4
Spider Silk Proteins from Transgenic Plants

Dr. Jürgen Scheller[1], **Dr. Udo Conrad**[2]
[1] Institute for Plant Genetics and Crop Plant Research (IPK), Leibniz Institute, Corrensstr. 3, 06466 Gatersleben, Germany; Tel.: +49-39482-5382; Fax: +49-39482-5138; E-mail: scheller@ipk-gatersleben.de
[2] Institute for Plant Genetics and Crop Plant Research (IPK), Leibniz Institute, Corrensstr. 3, 06466 Gatersleben, Germany; Tel.: +49-39482-5253; Fax: +49-39482-5138; E-mail: conradu@ipk-gatersleben.de

gfp	green fluorescent protein
HFIP	hexafluorisopropanol
MaSp	major spidroin
MiSp	minor spidroin
SD1, SO1SO1	synthetic spider silk genes/proteins
SLP	silk-like protein

1 Introduction

Plants and animals produce a diverse pattern of soft and hard tissues in order to ensure survival of the organisms under different environmental conditions. These specific properties of different tissues are based on a relatively limited set of starting materials as proteins, polysaccharides, and simple minerals. Specific tissues that provide mechanical features, permeability, optical properties and interaction with the environment must be developed under ambient conditions, often without external energy sources (Tirrell, 1996). Materials scientists are increasingly willing to investigate this fascinating biology for lessons in protein design and molecular evolution (Hayashi and Lewis, 2001), and indeed spider silk was one of the first such examples in this field. The evolutionary development of spiders was accompanied by the development of diversity, production and use of specific silks (for a review, see Hinman et al., 1992). Dragline silk, which is used by spiders to create the frame of their webs and also as safety lines, shows a combination of strength and toughness that is superior to that of many high-performance synthetic fibers (Table 1) (Tirell, 1996; Hinman et al., 2000; Vollrath and Knight, 2001). Spider silk genes encode proteins composed of iterated peptide motifs, and the consensus sequences are multiply repeated throughout the silk proteins (Hinman et al., 2000). Patterns of alternating Ala-rich blocks and Gly-rich amorphous blocks provide both strength and elasticity (Table 2) (Gosline et al., 1999; Hayashi and Lewis, 2000; Hinman et al., 2000), and such biomaterials might therefore be valuable for industrial and medical purposes. The modular nature of spider silk proteins has led to several attempts to design synthetic genes (for reviews, see Gosline et al., 1999; Hinman et al., 2000) and to express them in microorganisms. Because spider silk proteins consist largely of glycine and alanine, an extensive pool of these two amino acids must be provided if spider silk proteins are to be produced by rapidly growing microorganisms, including bacteria and yeasts. Another problem in the bacterial production of spider silk is genetic instability due to recombination, resulting from the highly repetitive genes that encode the spider silk proteins. Experiments to express high molecular-weight silk genes in *Escherichia coli* and *Pichia pastoris* resulted also in the synthesis of truncated proteins as a limiting factor for bioengineering of spider silk (Fahnestock et al., 2000). In a recent publication, Lazaris et al. (2002) reported the successful production in cultured mammalian cells of spider silk proteins ranging from 60 to 140 kDa in size. Moreover, fibers exhibiting toughness and modulus values comparable with those of native drag line silk, but with a lower tenacity, were spun from a concentrated aqueous solution of a single spider silk protein (60 kDa). Although this seemed to open the way for the production of spider silk from animal cells, mass production for technical purposes from animal cells seems to be both expensive and limited in amount, mainly due to the large fermentors, sterile conditions and expensive media that must be provided. In order to overcome all these and other limitations, plants were chosen for the production of spider silk proteins (Scheller et al., 2001). Slowly growing organisms such as plants can easily provide the extensive amino acid pools necessary for efficient spider silk protein synthesis. Moreover, plants are grown in large areas worldwide, not only for food production but also to provide resources for industrial purposes. Hence, mass production of plants is common, and large-scale plant production and

