Spider Dragline Silk

Molecular Properties and Recombinant Expression

Anna Rising

Faculty of Veterinary Medicine and Animal Science
Department of Biomedical Sciences and Veterinary Public Health and
Department of Anatomy, Physiology and Biochemistry
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2007

Acta Universitatis Agriculturae Sueciae
2007: 38

ISSN 1652-6880
ISBN 978-91-576-7337-4
Cover illustration: (Upper Left) Female *Euprosthenops australis* carrying an egg case (photo: Rising, A.); (Upper Right) Fibre made from recombinantly produced miniature spidroins. The tube diameter is 3 cm (photo: Rising, A.); (Lower Left) Scanning electron micrograph of the point of breakage after tensile testing of a recombinant fibre (from Stark *et al.*, 2007); (Lower Right) Helical wheel presentation of the five predicted α-helices in the spider silk N-terminal domain (from Rising *et al.*, 2006).

Tryck: SLU Service/Repro, Uppsala 2007

Abstract

Rising, A. 2007. *Spider dragline silk – molecular properties and recombinant production.* Doctor's dissertation.
ISSN: 1652-6880, ISBN: 978-91-576-7337-4

Spider dragline silk possesses several desirable features of a biomaterial; it has extraordinary mechanical properties, is biocompatible and biodegradable. It consists of large proteins, major ampullate spidroins (MaSp:s), that contain alternating polyalanine- and glycine-rich blocks between non-repetitive N- and C-terminal domains. No full length MaSp gene has been cloned, hence the knowledge of their constitution is limited. The spider stores the silk in a liquid form, which is converted into a fibre by a poorly understood mechanism.

Even truncated spidroins are difficult to produce recombinantly in soluble form. Most previous attempts to produce artificial spider silk fibres have included solubilization steps in non-physiological solvents and the use of spinning devises for fibre formation.

This thesis presents a novel method for production of macroscopic fibres under physiological conditions, without using denaturing agents. A miniature spidroin is identified that can be produced recombinantly in *E. coli* when fused to a soluble fusion tag. Upon enzymatic release of the fusion tag, the miniature spidroins spontaneously form macroscopic fibres in physiological solution. These fibres resemble native silk and their strength equals that of fibres spun from regenerated silk. Initial studies suggest that the fibres are biocompatible. This represents a major breakthrough for future biomaterial development.

Molecular studies of cDNA and genetic sequences encoding the dragline silk revealed an unexpectedly high level of heterogeneity and the presence of at least two MaSp1 genes. Furthermore, the *E. australis* MaSp2 was characterised for the first time, as well as a new MaSp-like spidroin.

Sequence analysis of previously published spidroin N-terminal domains compared with that of *E. australis* MaSp1, enabled identification of signal peptides and a ~130 residue non-repetitive domain common to dragline, flagelliform and cylindriform spidroins. Moreover, this highly conserved N-terminal domain was concluded to consist of five positionally conserved α-helices.

Structural studies using circular dichroism spectroscopy on recombinantly produced MaSp1 N- and C-terminal domains showed that these are folded, stable and soluble, and that salts or pH has no major effect on their secondary structures.

Keywords: Spider, Major ampullate spidroin, *Euprosthenops australis*, dragline, silk, recombinant expression

Author's address: Anna Rising, Department of Biomedical Sciences and Veterinary Public Health, SLU, Box 7028, SE-750 07 UPPSALA, Sweden and Department of Anatomy, Physiology and Biochemistry, SLU, Box 575, S-751 23 UPPSALA, Sweden.

To my family

Contents

Introduction, 9
Definition of silk, 9
Silks and silk glands in spiders, 9
The major ampullate gland, 12
The formation of a solid fibre, 13
The dragline silk proteins and their encoding genes, 14
Secretory proteins and signal peptides, 16
The dragline silk, 16
Mechanical properties, 18
Supercontraction, 20
Recombinant expression of dragline silk proteins, 21
Production in Escherichia coli, 21
Production in yeast, 23
Production in cell culture, 23
Production in plants, 23
Evolutionary aspects, 24
Bombyx mori silk, 26
Amyloid proteins and fibrils, 27

Present investigation, 29
Aims and scope of thesis, 29
Materials and experimental procedures, 29
Euprosthenops australis, 29
Circular dichroism (CD) spectroscopy, 30
Phylogenetic analysis, 31
Secondary structure predictions, 32
Results and discussion, 33
Macroscopic fibres self-assembled from recombinant miniature spider silk proteins (I), 33
Major ampullate spidroins from Euprosthenops australis: multiplicity at protein, mRNA and gene levels (II), 34
N-terminal nonrepetitive domain common to dragline, flagelliform and cylindriform spider silk proteins (III), 35
Structural properties of non-repetitive and repetitive parts of major ampullate spidroin 1 from Euprosthenops australis. Implications for fibre formation (IV), 37

Concluding remarks and future perspectives, 39

References, 42

Acknowledgements, 50

Appendix

Original articles I-IV

This thesis is based on the following papers, which will be referred to by their Roman numerals:

I. Stark M., Grip S.*, Rising, A.*, Hedhammar M., Engström W., Hjälm G., Johansson J. (2007) Macroscopic fibres self-assembled from recombinant miniature spider silk proteins. *Biomacromolecules*, (In press).
* Equal contribution

II. Rising A., Johansson J., Larson G., Bongcam-Rudloff E., Engström W., Hjälm G. (2007) Major ampullate spidroins from *Euprosthenops australis*: multiplicity at protein, mRNA and gene levels. (Submitted to *Insect Molecular Biology*).

III. Rising A., Hjälm G., Engström W., Johansson J. (2006) N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. *Biomacromolecules, 7*, 3120-3124.

IV. Stark M., Rising A., Grip S., Nordling K., Johansson J., Hedhammar M. (2007) Structural properties of non-repetitive and repetitive parts of major ampullate spidroin 1 from *Euprosthenops australis*. Implications for fibre formation (Manuscript).

Abbreviations

Three and one letter codes for the 20 naturally occurring amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Bases

A	adenine
C	cytosine
G	guanine
T	thymine

Spider species

A. bruennichi	*Argiope bruennichi*
A. diadematus	*Araneus diadematus*
A. gemmoides	*Araneus gemmoides*
A. trifasciata	*Argiope trifasciata*
D. tenebrosus	*Dolomedes tenebrosus*
E. australis	*Euprosthenops australis*
L. geometricus	*Latrodectus geometricus*
L. hesperus	*Latrodectus hesperus*
N. clavata	*Nephila clavata*
N. clavipes	*Nephila clavipes*
N. inaurata madagascariensis	*Nephila inaurata madagascariensis*

Other abbreviations

AcSp	Aciniform spidroin
ADF	*Araneus diadematus* fibroin
AFM	Atomic force microscopy
BHK	Baby hamster kidney
B. mori	*Bombyx mori*
CD	Circular dichroism
Chi	Crossover hotspot instigator
CySp	Cylindriform spidroin
E. coli	*Escherichia coli*
ECP	Egg case protein
ER	Endoplasmic reticulum
Flag	Flagelliform
FTIR	Fourier transform infrared
gpd	grams per denier
HFIP	hexafluoroisopropanol
kb	kilobases
MAC-T	Mammary alveolar cells with large-T antigen
MaSp	Major ampullate spidroin
MaSpL	Major ampullate spidroin like
MiSp	Minor ampullate spidroin
NMR	Nuclear magnetic resonance
PHD	Protein profiling of Heidelberg
RecA	Recombinase A
SCP	Spider coating peptide
SEM	Scanning electron microscopy
Spidroin	Spider silk protein
TEM	Transmission electron microscopy
TuSp	Tubuliform spidroin
w/v	Weight per volume
w/w	Weight per weight

Introduction

Spider silks have been known for long to possess extraordinary mechanical properties. Some spiders can spin seven different types of silks (Candelas & Cintron, 1981). The strongest among these, the dragline silk, is one of the toughest materials known to man (Gosline *et al.*, 1999). Thus, the spider silk is an interesting material to mimic for commercial purposes. Already at the beginning of the 18th century, techniques for manufacturing stockings and gloves from spider cocoon silk were described (Bon, 1710-1712). Bon also addressed the issue of medical implications and suggested that spider silk could be used to stop haemorrhage and for wound healing. Furthermore, old abandoned spider webs seem to have an inherent resistance to microorganisms (Foelix, 1996). Despite being in the focus of biologists and material scientists for centuries we are still not capable of producing fibres with the same toughness as dragline silk. In contrast to man-made high-performance materials, spider silk is produced at ambient temperature and pressure using renewable resources and a benign solvent.

