Spider Silk Proteins: Structure and Function Revealed by NMR Spectroscopy

This article discusses the structure and function of important regulatory elements in spider silk, the presence of which is vital for thread formation. Using nuclear magnetic resonance spectroscopy (NMR), both the three-dimensional structure as well as important findings on the function of these protein domains could be determined for the first time.

Introduction
To date, spider silk is one of the most tear-resistant biomaterials known, with properties that exceed modern high-performance synthetic materials by far. The evolutionary optimization of spider silk over more than 450 million years has created a material that is harder than steel, yet also elastic and flexible like rubber [1-3]. This variety is made possible by up to six different silk glands in the female spider, each of which produces different silk proteins with varying properties. The superb material properties, low weight and biocompatibility make spider silk a sought-after material.

This article primarily seeks to address the structure of N- and C-terminal domains in the very robust dragline thread of the spider. Based on this data, a thread formation model will finally be discussed, which reveals the essential function of both domains in the context of the overall silk proteins.

Properties of Different Silk Proteins
The spider uses silk threads from different silk glands for a variety of purposes [2,3]. For instance, the major ampullate (MA) silk gland produces silk for the robust dragline thread and the frame of the web (major ampullate spidroin (MAS)). An auxiliary thread (minor ampullate spidroin (MIS)) is produced by the minor ampullate (MI) gland, which is the first to be used, and which serves as a template for the final web. Apart from the frame, the spider web consists of flagelliform silk, which is produced in the flagelliform gland. This silk is very elastic, so that the kinetic energy of prey can be absorbed during capture. The spider also uses silk to wrap its eggs. This covering consists of an inner layer, which originates from the aciniform silk gland, and an outer layer, which is produced in the tubuliform gland.
Additionally, the spider requires a further material to stick together the individual threads and to coat the sticky spiral.

This material comes from the tubuliform gland. The spider then draws the required silk material from the respective gland, as required.

**Chemical Composition**

Silk proteins consist of long and repetitive amino acid sequence modules. A distinction is made between the modules that lend strength to the silk and those that lend elasticity [1]. A hard silk thread contains a high percentage of crystalline areas, which non-covalently join together different protein chains, but which are non-tensile due to the rigid structure. The optimal amino acid sequences for generating these crystalline areas are polyalanine and polyalanine/glycine blocks which presumably form β-sheet-type structures. An elastic thread is formed when blocks, made, for example, from GPGQQ (G:Glycine, P:Proline, Q:Glutamine), are included in greater numbers [1-3]. In solution and probably also in the micelle storage form (see below), these repetitive structures remain in an unfolded state. There are, however, non-repetitive protein domains structured in solution at the N- and C-terminal ends of the silk proteins [2-5] (fig. 1). The structure and function of these areas of spider silk proteins were unknown until recently and are discussed below.

**Structure and Function of the C-Terminal Non-Repetitive Domain**

The region with the highest conserved amino acid sequence among the spider silk proteins is the C-terminus with at least 45 % identity [4,6] between the least related pairs in this group (fig. 2a). It was shown that the C-terminus is contained in the finished thread and is undifferentiated [7]. On this basis, it has been assumed that this domain plays an essential role in controlling the solubility of silk proteins. This domain is also important for thread formation, which is induced by a change in the ionic composition and by mechanical stimuli [8]. In the process, the repetitive
sequence elements are aligned with one another, thereby enabling the formation of β-sheet structures. Despite a range of published work, the function of this domain remains unclear. Moreover, no three-dimensional structure has been achieved to date, due to the strong aggregation tendency and the low stability of these proteins.

Our method of choice is nuclear magnetic resonance spectroscopy (NMR), which allows the structure of molecules to be examined in solution (see below). As opposed to X-ray structural analysis, the proteins do not have to exist in the form of crystals, but can be investigated in their native-like environment.

Production of the C-terminal domain in the garden spider (Araneus diadematus) was achieved using a synthetic gene from the known natural sequence in the Escherichia coli bacterium, in which identification with the NMR-active nuclei $^{13}$C and $^{15}$N can also be achieved, depending on the composition of the culture medium. Since these spider silk proteins are very prone to aggregation, the first thing to do is to look for appropriate co-solvents, which stabilize the protein in its soluble form. It emerges that 1 % (v/v) Trifluoroethanol (TFE) is sufficient to increase the stability of the proteins from one day to more than seven days. For this reason, a complete NMR analysis was possible (fig. 2b). With the aid of two- and three-dimensional NMR experiments, several thousand paired distances could be established between the protons in the protein, which, in turn, could be used for structure calculation [6].