harvesting technologies are well developed. In the past, plants have been used successfully for the production of different transgenic protein products (for a review, see Giddings, 2001),and the storage organs of crop plants (e.g., seeds and tubers) have long been used for the stable expression of xenogenic proteins (Fiedler and Conrad, 1995; Artsaenko et al., 1998; Perrin et al., 2000; Saalbach et al., 2001). In several cases, the stable accumulation of functional proteins to high levels has been achieved by retention in the endoplasmic reticulum (ER) of plant cells (Wandelt et al., 1992; for a review, see Conrad and Fiedler, 1998). Spider silk proteins have been successfully produced using this approach in both transgenic tobacco and potato plants. Furthermore, the extreme heat stability of these plant-produced synthetic spider silk proteins has been utilized in the development of simple purification procedures.

2 Historical Outline

During a 400-million-year period of evolution, spiders have developed a set of fibers which have a wide range of mechanical properties (Shear et al., 1989). Such diversity has been enforced by the central role that silk plays in the spider's life, for example prey capture, shelter construction, and reproduction.

The various physical and mechanical properties of silk have been studied extensively over the past 50 years. In addition, much effort has been made in attempting to understand the spinning process and to develop artificial spinning techniques. However, until now very few spider silk genes have been partially sequenced, and most of these are from orb-weaving spiders. Based on the results of these studies, several attempts of heterologous expression of spider silk proteins have been described, and this section will elucidate on the historical background.

Tab. 1 MaSpI, MaSpII and Flag: uses in silk and consensus amino acid repeats of sequenced areas

Silk	*Use*	*Spinneret*	*Known proteins*	*Consensus amino acid repeat*
Major ampullate dragline	Web frame and radii	Anterior	MaSpI	GGAGQGGYGGLGGQGAGR GGLGGQ$(GA)_2A_5$
			MaSpII	$(GPGGYGPGQQ)_2GPSGPGSA_8$
Flagelliform	Elastic capture spiral	Posterior	Flag	$(GPGGX)_{43\text{-}63}$-$(GGX)_{12}$-flag spacer

Tab. 2 Mechanical properties of spider silk and other structural materials

Material	*Strength (N m^{-2})*	*Elasticity (%)*	*Energy to break (J kg^{-1})*
Dragline silk	4×10^9	35	1×10^5
Flagelliform silk	1×10^9	>200	1×10^5
Kevlar	4×10^9	5	3×10^4
Rubber	1×10^6	600	8×10^4
Tendon	1×10^9	5	5×10^3
Nylon, type 6	4×10^7	200	6×10^4

X-ray studies of various fibers from the orders Lepidoptera (moths) and Araneae (spiders) were first performed in 1960, by Warwicker, who proposed that all fibroins were composed of crystalline regions of antiparallel sheets of polypeptide chains and nonstructured amorphous regions. The crystalline arrays were thought to be responsible for the stiffness of the fiber. In 1984, Gosline et al. found that the amorphous regions (55–60% of spider silk) were more or less kinetically free, random-coil molecules that could change their shape under the influence of an external load (Gosline et al., 1984). In other words, these regions could behave as rubber networks and exhibit entropic elasticity.

Denny, in 1976, and Wainwright et al., in 1982, demonstrated the remarkable mechanical properties of spider silk fibers and compared them with the properties of other structural materials. Dragline silk was found to be stronger than high- tensile steel, and on a weight basis almost matched the stiffness or strength of the high-performance p-aramid fiber, Kevlar. The major advantage of dragline silk is its unique combination of strength and the high degree to which it can be extended before reaching its break point (see Table 1).

In 1998, Liivak et al. constructed a minimized wet-spinning apparatus to spin fibers of silk proteins. The apparatus was capable of spinning fibers from solutions containing 10 mg of purified protein. Spinning experiments were performed with solubilized silk from the silkworm *Bombyx mori* (Liivak et al., 1998) and the golden orb-weaver *Nephila clavipes* (Seidel et al., 1999), and regenerated silk was produced. Interestingly, recombinant spider silk proteins could not be processed into a fiber. The starting material was solubilized in hexafluorisopropanol (HFIP) in a sophisticated manner described previously by Lock et al. (1993) and Fahnestock (1994). A typical working concentration for spinning was 2.5% (w/w) silk fibroin in HFIP. The spinning solution was pressed through a small needle (80–250 µm inner diameter) into a precipitation bath (methanol for silk proteins of *B. mori* or acetone for silk proteins of *N. clavipes*), whereupon the proteins immediately precipitated and coagulated into a fiber. The best regenerated fibers had a maximum stress which approached that measured for native fibers of *B. mori*.