Definition of silk

Silks are defined as ”fibrous proteins containing highly repetitive sequences of amino acids and are stored in the animal as a liquid and configure into fibres when sheared or spun at secretion” (Craig, 1997). Spider silks are composed of proteins that generally show a repetitive core region flanked by non-repetitive N- and C-terminal domains (Hayashi & Lewis, 2001; Motriuk-Smith *et al.*, 2005).

Silks and silk glands in spiders

Silk is produced solely by arthropods, and only by animals in the classes Insecta, Myriapoda and Arachnida (Craig, 1997). The larvae of insects from many groups secrete a great variety of silks, but only one type of silk is produced by a single individual. In contrast, among the order Araneae, that belongs the class Arachnida (Figure 1), up to seven different silks can be spun by individual spiders of certain species. The suborder Araneomorphae includes the great majority of spiders, about 35 000 species in 90 families distributed all over the world, which display their possibilities to adapt to various ecological conditions (Kovoor, 1987). The majority of spider silks studied are spun by members of the Araneomorphae (Gatesy *et al.*, 2001; Gosline, DeMont & Denny, 1986).

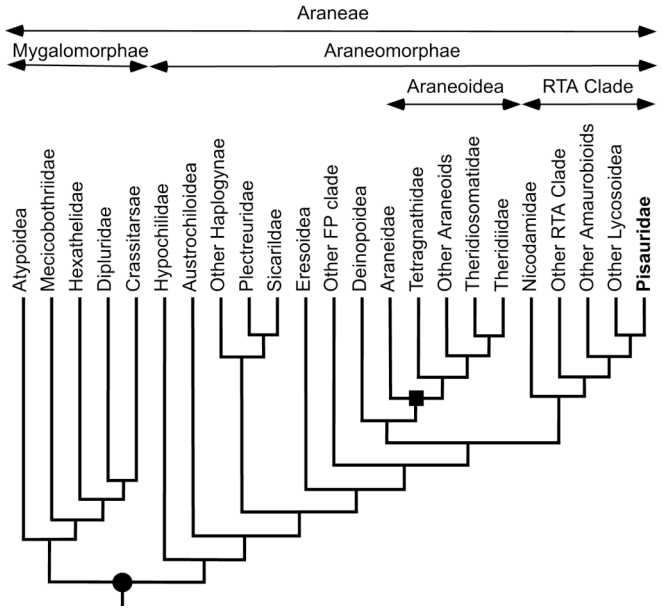

Figure 1. Simplified morphological phylogeny of the Araneae. The *Euprosthenops australis* belongs to the Pisauridae (shown in bold). Modified from Challis, Goodacre & Hewitt (2006). Nodes that are calibrated by fossil evidence (●) 240 million years ago and (■) 125 million years ago (Selden & Gall, 1992; Selden, 1990).

The most prominent and functionally significant silk glands are found in the opisthosoma (abdomen) of the spider (Figure 2). The silk glands are probably derived from epidermal invaginations of the fourth and fifth segments of the opisthosoma which bear the spinnerets. Each gland is connected to a secretory duct that leads to a spool or spigot on the spinneret (Kovoor, 1987). Orb weaving spiders, that belongs to the Araneoidea (Figure 1), can have up to seven different pairs of silk glands (Candelas & Cintron, 1981), each producing a silk with specific purpose and unique mechanical properties. The dragline silk is produced in the major ampullate gland and is used to make the framework of the web and also as a lifeline. The minor ampullate gland is morphologically similar to the major ampullate gland, and synthesises fibres for the web radii (Gosline, DeMont & Denny, 1986). The flagelliform silk, the most elastic fibre, is produced by a gland with a corresponding name and makes up the capture spiral of the web. The capture spiral is coated with a sticky silk produced by the aggregate glands (Kovoor, 1987). The piriform silk is used as a cement to attach fibres to a surface or to connect different threads in the web (Gosline, DeMont & Denny, 1986; Kovoor, 1987). For the production of the egg case the orb weaving spider uses two types of silk. The outer layer is made up by silk from the cylindriform (also called tubuliform) glands (Candelas, Ortiz & Molina, 1986), whereas the inner layer is composed of silk derived from the aciniform gland (Vollrath, 2000). The aciniform silk is also used for wrapping prey. The different spider silks are summarised in Table 1.

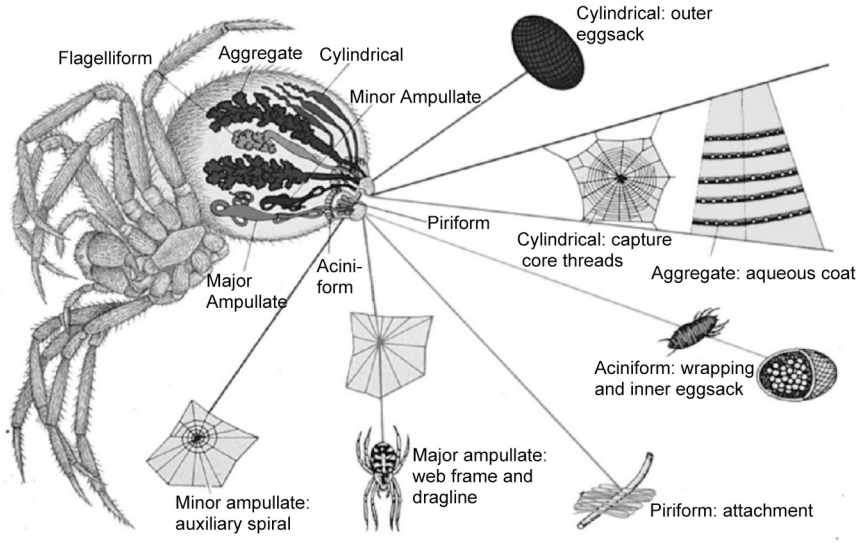

Figure 2. The silk glands and silks produced by the orbweaving spider *Araneus diadematus*. Modified from Vollrath (1992).

Members of the Araneomorphae have one to four pairs of spinnerets and almost all have at least three categories of silk glands; ampullate, aciniform and piriform glands (Kovoor, 1987; Kovoor & Lopez, 1983). The major ampullate glands probably evolved at the divergence of Araneomorphae, approximately 240 million years ago, whereas the flagelliform glands are thought to have originated at a later stage, some 125 million years ago (Figure 1, Challis, Goodacre & Hewitt, 2006).

