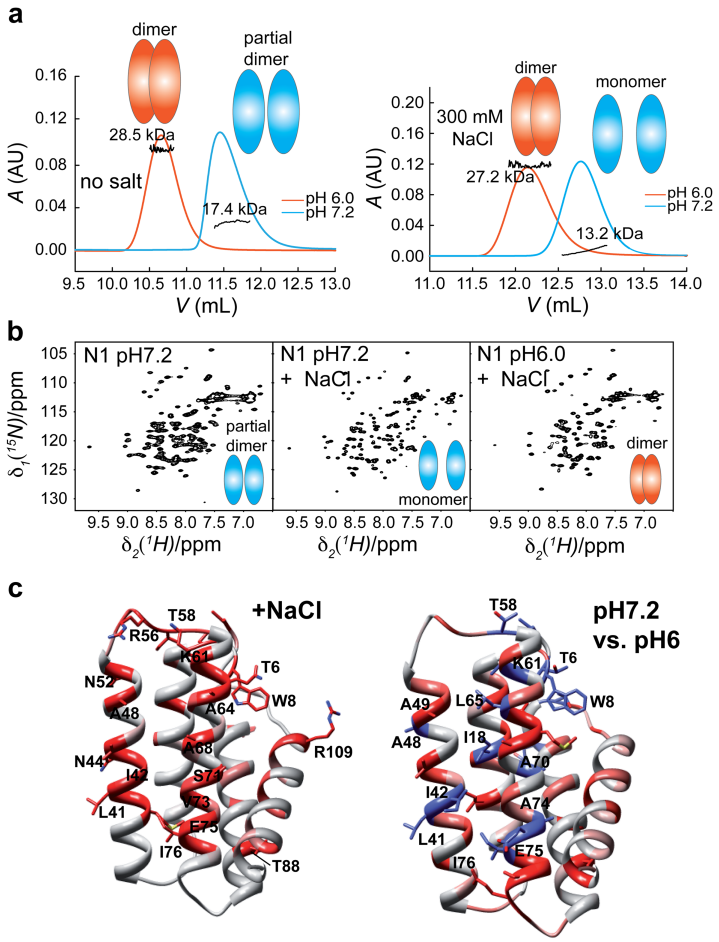

Figure 8. pH and salt-dependent dimerization of the N-terminal NR domain.(a) Analytical size exclusion runs with N1 at different salt concentrations and pH values. The molecular mass of each species was detected with multi-angle static light scattering. The calculated molecular masses are 13.6 and 27.2 kDa for the monomer and the dimer, respectively. (b) 2D [^{1}H,^{15}N]-HSQC spectra of N1 at different sodium chloride concentrations and different pH. (c) Effect of salt and a change in pH on the backbone amide resonances in a 2D [^{1}H,^{15}N]-HSQC spectrum mapped onto the structure of N1. NMR resonances of residues colored in blue disappear at pH6. Figure is adapted from Ref. [29].

whereas the C-terminal NR domain just facilitates the formation of stable protein micelles at this stage [15,16,27,31]. The presence of sodium chloride during silk protein storage further stabilizes the two NR domains and leads to stable micelle formation and prevents phase separation of concentrated silk solutions [15].

The functions of these two regulatory elements change quite drastically during fiber assembly in the spinning funnel of the spider (Figure 9, lower panel). During that process, the pH drops from pH 7 to ~pH 6, the sodium chloride concentration is also reduced whereas the phosphate concentration goes up, water is being removed, and finally there are shear forces accompanied by elongational flow in the spinning duct [18]. The N-terminal NR domain basically acts as a pH and salt sensor and dimerizes at lower pH and low salt [28,29,32,33], leading to multivalent anchoring of individual silk protein molecules. The function of the C-terminal NR domain during silk assembly is more complex. This domain is very stable against thermal denaturation but extremely sensitive to shear forces [27]. Under shear, it drives the controlled aggregation process of the entire silk protein and leads to the formation of anchor points defining the correct register for the alignment of the repetitive sequence elements, which finally enables the formation of an aligned fiber as compared with amorphous aggregates obtained in its absence [27]. The alignment of the repetitive sequence elements is driven by the elongational flow present in the spinning duct of the spider [18], causing these elements to undergo a structural transition from unfolded to β-sheet as proven by solid state NMR [21]. The N-domains and C-terminal domains act in concert and perform different parts of the same process. Together they sense changes in salt, pH, and shear forces; only the fine-tuned interplay between these parameters enables the assembly of a stable silk fiber with its extraordinary mechanical properties [1].