Although the spinning mechanism of a spider is not fully understood, major efforts have been made to shed light upon the process (Vollrath and Knight, 2001). In contrast to the harsh organic solvents and high pressures currently needed for the artificial spinning of solubilized silk, spiders have developed mechanisms to spin fibers from a highly concentrated, water-based protein solution. In 1991, Kerkam et al. emphasized that the spinning dope was formed to behave like a liquid crystal, where the single molecules were already aligned approximately parallel to one another, but not fully packed. This allowed the viscous solution ($>50\%$ protein) to be processed using a slow flow rate and low pressure (Knight and Vollrath, 1999). In 1999, Vollrath and Knight described a process called "internal tapedown taper", where the perpendicular solutions were rapidly extended and the thread was pulled away from the walls of the duct. It is possible that the higher stress forces during this extension join together the dope molecules with hydrogen bonds to produce the anti-parallel beta-conformation of the final thread. During this phase separation, the silk becomes increasingly hydrophobic and loses its water. In 2001, Knight and Vollrath showed that a slight acidification of the duct solution further promoted this process of silk formation.

In 1990, Xu and Lewis isolated a partial cDNA clone encoding a dragline silk fibroin (*MaSpI*). For this purpose, they constructed a cDNA library from the silk glands of *N. calvipes*. The authors sequenced the 2.4 kbp of the 3′ end of the original *MaSpI*-mRNA. The predicted amino acid sequence showed a highly repetitive structure of the protein, though the repeat units were not rigidly conserved and consisted of polyalanine segments of six to nine amino acids followed by Gly-Gly-Xaa repeats, with Xaa being alanine, tyrosine, leucine, or glutamine. The sequence contained virtually no proline, which was known to account for 3.5% of the amino acid composition of spider silk. In addition, the authors carried out a Northern blot (RNA) analysis which showed bands at 7, 9 and 12 kb. These results indicated the existence of multiple spider silk genes with similar sequence patterns.

In 1992, Hinman and Lewis isolated a partial cDNA clone for a second dragline silk fibroin (*MaSpII*), thereby showing dragline silk to be composed of two proteins, MaSpI and MaSpII. The full length of the *MaSpII*-clone was 2 kbp. The predicted amino acid sequence showed a highly repetitive structure of the protein, with the structural elements consisting of polyalanine segments of six to nine amino acids, followed by repeating pentapeptides such as Gly-Tyr-Gly-Pro-Gly, Gly-Pro-Gly-Gly-Tyr and Gly-Pro-Gly-Gln-Gln.

In 1998, Hayashi and Lewis isolated silk-cDNA clones for a gene *Flag* from flagelliform silk (elastic capture spiral silk), and later, in 2000, demonstrated the genomic intron-exon structure of this gene. *Flag* has a coding sequence of 15,500 bp, and the entire *Flag* locus spans over 30 kb and consists of 13 exons. The flagelliform protein is largely composed of iterated sequences, with the dominant repeat Gly-Pro-Gly-Gly-Xaa appearing up to 63 times in tandem arrays. The authors proposed an "elastin" spring-like helix (β-spiral) of Flag as the basis for the higher elasticity of flagelliform silk than dragline silk (200% versus 35%). The protein does not contain polyalanine segments. *MaSpI* as well as *MaSpII* and *Flag* have highly conserved nonrepetitive C-terminal regions of unknown function. Beyond that, FLAG contains a nonrepetitive N-terminal region of unknown function, and it is likely that MaSpI and MaSpII also contain nonrepetitive N-terminal regions (for an overview, see Table 2).