Table 1. Spider silks

Gland	**Function**	**Core fibre proteins**	**Ensemble repeats**
Major ampullate	Web frame, life line	MaSp1	$(A)_n$ GA and GGX
		MaSp2	$(A)_n$, GGX, GPGXX
Minor ampullate	Web reinforcement and temporary capture silk	MiSp1	GGX, $(GA)_n$, $(A)_n$, spacer
		MiSp2	GGX, $(GA)_n$, $(A)_n$, spacer
Flagelliform	Capture spiral	Flag	$(GPGXX)_n$, GGX, spacer
Aciniform	Wrapping silk, inner egg case	AcSp1	$(S)_n$, GGX
Tubuliform	Outer egg case	TuSp1	$(S)_n$, $(T)_n$, $(SX)_n$
		ECP-1	$(GA)_n$, short $(A)_n$
		ECP-2	$(GA)_n$, short $(A)_n$
Aggregate	Sticky coating for capture spiral	SCP-1	none
		SCP-2	none
Piriform	Attachment disk and joining fibre	Unknown	

Modified from Hu, *et al.*, 2006. Data on aggregate silk from Hu, *et al.,* 2007.

The major ampullate gland

The major ampullate gland consists of three distinct parts: the tail, the sac and the duct (Figure 3). The proteins that constitute the final fibre are produced in the tail and sac (Bell & Peakall, 1969; Plazaola & Candelas, 1991; Sponner *et al.*, 2005a).

The main, and probably sole, function of the tail is to synthesise silk protein (Bell & Peakall, 1969). The epithelium of the tail presents a simple columnar arrangement, consisting of a single type of secretory cell. The cells have large nuclei, well-developed rough endoplasmic reticulum (ER) and a large number of secretory granules, features that are compatible with a high rate synthesis of secretory proteins. The granules are accumulated in the apical region of the cell, where they discharge their content in a merocrine fashion (Plazaola & Candelas, 1991).

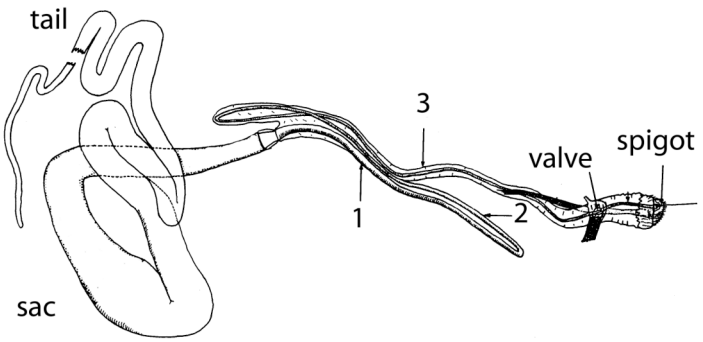

Figure 3. The major ampullate gland is composed of three distinct parts; the tail, the sac and the duct. The three limbs of the duct are indicated (1-3). The valve is located between the third limb of the duct and the spigot. Modified from Vollrath & Knight (1999).

The volume of the sac is larger than that of the tail. In the sac the secreted proteins are stored in a highly concentrated (~30-50%, w/w) aqueous solution called the dope (Chen, Knight & Vollrath, 2002; Hijirida *et al.*, 1996). The overall rate of protein synthesis in the single layered epithelium of the sac is only one quarter of that found in the tail. In addition, the total volume of the epithelial cells of the tail is ten times larger than that of the sac. Consequently, the amount of protein synthesised in the sac is only a few percent of the total amount synthesised by the gland (Bell & Peakall, 1969).

If the sole function would be to connect the sac to the spinneret, the spider would have managed with a five times shorter duct (Bell & Peakall, 1969). It is progressively narrowing and has three limbs folded into an S-shape (Knight & Vollrath, 1999). The aqueous solution of proteins stored in the sac undergoes conversion to a water insoluble fibre in the third limb of the duct (Work, 1977).

The wall of the duct is composed of a single layer of epithelial cells producing a cuticular intima (Vollrath, Knight & Hu, 1998). The epithelium of the distal part of the duct shows morphological features that suggest a role in water and ion transport (*i.e.* microvilli, desmosomes, infolding of the basal membrane) (Bell & Peakall, 1969; Vollrath, Knight & Hu, 1998). The duct ends in a spigot on the anterior spinneret (Kovoor, 1987; Wilson, 1969). A muscle-controlled valve is located at the end of the duct, just before it enters the spinneret. It probably acts as a clamp gripping the dragline silk when the spider suspends itself from a support or arrests a fall without using its legs (Wilson, 1962; Wilson, 1969). An additional function of the valve might be to act as a pump to restart spinning if the dragline is broken internally (Vollrath & Knight, 1999).

The formation of a solid fibre

How the spider manages to keep the silk proteins (spidroins) in a highly concentrated aqueous solution and the mechanisms behind the conversion into a solid fibre is not completely understood. Factors such as a lowered pH, prealignment of the protein molecules, shear forces and changes in ion concentration along the spinning apparatus have been suggested to contribute to the process (see further below).

The dope in the proximal part of the sac contains numerous small droplets, less than 1 μm in diameter, which are distributed throughout the otherwise homogenous silk matrix (Knight & Vollrath, 1999; Vollrath & Knight, 1999). The droplets have been suggested to fuse and stretch in the spinning apparatus to form the elongated canaliculi sometimes observed in the core of the final fibre (Augsten, Muhlig & Herrmann, 2000; Frische, Maunsbach & Vollrath, 1998; Vollrath & Knight, 1999). However, other researchers have been unable to identify such structures in the fibre (Thiel, Kunkel & Viney, 1994). The secondary structure of the proteins in the dope has not been unequivocally determined. ^{13}C nuclear magnetic resonance (NMR) spectroscopy, together with Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopy showed that the proteins are in a state of dynamically averaged helical conformation (Hijirida*, et al.*, 1996). The absence of β-sheets in the dope has also been confirmed by using Congo Red staining (Knight, Knight & Vollrath, 2000). A predominantly random coil conformation has been proposed by using NMR (Hronska *et al.*, 2004; Lawrence, Vierra & Moore, 2004), whereas mainly random coil and α-helical conformation of the proteins in the proximal part of the gland and a β-sheet rich structure in the distal part have been demonstrated by CD (Dicko, Vollrath & Kenney, 2004; Kenney *et al.*, 2002).

In summary, the major ampullate silk proteins in the dope seem difficult to assign specific conformations. A contributing factor to this might be that conformational changes of the proteins in the dope can be induced by shearing and/or dehydration (Chen, Knight & Vollrath, 2002), making the dope susceptible to handling. However, it can be concluded that the spidroins are stored in a less ordered conformation than in the fibre, where mainly β-conformation is seen (Simmons, Michal & Jelinski, 1996; Warwicker, 1960).