The C-terminal domains form dimers, which are linked together via a disulfide bridge. A critical point in determining the structure of homodimeric proteins by means of NMR spectroscopy is the differentiation of NOE contacts within a protomer, from contacts between the protomers. This can be achieved with isotope-filtered NOESY experiments. Usually unmarked proteins and proteins labeled with $^{13}$C/$^{15}$N are mixed at a ratio of 1:1 and then the contacts of $^{13}$C/$^{15}$N-bound protons are selectively measured against $^{12}$C/$^{14}$N (unlabeled)-bound protons. In this special case, it isn't as easily achieved, due to the covalent link via the disulfide bridge. Hence, re-folding of the protein under reducing conditions was performed after mixing isotope-labeled and unlabeled proteins at a ratio of 1:1. Once clear NOE contacts were present, the structure of this protein could be calculated (fig. 2c).

The structure of this homodimeric protein, which is covalently fixed via a disulfide bridge, represents a new folding topology [6] (fig. 2d). In the center of the structure are two long α-helices each from different protomers, which are connected at the N-terminal end via the disulfide bridge. The remaining four helices for each protomer form an annulus around the nucleus of the structure. In each protomer, there are two salt bridges that are made up of the pairs Arg43-Asp93 and Arg52-Glu101. The
area of the structure in which the salt bridges are located is also the least robust part. In thread formation, these salt bridges are broken by the occurring shearing forces. This leads to destabilisation of the entire protein and ultimately to its unfolding. This step is essential for thread formation.

**Structure and Function of the N-Terminal Non-Repetitive Domain**
The crystal structure of the N-terminal domain of MA silk from Euprosthenops australis revealed that it acts as a pH sensor that benefits thread formation during the spinning process at low pH (pH6), and inhibits this process during silk protein supply at neutral pH (pH7.2) [9]. However, it was not clear which processes this function is based on.

Using the N-terminal domain of another orb-web spider, the black widow (*Latrodectus hesperus*), we employed NMR spectroscopy and static light scattering to clearly demonstrate that this protein exists as a monomer at neutral pH, and as a dimer at low pH (fig. 3) [10]. The presence of sodium chloride, as it appears in silk protein storage, also benefits the monomeric form. This data shows that this domain couples its aggregate state with the pH-value and, in doing so, contributes to multivalent cross-linking of the individual silk proteins during thread formation, which is an important reason for the high strength of the fiber.

**Model of Fiber Assembly**
The N- and C-terminal non-repetitive domains of the MA spidroin are crucial for fiber formation. The silk proteins are stored in the spider as micellar structures (fig. 4). In these structures, the relatively hydrophilic non-repetitive domains are located on the exterior (oriented towards the solvent) rendering the surface of the micelle polar and charged. The repetitive elements stay unstructured inside, which causes a protective barrier of fiber-forming elements and thereby prevents an early aggregation of the proteins inside the spider. In addition, the N-terminal domains at the surface of the micelle repel each other, because its charge at ~pH7 does not permit dimerization (fig. 4 left). This storage form is stabilized through the presence of increased sodium chloride concentrations. In fiber formation, NaCl is exchanged for sodium phosphate, shear forces occur, and water is drawn out of the protein mass. Through these stimuli, the C-terminal domain unfolds and exposes hydrophobic surfaces, which enable non-covalent cross-linking of different silk proteins. By this process, the repetitive elements are also brought into the appropriate register, which is essential for efficient fiber formation.

The N-terminal domain is less aggregation-prone than the C-terminus. The function of this part of the protein involves pH-dependent dimerization with the N-terminal ends of other protein strands, which theoretically results in an infinitely long chain
with optimum mechanical properties (fig. 4).

To date, it has not yet been possible to produce spider silk proteins with properties similar to the natural proteins. Problems determine the expression of long, highly repetitive sequences and also the length of the proteins themselves (Mr=200-300 kDa). The optimal conditions for spinning the thread are also unknown. Only after overcoming these problems a large-scale production and application of this interesting material may be possible.

**Literature**


For more references please contact the authors.

**Nuclear Magnetic Resonance (NMR) Spectroscopy**

NMR is a spectroscopic method that detects the electronic environment of individual nuclei and their interaction with neighboring nuclei. Only isotopes of uneven mass or atomic number can be investigated - for example, $^1\text{H}$, $^{13}\text{C}$ and $^{15}\text{N}$. The naturally occurring carbon or nitrogen isotopes ($^{12}\text{C}$ and $^{14}\text{N}$) are not NMR active. Therefore, the $^{13}\text{C}$ and $^{15}\text{N}$ isotopes must first be enriched in order to detect these nuclei. This can be achieved by adding $^{13}\text{C}$-glucose or $^{15}\text{N}$-amonium salts to the culture medium in which the bacteria that produce the respective protein are cultured.

So-called Nuclear Overhauser Effect Spectroscopy (NOESY) is used for measuring pairwise distances between nuclei. Hence, distances of up to 6 Å can be detected, usually between protons. Hundreds to thousands of these distances are necessary for structural analysis.

**Teaser Image: Copyright: Antje Lindert-Rottke - Fotolia.com**

**Authors**

Dr. Franz Hagn, Prof. Dr. Horst Kessler

**Contact**

*Harvard Medical School*