Summary

In this work, the structure and function of the NR domains of spider dragline silk proteins have been presented, and their role in fiber formation has been discussed. It was shown that both domains have rather complementary functions and individually promote controlled fiber formation and its final mechanical properties. However, to artificially produce spider silk materials, which are identical to the natural silks, one not only has to optimize their sequences but also must find optimal spinning conditions. The silk assembly mechanism summarized here may provide valuable insights to achieve these goals in the near future.

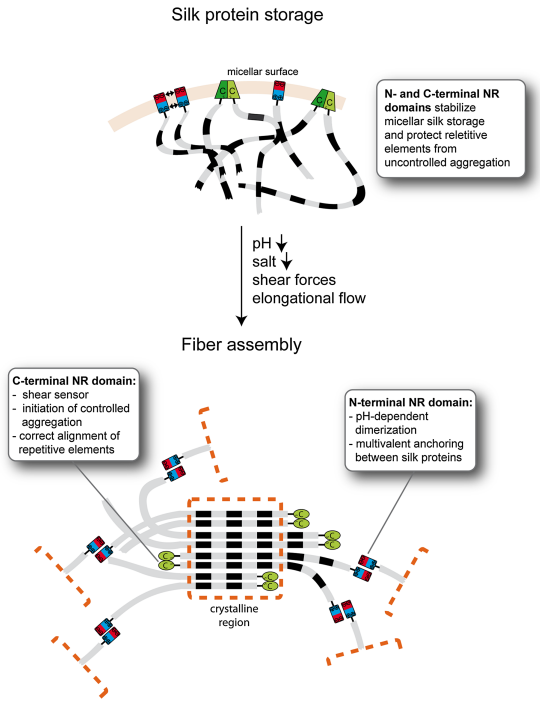

Figure 9. Model of the influence of the non-repetitive domain on fiber assembly.Top panel: because of their polar and/or charged nature, the N-terminal and C-terminal NR domains lie on the surface of protein micelles during silk protein storage in the gland of the spider. Lower panel: during fiber assembly, shear forces and elongational flow are present in the spinning duct, the pH drops, and salt and water are removed from the protein mass, leading to a controlled aggregation process of the silk proteins, governed by the C-terminal NR domain. The soluble N-terminal NR domain serves as a pH-dependent site of multivalent anchoring between silk proteins, leading in principle to protein chains of infinite length, which in turn leads to improved mechanical properties of the fiber.

Acknowledgements

All experiments were performed in the lab of Horst Kessler at the Technische Universität München. I want to thank him for continuous support and guidance as well as for proofreading of the manuscript. Thomas Scheibel, Lukas Eisoldt, Christopher Thamm, and John Hardy are gratefully acknowledged for collaboration in the spider silk field and the kind gift of plasmids used for initial work. I want to thank the Elitenetzwerk Bayern and the elite graduate program CompInt managed by Andreas Bausch and Georg Abstreiter (TU München) for their support. I also want to thank Katie Edmonds (Harvard Medical School) for proofreading of the manuscript. The Bavarian NMR Centre is gratefully acknowledged for NMR time. The chemical shift assignments and structural coordinates are deposited at the BMRB (16249 and 17131) and the RCSB database (2khm.pdb), respectively. Postdoctoral funding by EMBO (ALTF-265-2010) and Human Frontier Science Program (LT000297/2011-L) long-term fellowships is gratefully acknowledged.