In the distal part of the sac and the first two limbs of the duct, the dope appears to be liquid crystalline (Kerkam *et al.*, 1991; Knight & Vollrath, 1999), forming a nematic phase, *i.e.* it forms a substance that flows as a liquid but maintains some of its orientational order characteristic of a crystal with the long axes of neighbouring molecules located approximately parallel to one another. This mechanism is proposed to allow the dope to flow through the sac and duct while the molecules align (Vollrath & Knight, 2001). The elongational flow rate in the duct is likely to increase as the duct narrows (Knight & Vollrath, 1999). The forces generated during this stage probably cause the proteins to align to form the β-sheets seen in the final fibre (Knight, Knight & Vollrath, 2000; Vollrath & Knight, 2001). In the third limb of the duct, the solid fibre is formed as the dope suddenly narrows and pulls away from the walls of the duct (Vollrath, Knight & Hu, 1998; Work, 1977) emitting water (Peakall, 1969; Tillinghast, Chase & Townley, 1984; Vollrath & Knight, 2001). The phase separation might be further facilitated by a drop in pH, caused by proton pumps in the distal part of the duct (Dicko, Vollrath & Kenney, 2004; Knight & Vollrath, 2001; Vollrath, Knight & Hu, 1998) and changes in ion concentration (Knight & Vollrath, 2001). In two different reports the pH has been found to decrease from 7.2 in the tail to 6.3 at the beginning of the duct (Dicko, Vollrath & Kenney, 2004), and from 6.9 in the sac to 6.3 in the third limb of the duct (Knight & Vollrath, 2001). When the silk dope travels down the duct, the most significant changes in ion concentration are observed as a concomitant decrease in Na^+ and Cl^- concentration and an increase in K^+ concentration (Knight & Vollrath, 2001; Tillinghast, Chase & Townley, 1984). The exchange of Na^+ for K^+ could facilitate the conversion of structural water on the proteins to bulk water since K^+ is slightly more chaotropic (Knight & Vollrath, 2001). The uptake of water by the epithelium in the duct may be associated with the reabsorption of Na^+ and Cl^- (Knight & Vollrath, 2001).

In vitro experiments on recombinant proteins and native dope have proposed that lowered pH (Dicko *et al.*, 2004; Vollrath, Knight & Hu, 1998), increased amounts of phosphate ions (Huemmerich *et al.*, 2004a) as well as certain cations, such as K^+ (Chen, Knight & Vollrath, 2002; Dicko*, et al.*, 2004), may induce changes in the proteins secondary structure and/or induce aggregation.

Possibly, the formed fibre is coated before it leaves the spinneret by a layer of glycoproteins (Augsten, Muhlig & Herrmann, 2000; Sponner *et al.*, 2005b; Vollrath & Knight, 1999).

The dragline silk proteins and their encoding genes

The dragline silk is unusual in its amino-acid composition, a majority of the residues have short or no side chain (Ala, Gly) (Casem, Turner & Houchin, 1999; Peakall, 1969). The fibre is composed of at least two similar proteins, most often referred to as major ampullate spidroin (MaSp) 1 and MaSp2 (Hinman & Lewis, 1992). However, since some sequences do not conform to the general description of these proteins, alternative nomenclature exist. Such sequences include *Araneus diadematus* fibroin 3 (ADF-3) and ADF-4, two MaSp2-like proteins that are

believed to constitute the bulk of the dragline fibre in *Araneus diadematus* (Guerette *et al.*, 1996). Another example is the fibroin 1 and 2 of *Dolomedes tenebrosus* that could not be distinguished from minor ampullate spidroins (Gatesy*, et al.*, 2001). The proteins are large, their size is estimated 200-720 kDa, depending on the experimental conditions used (Jackson, 1995; Mello, 1994; Sponner*, et al.*, 2005a).

The size of the encoding mRNAs has been estimated to 7.5-12 kb, as judged by Northern blotting (Guerette*, et al.*, 1996; Hayashi, Shipley & Lewis, 1999). Only partial dragline silk genomic and cDNA sequences are currently available. This is probably due to the large GC-rich transcripts/genes as well as the repetitive nature of the coding sequence. Most of the sequences available encode the C-terminal part of the proteins, since technical cloning issues favour 3´amplification of mRNA. The abundance of the nucleotides G and C results in high amounts of alanine and glycine in the proteins. Both transcripts show a codon usage with a preference for A or T as the third nucleotide, which lowers the total amount of G and C. As a consequence the DNA melting temperature drops and, most likely, the degree of secondary structure of the mRNA is reduced (Hinman & Lewis, 1992; Mita *et al.*, 1988).

MaSp1 and MaSp2 are similar in amino acid composition and probably share a common architecture. They both have non-repetitive N-terminal and C-terminal regions flanking an extensive repetitive part (Hinman & Lewis, 1992; Motriuk-Smith*, et al.*, 2005; Xu & Lewis, 1990). The repetitive part probably comprises hundreds of alanine-blocks (4-12 residues) interspersed with glycine-rich repeats of different length. The main motifs in the sequence are for MaSp1 $(A)_n$, GA and GGX, whereas MaSp2 is dominated by the motifs $(A)_n$ and GPGXX (Gatesy*, et al.*, 2001; Hayashi, Shipley & Lewis, 1999).

All hitherto identified dragline silk proteins share a conserved C-terminal domain which is approximately 100 amino acid residues long (Challis, Goodacre & Hewitt, 2006). This domain contains a mainly hydrophobic stretch of about 20 amino acid residues that is predicted to form an amphipathic α-helix. A Cys residue is located just N-terminally of the hydrophobic stretch and in direct vicinity of an Asp residue. It has been proposed that this helical part is involved in hydrophobic interaction between spidroins, thereby facilitating disulfide bridge formation (Challis, Goodacre & Hewitt, 2006; Sponner *et al.*, 2005c). The role of the C-terminal domain is not determined, though it has been shown to be crucial in fibre formation (Ittah *et al.*, 2006). Apart from being responsible for intermolecular disulfide bridge formation, it has also been proposed to be important in maintaining the aqueous state of silk prior to extrusion and to be responsible for and/or have a role in signalling (Beckwitt & Arcidiacono, 1994; Sponner*, et al.*, 2005c). It might also be required for recruiting accessory proteins such as chaperones, in order to facilitate correct folding (Challis, Goodacre & Hewitt, 2006).

In 2005, Motriuk-Smith *et al.* presented the first N-terminal sequences originating from dragline silk proteins. Several tentative start codons were found

and two possible isoforms derived from different translational start sites were suggested. Signal peptides were not unequivocally identified (Motriuk-Smith, *et al.*, 2005). This notion was recently amended when a common N-terminal domain and signal peptides could be identified in several different spider silks, including MaSp:s (*cf.* III).

Secretory proteins and signal peptides

Since spidroins are secretory proteins, they should enter the secretory pathway. This requires a signal peptide that directs the protein to the ER. Secretory proteins have signal peptides located in the N-terminal region. When the signal peptide emerges from the ribosome, the translational complex is directed to the ER. As the protein is synthesised it is translocated across the ER membrane through a protein pore (translocon). The signal peptide is cleaved off and the mature protein subsequently released into the ER lumen. The structural features and functions of signal peptides are conserved between different eukaryotic organisms (von Heijne, 1988). They consist of three regions; a hydrophilic n-region usually with net positive charge (1-5 residues long), followed by a hydrophobic h-region (7-15 residues), and finally a more polar c-region containing the signal peptidase cleavage site (3-7 residues) (von Heijne, 1990). The high degree of conservation has made it possible to develop software, *e.g*. SignalP (Bendtsen *et al.*, 2004), that predicts whether a particular sequence is likely to function as a signal peptide or not.