Other groups (Fahnestock et al., 1994, patent application; Prince et al., 1995; Lewis et al., 1996; Arcidiacono et al., 1998) also reported the expression of synthetic spider silk proteins in *E. coli*, while in 1997 Fahnestock and Bedzyk showed the production of synthetic spider silk proteins in *P. pastoris*. However, the latter authors did not demonstrate any spinning of proteins into fibers, perhaps due to the low expression levels and poor solubility of the recombinant spider silk proteins in water-based buffers.

The silk of the silkworm consists of both heavy and light chains, and the fibers are less strong in comparison with dragline silk. The genomic sequence of the L-chain of *B. mori* was published by Kikuchi et al. (1992), and that of the heavy chain later by Zhou et al. (2000). In 1999, Yamao et al. reported a stable genetic transformation of the silkworm by baculovirus-based gene targeting. These authors cloned the *B. mori* fibroin *light* (*L*)*-chain* and inserted the green fluorescent protein (*gfp*). The chimeric *L-chain-gfp* gene was then inserted into the genomic L-chain via homologous recombination, the chimeric gene expressed in the posterior silk gland, and the gene product then spun into the cocoon layer. The authors claimed that the introduction of foreign genes downstream of a powerful promoter, such as the *fibroin*

promoter, would allow large-scale production of recombinant protein in the silkworm. In this way, it might be possible to transform a silkworm into a spider silk-spinning insect.

In 2001, Scheller et al. constructed stable transgenic tobacco and potato plants for the production of spider silk proteins. These authors were able to demonstrate the accumulation of up to 2% spider silk protein of total soluble proteins, which were encoded by synthetic genes in the range from 420 to 3600 bp. The genes were assembled in a manner such that a very high homology to the native genes ($>90\%$) was achieved. The recombinant proteins (10–100 kDa) were resistant to heat stress and could be selectively precipitated using ammonium sulfate.

In 2002, Lazaris et al. showed the production of known partial native spider silk genes *MaSpI*, *MaSpII* and *Adf3* in mammalian cells with molecular masses of 60 to 140 kDa. Additionally, these authors reported the spinning of silk monofilaments from a concentrated aqueous solution ($>20\%$ protein) of recombinant spider silk protein. The water-insoluble fibers exhibited toughness and modulus values which were comparable with those of native dragline silk, but with lower tenacity. A future aim is to produce recombinant spider silk proteins in the milk of goats.

3
Molecular Genetics and Biochemistry

Unlike silk from the silkworm, spider silk cannot be obtained in sufficient quantities from spiders. As mentioned above, many attempts have been made to develop alternative strategies for obtaining spider silk or spider silk proteins in large quantities. Based on a new gene synthesis approach, synthetic spider silk genes were constructed such that the recombinant proteins exhibited homologies of $>90\%$ in comparison with the native model (MaSpI). The next step was to generate transgenic tobacco and potato plants that expressed remarkable quantities of recombinant synthetic *N. clavipes* dragline proteins (MaSpI) in the ER. Using the ER-expression system, spider silk proteins of up to 100 kDa could be detected in tobacco and potato leaves, and also in potato tubers. The recombinant spider silk genes exhibited extreme heat stability, and this property was used subsequently to purify the spider silk proteins by a simple and efficient procedure.

3.1
Construction of Synthetic Spider Silk Genes and Transgenic Plants

For the initial expression analysis of recombinant spider silk proteins in transgenic plants, synthetic spider silk genes of different molecular masses were designed to match the known *N. clavipes* MaSp1 cDNA. In future, it might be possible to alter the modular structure of the synthetic genes by combining them with other sequences – a strategy that would lead to the production of spider silk-like proteins with different and novel properties. In addition, different synthetic fusion genes could be obtained relatively simply using this approach. With this technique, 18 synthetic oligonucleotides were used to assemble basic cassettes that were then inserted into pUC19-derived plasmids and cloned in *E. coli*. In a second step, the basic cassettes were then combined to obtain increasingly larger segments and finally synthetic spider silk genes (Figure 1A). Six synthetic spider silk gene constructs coding for proteins in the range from 10 to 100 kDa were used to test the plant expression system and to provide a panel of spider silk proteins for the optimi-