References

1 Gosline JM, Guerette PA, Ortlepp CS, Savage KN. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* 1999; **202**: 3295–3303.
2 Heim M, Keerl D, Scheibel T. Spider silk: from soluble protein to extraordinary fiber. *Angew. Chem. Int. Ed.* 2009; **48**: 3584–3596.
3 Lewis RV. Spider silk: ancient ideas for new biomaterials. *Chem. Rev.* 2006; **106**: 3762–3774.
4 Gosline JM, Demont ME, Denny MW. The structure and properties of spider silk. *Endeavour* 1986; **10**: 37–43.
5 Vollrath F. Spider webs and silk. *Sci. Am.* 1992; **266**: 70–76.
6 Rising A, Hjalm G, Engstrom W, Johansson J. N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. *Biomacromolecules* 2006; **7**: 3120–3124.
7 Motriuk-Smith D, Smith A, Hayashi CY, Lewis RV. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules* 2005; **6**: 3152–3159.
8 Sponner A, Vater W, Rommerskirch W, Vollrath F, Unger E, Grosse F, Weisshart K. The conserved C-termini contribute to the properties of spider silk fibroins. *Biochem. Biophys. Res. Commun.* 2005; **338**: 897–902.
9 Romer L, Scheibel T. The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* 2008; **2**: 154–161.
10 Ittah S, Cohen S, Garty S, Cohn D, Gat U. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. *Biomacromolecules* 2006; **7**: 1790–1795.
11 Knight DP, Vollrath F. Liquid crystals and flow elongation in a spider's silk production line. *P. Roy. Soc. Lond. B Bio.* 1999; **266**: 519–523.
12 Knight DP, Vollrath F. Changes in element composition along the spinning duct in a Nephila spider. *Naturwissenschaften* 2001; **88**: 179–182.
13 Jin HJ, Kaplan DL. Mechanism of silk processing in insects and spiders. *Nature* 2003; **424**: 1057–1061.
14 Vendrely C, Scheibel T. Biotechnological production of spider-silk proteins enables new applications. *Macromol. Biosci.* 2007; **7**: 401–409.
15 Exler JH, Hümmerich D, Scheibel T. The amphiphilic properties of spider silks are important for spinning. *Angew. Chem. Int. Ed.* 2007; **46**: 3559–3562.
16 Hümmerich D, Helsen CW, Quedzuweit S, Oschmann J, Rudolph R, Scheibel T. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 2004; **43**: 13604–13612.
17 Slotta U, Hess S, Spiess K, Stromer T, Serpell L, Scheibel T. Spider silk and amyloid fibrils: a structural comparison. *Macromol. Biosci.* 2007; **7**: 183–188.
18 Rammensee S, Slotta U, Scheibel T, Bausch AR. Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. U.S.A.* 2008; **105**: 6590–6595.
19 Knight DP, Knight MM, Vollrath F. Beta transition and stress-induced phase separation in the spinning of spider dragline silk. *Int. J. Biol. Macromol.* 2000; **27**: 205–210.
20 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**: 2399–2407.
21 van Beek JD, Hess S, Vollrath F, Meier BH. The molecular structure of spider dragline silk: folding and orientation of the protein backbone. *Proc. Natl. Acad. Sci. U.S.A.* 2002; **99**: 10266–10271.
22 Ittah S, Michaeli A, Goldblum A, Gat U. A model for the structure of the C-terminal domain of dragline spider silk and the role of its conserved cysteine. *Biomacromolecules* 2007; **8**: 2768–2773.
23 Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 1999; **34**: 93–158.
24 Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* 1994; **4**: 171–180.
25 Gemmecker G, Jahnke W, Kessler H. Measurement of fast proton-exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.* 1993; **115**: 11620–11621.
26 Tugarinov V, Kay LE. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* 2003; **125**: 13868–13878.
27 Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, Scheibel T, Kessler H. A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* 2010; **465**: 239–242.
28 Askarieh G, Hedhammar M, Nordling K, Saenz A, Casals C, Rising A, Johansson J, Knight SD. Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* 2010; **465**: 236–238.

29 Hagn F, Thamm C, Scheibel T, Kessler H. pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk—implications for fiber formation. *Angew. Chem. Int. Ed Engl.* 2011; **50**: 310–313.
30 Vollrath F, Knight DP. Liquid crystalline spinning of spider silk. *Nature* 2001; **410**: 541–548.
31 Lin Z, Huang W, Zhang J, Fan JS, Yang D. Solution structure of eggcase silk protein and its implications for silk fiber formation. *Proc. Natl. Acad. Sci. U.S.A.* 2009; **106**: 8906–8911.
32 Gaines WA, Sehorn MG, Marcotte WR, Jr. Spidroin N-terminal domain promotes a pH-dependent association of silk proteins during self-assembly. *J. Biol. Chem.* 2010; **285**: 40745–40753.
33 Landreh M, Askarieh G, Nordling K, Hedhammar M, Rising A, Casals C, Astorga-Wells J, Alvelius G, Knight SD, Johansson J, Jornvall H, Bergman T. A pH-dependent dimer lock in spider silk protein. *J. Mol. Biol.* 2010; **404**: 328–336.
34 Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982; **157**: 105–132.
35 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–2948.
36 Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 2003; **31**: 3320–3323.