The dragline silk

The structure of dragline silk has not yet been established. Several models have been proposed. A skin-core structure was first suggested after examining dragline silk by light microscopy after wetting and stretching of the fibre (Work, 1984). Later the core region was divided into two separate layers, appearing as two concentric regions in cross-sections of the fibre, as judged by results obtained by atomic force microscopy (AFM). The two layers of the core were suggested to have a fibrillar morphology and to be surrounded by a thin and easily fractured skin (Li, McGhie & Tang, 1994). A model consisting of a fibrillar core with a thin outer skin has also emerged from the use of ^{13}C NMR, scanning electron microscopy (SEM), transmission electron microscopy (TEM), and confocal scanning light microscopy (Augsten, Muhlig & Herrmann, 2000; van Beek *et al.*, 2002). Glycoproteins in the skin and inside the fibre was detected by labelling with Concavalin A gold (Augsten, Muhlig & Herrmann, 2000). However, no skin-core composition or fibrillar structure could be observed in other studies where dragline fibres were examined by light microscopy and TEM (Thiel, Guess & Viney, 1997; Thiel, Kunkel & Viney, 1994). By submersing spider dragline silk in urea and examining it by light microscopy, yet another model of the structural composition was presented by Vollrath and co-workers. According to this model there are four layers, one of which is composed of a microfibrillar network (Vollrath *et al.*, 1996). The fibrillar network could not be identified by Frische *et al*. (1998), using TEM. However, a thin outer layer of higher electron density was identified

(Frische, Maunsbach & Vollrath, 1998). In another study, by using labelled antibodies directed towards MaSp1 and MaSp2, respectively, and TEM of cross sectioned dragline silk, a three layered model was proposed. According to this, the MaSp1 and MaSp2 are located in a core region, surrounded by an outer layer of unknown proteins (skin) which in turn is coated with glycoproteins (Sponner*, et al.*, 2005b). These partly contradictory results and the number of different models proposed suggest that the structure of the fibre is difficult to determine. Possibly, the fibre is sensitive to chemicals, temperature, changes in humidity and/or sectioning techniques used.

The distribution of the spidroins in the dragline silk is interesting in the sense that MaSp1 is found almost uniformly within a core region whereas MaSp2 is missing in the periphery and tightly packed in certain core areas. However, in the duct when the dope is in liquid form, the two proteins seem to be evenly distributed. It is not known how the proteins are separated from one another, nor is the functional significance of the protein distribution in the fibre understood (Sponner*, et al.*, 2005b). However, the fact remains that MaSp1 is more abundant in the fibre than MaSp2 (Hinman & Lewis, 1992; Sponner*, et al.*, 2005a).

Molecular studies of dragline silk by X-ray diffraction (Bram *et al.*, 1997; Warwicker, 1960), Raman spectroscopy (Gillespie, Viney & Yager, 1994), Fourier transform infrared (FTIR) spectroscopy (Dong, Lewis & Middaugh, 1991) and NMR (Kummerlen *et al.*, 1996; Lawrence, Vierra & Moore, 2004; Simmons, Ray & Jelinski, 1994) have shown that spider dragline silk is composed of alanine-rich anti-parallel β-sheets. Simmons *et al.* (1996) proposed a model for the molecular structure of spider dragline silk where two crystalline fractions composed of stretches of alanine were embedded in a glycine rich amorphous matrix (Figure 4). From solid state ^{2}H-NMR, the two crystalline fractions were found to be either highly oriented or poorly oriented and less densely packed. About 40% of the alanines were concluded to be present in β-sheets that are highly oriented parallel to the fibre axis. No increase in the amount or orientation of the crystals due to stretching could be observed (Simmons, Michal & Jelinski, 1996).

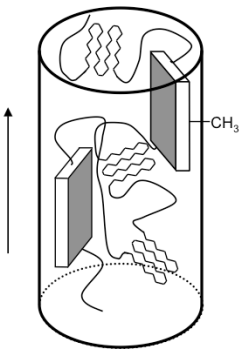

Figure 4. Schematic presentation of the molecular arrangement in the dragline silk. Highly oriented alanine-rich crystals of β-sheets are shown as blocks, jagged lines indicate weakly oriented yet crystalline unaggregated β-sheets in an amorphous glycine-rich matrix (curved lines). The arrow indicates the fibre axis. Adapted from Simmons, Michal & Jelinski (1996).

The description of the glycine-rich matrix as amorphous should not be interpreted as a complete lack of organised structure. The motif GGX has been assigned a three-dimensional structure by experimental data. It probably forms a 3_1-helical structure, as judged by NMR and FTIR (Dong, Lewis & Middaugh, 1991; Kummerlen*, et al.*, 1996). It has been speculated that these 3_1-helices form interhelical hydrogen bonds to reinforce the highly oriented polymer network (Kummerlen*, et al.*, 1996). However there are also reports suggesting that some GGX-motifs are incorporated in the β-sheets (van Beek*, et al.*, 2002). The GPGXX pentapeptide is suggested to form a spiral, similar to the β-turn spiral of elastin (Hayashi, Shipley & Lewis, 1999; Hinman & Lewis, 1992). The presence of turn structures have also been proposed (Michal & Jelinski, 1998).

Dragline silk proteins contain Tyr in the glycine-rich stretches of the repetitive part. These have been hypothesised to form di-tyrosine crosslinks in the fibre (Vollrath & Knight, 1999). In the dope, tyrosine residues are suggested not to be buried but instead accessible to the solvent, which supports the hypothesis that spider silk protein aggregation may occur via Tyr-Tyr interactions (Dicko*, et al.*, 2004). However no di-tyrosines could be detected in hydrolysates of dragline silk (Vollrath & Knight, 1999).

Mechanical properties

The mechanical properties of the dragline silk are thought to be conferred largely to the secondary structure of the amino acid motifs in the repetitive part of the proteins. In the current model of dragline silk, the alanine-crystals link the proteins together and give the fibre its strength. The extensibility of the fibre is probably due to the glycine-rich amorphous matrix, where the GPGXX and GGX motifs possibly form spring-like structures and 3_1 helices respectively (Hayashi, Shipley & Lewis, 1999). The variability in length and properties of the repetitive domains between dragline silks of different species may, at least partly, explain differences in mechanical properties (Gosline*, et al.*, 1999). However, the mechanical characteristics of dragline silk differs substantially, not only between species, but also within species and even within an individual (Madsen, Shao & Vollrath, 1999). The dragline silk is affected by a number of variables, such as silking rate, starvation, reeling or natural spinning, anaesthesia of the spider, temperature and humidity. Other factors that could influence the values obtained include inaccurate determination of fibre diameter, gross defects that cause fibre failure and strain rate (Cunniff *et al.*, 1994; Gosline*, et al.*, 1999; Madsen, Shao & Vollrath, 1999; Madsen & Vollrath, 2000; Vollrath, Madsen & Shao, 2001).

Figure 5 illustrates an example of a stress-strain curve. The stress (σ) is the force (F) per cross sectional area, defined as $\sigma=F/A$, where A is the initial cross sectional area of the fibre. The strain (ε) is the deformation, defined as $\varepsilon=\Delta L/L_0$, where L_0 is the fibres initial length and ΔL is the change in fibre length. The slope of the stress-strain curve gives the stiffness of the material. The strength (σ_{max}) and the extensibility (ε_{max}) are the maximum values of stress and strain at the point where the material fails. The area under the stress-strain curve gives the required energy to break (toughness) of the material. The point where increased stress will

result in inelastic deformation of the material is called the yielding point. The mechanical properties of dragline silk, some other biomaterials and a few high performance man-made materials are listed in Table 2. In comparison, the dragline silk does not display a superior strength to several other materials. However, because of its extensibility, it outperforms all other materials when it comes to toughness. To catch a flying insect, the ability to absorb energy as well as the manner in which it is absorbed is important. The energy could either be stored through elastic deformation or it could be dissipated as heat through friction. The hysteresis, defined as the ratio of energy dissipated to energy absorbed, for dragline silk is approximately 65%, meaning that a majority of the kinetic energy absorbed is transformed into heat and will not be available to throw the prey out of the web through elastic recoil (Gosline*, et al.*, 1999). In contrast, tendons have a hysteresis of approximately 9% reflecting their function as energy stores during locomotion (Pollock & Shadwick, 1994).

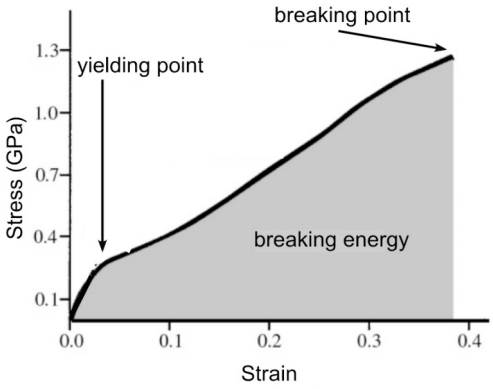

Figure 5. An example of a stress-strain curve of a dragline silk from *Nephila edulis*. The yielding point and breaking point are indicated by arrows. The breaking energy can be calculated from the area under the curve (shown in grey). Modified from Vollrath, Madsen & Shao (2001).

The dragline silk studied in I-IV is spun by *Euprosthenops australis*. The dragline of an *Euprosthenops* of unknown species has been found to be the strongest (1.5 GPa) and among the least extendable (17%) of dragline silks reported in the literature (Table 2; Madsen, Shao & Vollrath, 1999). One exception is a study where the tensile strength of *Araneus gemmoides* and *Nephila clavipes* draglines were reported to exceed 4 GPa (Stauffer, Coguill & Lewis, 1994). However, this value is much higher than 0.8-1.5 GPa that has been found in a range of studies by other researchers (Table 2; Blackledge & Hayashi, 2006). From the stress-strain graph in Madsen *et al.* (1999) the toughness of *Euprosthenops sp.* dragline silk can be calculated to approximately 170 MJ/m^3.

Table 2. Mechanical properties of dragline silk and other materials

Material	Strength σ_{max} (GPa)	Extensibility ε_{max} (%)	Toughness (MJ/m^3)	Ref.
Dragline silk[l]	0.8-1.5	15-39	96-230	a, b, c, d, e, f, g
Bombyx mori silk	0.6	18	70	e
Mammalian tendon	0.12	2.7	6	h, i
Kevlar	3.6	2.7	50	i, j
High tensile steel	1.5	0.8	6	i, j

[l] Only dragline silks from web building spiders are included.
[a] Sirishaisit, *et al.*, 2003
[b] Vollrath, Madsen & Shao, 2001
[c] Madsen, Shao & Vollrath, 1999
[d] Lawrence, Vierra & Moore, 2004
[e] Gosline, *et al.*, 1999
[f] Madsen & Vollrath, 2000
[g] Swanson, *et al.*, 2006
[h] Pollock & Shadwick, 1994
[i] Gosline, *et al.*, 2002
[j] Gordon,1988

Supercontraction

When dragline silk is immersed in water, it shortens its length by approximately 20-50%, the diameter increases and its mechanical properties change (Shao & Vollrath, 1999). This phenomenon is known as supercontraction and results in a reduction of stiffness whereas the extensibility is drastically increased. The silk becomes more rubber-like, possibly due to changes in the amorphous regions (Gosline, Denny & Demont, 1984). The purpose of the supercontraction, if any, and its impact on the web is not known. In the web the fibres are usually attached to rigid structures, and this limits their capacity to shorten significantly (Gosline*, et al.*, 1999). Supercontraction could be a constraint of combining strength and extensibility. In this case it would not fill any specific purpose and hence would not be subject to evolutionary pressure (Liu, Shao & Vollrath, 2005).

Studies utilizing X-ray diffraction (Grubb & Ji, 1999), Raman spectroscopy (Shao & Vollrath, 1999), NMR (Eles & Michal, 2004; Simmons, Michal & Jelinski, 1996), and birefringence (Fornes, Work & Morosoff, 1983) suggest that supercontraction is driven by the reversible disorientation of the molecular chains in the amorphous region. Water molecules break hydrogen bonds in the amorphous region but are unable to penetrate the crystalline areas. The crystals rotate upon supercontraction but are otherwise unaffected by the presence of water (Simmons, Michal & Jelinski, 1996; Work & Morosoff, 1982). Reorientation of the sheet regions when wetted has also been observed in minor ampullate silk, that does not supercontract. Thus it is probably the Gly-rich parts and not the crystalline regions that are responsible for the supercontraction of major ampullate silk (Parkhe *et al.*, 1997).

Recombinant expression of dragline silk proteins

Unlike silk worms, spiders are territorial and therefore difficult to rear. Hence, there is a need for alternative ways of producing spider silk. This has been achieved by recombinant production in a variety of organisms including bacteria, yeast, mammalian and insect cells, tobacco, potato, and even transgenic goats (Fahnestock & Bedzyk, 1997; Huemmerich *et al.*, 2004b; Lazaris *et al.*, 2002; Prince *et al.*, 1995; Scheller *et al.*, 2001; Williams, 2003). Since no full length MaSp gene has been cloned and due to limitations of the production systems available, only sequences encoding partial spidroins have been expressed. Problems with solubility have been a major obstacle because it leads to significant loss of protein during purification and the necessity to use denaturing agents. Other common problems involve low expression levels and truncation (especially of longer proteins) as well as instability of the cDNA inserts in prokaryotes (Arcidiacono *et al.*, 1998; Fahnestock & Bedzyk, 1997; Fahnestock & Irwin, 1997; Prince*, et al.*, 1995). Truncation and low levels of expression could, at least partly, be dependent on the restricted amino acid usage, which puts high demands on abundance of specific tRNAs. The cells producing the dope in spiders have specialised tRNA pools to meet the demand of alanine and glycine during translation, something most expression systems lack (Candelas, 1990). The spidroin mRNAs are likely to have a high degree of secondary structure, possibly causing translational pauses and fall-offs at the ribosome (Fahnestock & Irwin, 1997). The results of expression of dragline silk proteins in different systems are summarised in Table 3.

Production in Escherichia coli

Recombinant spidroins ranging from 12 to 163 kDa have been produced in *E. coli*. Synthetic (designed iterated repetitive modules) as well as partial native sequences have been expressed and purified. In one study, an exceptionally high expression level approaching 300 mg/L, was obtained. However, the quantification method used was not stated and there were considerable size heterogeneity of the proteins produced (Fahnestock & Irwin, 1997). Problems with instability of the inserted genes and truncation of protein synthesis leading to an array of protein species of different length have also been observed by others (Arcidiacono*, et al.*, 1998). These problems seem to increase with increasing size of the inserted gene. Average translation termination rates have been reported to be between 1 in 300 to 1 in 1100 codons, though the shorter protein species could also be the result of transcriptional problems and/or proteolytic activities (Fahnestock & Irwin, 1997). In another study, the appearance of lower molecular weight bands was observed especially after delays in the purification process, why these were presumed to result from proteolytic cleavage (Lewis *et al.*, 1996). Problems with solubility of the expressed proteins have forced the purification procedures to include the use of denaturing agents, such as guanidine hydrochloride or urea (Arcidiacono*, et al.*, 1998; Arcidiacono *et al.*, 2002; Fahnestock & Irwin, 1997). To obtain fibres, artificial spinning procedures have been used, requiring the proteins to be dissolved in urea, hexafluoroisopropanol (HFIP) or formic acid. None of these reports show any data on mechanical properties of the recombinant fibres.

Table 3. Expression of partial dragline silk proteins

Organism	Type	Size (kDa)	Yield (mg/L)	Purity (%)	Comments	Fibers produced	Ref.
E.coli	Synthetic MaSp1 repeats *N. clavipes*[#]	15-41	2-15	70	Poor solubility. Precipitation during purification.	No	(1)
”	MaSp1 C-terminal part *N. clavipes* and	43	4	50-95	Denaturing agents used in purification	Yes. Microspinner in coagulation bath	(2)
	MaSp1/2 repeats hybrid	55	ns	42-90	”	”	
”	Repetitive synthetic or C-terminal domains/combined synthetic and C-terminal domains from ADF-3 and 4	12-60	10-30 (140-360, fermentor)	ns	Denaturing agents used in purification	No	(3)
”	Synthetic MaSp1 and MaSp2 repeats *N. clavipes*	65-163	ns (300mg/L expressed)	>90	Denaturing agents used in purification or to dissolve precipitates	Yes. Details not shown	(4)
”	Repetitive part MaSp1 *N. clavipes*	50	2.5	ns	Denaturing agents used in purification	Yes. Electrospun from HFIP solution	(5)
”	Synthetic MaSp 2 repeats	58	1-10 mg/g cells*	>99	Dissolved in formic acid for spinning	Yes. Forced through needle into coagulation bath	(6)
P. pastoris	Synthetic MaSp1 repeats *N. clavipes*	<3000 codons[‡]	ns (<1000mg/L expressed)	>95	Significant loss of protein due to precipitation. Denaturing agents used in purification	Possibly by O’Brien *et al*. 1998. Details not shown.	(7)
BHK and MAC-T cells	ADF-3 C-terminal part from *A. diadematus*	60 (60-140)	ns (25-50 mg/L expressed)	70-90	Precipitation/denaturation during puification. Urea used in spin dope preparation	Yes. Coagulation bath	(8)
	MaSp1 C-terminal part *N. clavipes*	59-101	ns	ns		No	
	MaSp2 C-terminal part *N. clavipes*	59	ns	ns		No	
Sf9 cells	ADF-4 C-terminal part	54	ns	ns	Insoluble	Spontaneously forms intracellular fibres	(9)
	ADF-3 C-terminal part	56	ns	ns	Soluble	No	
Tobacco and potato	Synthetic MaSp1 repeats *N. clavipes*	<98	ns (2% of tsp[¥] expressed)		Soluble. Purification procedure involved precipitation, proteins dissolved in guanidine HCl	No	(10)
Tobacco	MaSp1 and MaSp2 C-terminal part *N. clavipes*	~60	ns (≤0.01% of tsp[¥] expressed)		Soluble. No purification procedure reported.	No	(11)

#: Additional constructs were expressed but not further characterised; ‡: only the number of codons is given; ns: not specified; *: yield given as mg/g cells; ¥: expression given as % of total soluble protein (tsp)

(1) Prince *et al.* 1995; (2) Arcadiacono *et al.* 1998, Arcadiacono *et al.* 2002; (3) Huemmerich *et al.* 2004a; (4) Fahnestock & Irwin 1997, Obrien *et al.* 1998; (5) Bini, *et al.* 2006; (6) Lewis *et al.* 1996; (7) Fahnestock & Bedzyk 1997; (8) Lazaris *et al.* 2002; (9) Huemmerich *et al.* 2004b; (10) Scheller *et al.* 2001; (11) Menassa, *et al.* 2004.

Production in yeast

Expression of synthetic multimers of repeats from MaSp1 and 2 of *N. clavipes*, up to 3000 codons long have been expressed in *Pichia pastoris*. The proteins could be synthesised without truncation and the genes inserted were in general stable through 100 doublings under no selective pressure. Genes longer than 1600 codons were translated less efficiently, though shorter proteins could be produced at high expression levels (1g/L). Solubility of the expressed proteins was an issue in the sense that decreased solubility with time caused the need for resolubilisation procedures using guanidine HCl. No information regarding the fibre forming ability was reported (Fahnestock & Bedzyk, 1997).

Production in cell culture

Partial MaSp1 and 2 from *N. clavipes* and ADF-3 from *A. diadematus* have been expressed in immortalised mammary alveolar cells with large-T antigen (MAC-T) and baby hamster kidney (BHK) cells. One of these, ADF-3, yielded a 60kDa recombinant protein of 70-90% purity when expressed in MAC-T and BHK cells. The level of expression was 25-50 mg/L in a hollow fiber reactor. The recombinant protein was recovered from the culture media by precipitation and could be redissolved in aqueous buffer. Fibres derived from recombinant ADF-3 protein were spun from dopes containing urea at protein concentrations of 10-28% (w/v) using a coagulation bath of methanol and water. By subjecting the fibres to post spinning draw in the coagulation bath (and also in some cases in water), the mechanical properties were enhanced. There was great variability among the fibres tested. The average tenacity was 1.65 grams per denier (gpd) and the toughness 0.645 gpd compared to 7-11 gpd and 0.6-1.3 gpd respectively for native *Araneus* dragline silk. The extensibility was approximately twice the value of native silk (Lazaris*, et al.*, 2002).

C-terminal parts of about 60 kDa of ADF-3 and 4 have been expressed in insect cells (Sf9 cells). ADF-3 was expressed as a soluble protein while ADF-4 aggregated intracellularly. Filaments, with diameters of 200 nm to 1 μm and lengths of up to 100 μm, constituted of ADF-4 were found in the cytosol. These fibres were too short to be subjected to mechanical testing (Huemmerich*, et al.*, 2004b).

Production in plants

Parts of spider dragline silk proteins, with sizes up to 100 kDa, have been expressed in tobacco and potato leaves (Menassa *et al.*, 2004; Scheller*, et al.*, 2001). The proteins were found in the soluble fraction of plant extracts. The purification procedure involved precipitation and the proteins were redissolved in buffer containing guanidine HCl (Scheller*, et al.*, 2001). No formation of fibres from these proteins has been reported (Menassa*, et al.*, 2004; Scheller*, et al.*, 2001).

Evolutionary aspects

Construction of phylogenetic trees is a common way of estimating the evolutionary past. Such trees can be based on comparison of morphology, protein or DNA sequences (Baldauf, 2003).

Homologues (*i.e.* genes that are similar due to a common ancestry) can be either orthologues or paralogues. Orthologues originate from a single ancestral gene in the last common ancestor of the compared genomes. In contrast, paralogues have arisen by gene duplication (Baldauf, 2003; Koonin, 2005). Different spider silk genes have most likely arisen by gene duplication and subsequent diversification, meaning that they are paralogous copies. As such they are expected to cluster in phylogenetic trees according to specific silk type, rather than by species. This is the case if gene duplication occurs before speciation. However, if the genes are paralogous and subjected to concerted evolution (homogenization through unequal crossing-over and gene conversion, Figure 6), the genes could be more similar within a species than between species. In this case, the genes would cluster according to species.

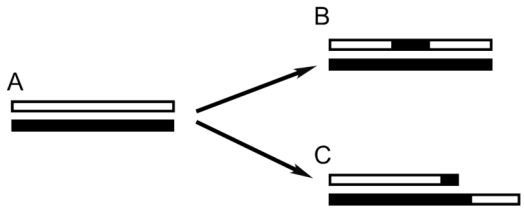

Figure 6. Comparison of gene conversion and unequal crossing-over. (A) Two DNA molecules. (B) Gene conversion; a segment from the black DNA molecule is replacing a part of the white DNA molecule. The black DNA remains unaltered. (C) Unequal crossing-over; the two DNA molecules switch segments of unequal size. Both are altered.

Evolutionary studies of dragline silk have been hampered by the lack of complete sequences of the MaSp genes. Most studies have concentrated on the non-repetitive C-terminal region as the repetitive region is incompletely sequenced and difficult to align. Hence, there is limited information regarding the origin and evolution of dragline silk.

In recent phylogenetic tree analysis including the C-terminal domain of major ampullate, minor ampullate, tubuliform, aciniform and flagelliform spidroins of orb weaving spiders, the spidroins cluster according to silk type and not to species (Challis, Goodacre & Hewitt, 2006; Garb *et al.*, 2006). As previously mentioned, such pattern would be expected with paralogous genes, where gene duplication has occurred prior to the divergence of the species and recombination events have been rare.

The repetitive region, however, shows a different pattern. If the repetitive parts of the genes evolved primarily by point mutations and localised insertion and

deletion events, it is predicted that positionally corresponding repeats would be more alike than repeats along one and the same gene. In an analysis of repeats from *N. clavipes* MaSp1 alleles, the last four repeats before the C-terminal domain were the only ones to group by position, thus instead suggesting concerted evolution (Hayashi, 2002). This has also been proposed by others studying both MaSp1- and MaSp2-analogue sequences (Beckwitt, Arcidiacono & Stote, 1998). In repetitive genes, *e.g.* spidroin genes, an advantageous repeat probably spreads along the gene by mechanisms like gene conversion and unequal crossing-over (Beckwitt, Arcidiacono & Stote, 1998; Hayashi & Lewis, 2000; Liao *et al.*, 1997). This results in that repetitive parts of the silk genes are more similar within one gene than between homologous genes from different species. In the silkworm silk genes, crossover hotspot instigator (Chi) like sequences are thought to create recombination hotspots that destabilise the genes, which suggests similar mechanisms might be acting in spider silk genes (Mita, Ichimura & James, 1994). The Chi sequence is an octamer (5´GCTGGTGG-3´) that was originally identified as a recognition site for RecA-mediated recombination in *E. coli* (Lao & Forsdyke, 2000). Since the corresponding amino acid sequence could be Ala-Gly-Gly, the Chi sequence and similar sequences are likely to occur in silk genes. Chi-like sequences serving as recombination signals have been identified in a number of genomes, *e.g.* human and yeast (Jeffreys, Wilson & Thein, 1985; Tracy, Baumohl & Kowalczykowski, 1997).

In contrast to the dragline silk genes, the flagelliform silk gene has been sequenced and characterised in its full length (Hayashi & Lewis, 2000). It spans over about 30 kb and is composed of 13 exons. The coding sequence of about 15.5 kb follows the typical pattern with non-repetitive N- and C-terminal regions and an extensive repetitive central part consisting of ensemble repeats. The sequence analysis of this gene revealed that on average the repeated exons are more alike within one species than between two different *Nephila* species. Most of the divergence observed between exons consists of short insertions or deletions of entire amino acid motifs. Both differences in length and the homogenisation seen within the gene was suggested to be caused by a combination of gene conversion and unequal crossing-over at repetitive exons (Hayashi & Lewis, 2000). However, Higgins and co-workers (2007) came to a different conclusion when they analysed multiple alleles of one intron-exon segment of the flagelliform gene from four different populations of *N. clavipes*. No evidence of gene conversion was found. Instead the alleles displayed high levels of nucleotide and amino acid substitutions among repeating motifs (Higgins *et al.*, 2007).

The N-terminal part of spider silk proteins was first described for flagelliform silk (Hayashi & Lewis, 1998). Since then, N-terminal regions from another flagelliform spidroin, four cylindriform spidroins and three MaSp:s have been reported (Hayashi & Lewis, 2000; Hu *et al.*, 2006; Motriuk-Smith*, et al.*, 2005; Zhao *et al.*, 2005; Zhao *et al.*, 2006). Several of these sequences were analysed for phylogenetic relationships in III.

Bombyx mori silk

The silkworm *B. mori* spins a cocoon silk with favourable mechanical properties (Table 2). The silk has been used commercially in the textile industry for centuries and, more recently, also for biomedical applications (Altman *et al.*, 2003). The fibre consists of two fibroin monofilaments surrounded by a coat of hydrophilic proteins (sericin). The sericin serves as glue sticking the fibres together in the cocoon. The monofilaments are constructed from three proteins; the heavy- and light chain fibroin (H, 350 kDa and L, 26 kDa, respectively) and the glycoprotein P25 (30kDa). The heavy and light chains are held together by a disulfide bond, while the P25 associates to these by non-covalent interactions (Inoue *et al.*, 2000).

The silk glands are paired and correspond to the salivary glands of other insect larvae. Each gland consists of three parts: a thin posterior part, a wider middle part and an anterior part. The P25 and the heavy and light chain fibroins are produced in the posterior part and may be secreted as a complex ($H_6L_6P25_1$) (Asakura *et al.*, 2007; Inoue*, et al.*, 2000). In the middle part of the gland the proteins can be stored in concentrated (30% w/v) form as a weak gel (Jin & Kaplan, 2003). The sericin is produced in liquid form, mainly in the anterior part of the gland (Matsunami *et al.*, 1998). The glands are connected to separate convergent ducts that fuse to a single duct. A monofilament is formed in the respective duct, before they fuse. As in spiders, the formation of the fibre filaments is proposed to be caused by decreased pH, ion exchange and shear forces (Ochi *et al.*, 2002; Terry *et al.*, 2004; Zhou *et al.*, 2005). The seracin remains in liquid form surrounding the solidified filaments. Shortly after the point of fusion of the ducts there is a structure called the silk press. It is constructed of two muscle-controlled stiff plates that can control the fibre diameter and probably also increase shear forces. The fibre is finally extruded through a spigot (pore) on the lower lip of the mouth (Asakura*, et al.*, 2007). In comparison, spiders have innervated muscular spinnerets that function in association with a spigot, allowing greater flexibility in the fibre production (Craig, 1997). The *B. mori* silk fibre is actively pulled from the spigot by side-to-side movement of the head, producing a figure eight pattern (Magoshi, Magoshi & Nakamura, 1994).

The heavy chain fibroin gene consists of two exons (67 and 15750 bp, respectively) and one intron (971 bp). The deduced amino acid sequence is composed of a highly repetitive core flanked by non-repetitive N- and C-terminal domains (~150 and 50 residues respectively). The N-terminal domain harbours a signal peptide and may adopt a globular structure (Wang *et al.*, 2006; Zhou *et al.*, 2001). Although similar in size, the *B. mori* and spider silk N-terminal domains differ both in amino acid sequence and in predicted secondary structure (unpublished observation). The core region is composed of 12 repetitive and 11 non-repetitive regions that alternate. The repetitive regions are hydrophobic and 159-607 amino acid residues long, while the non-repetitive regions are hydrophilic, shorter (~30 residues) and seem more conserved. Thus, the hydrophobic regions are much larger than the ones in dragline silk. The most common motif of the repetitive region is the GX dipeptide, where X is Ala in 65%, Ser in 23% and Tyr in 10% of the repeats (Bini, Knight & Kaplan, 2004; Ha *et al.*,

2005; Zhou *et al.*, 2000). The C-terminal domain is shorter and the amino acid composition is different compared to spider silk C-terminal domains (Bini, Knight & Kaplan, 2004). The *B. mori* C-terminal domain contains three Cys residues, two of which are engaged in an intra-molecular disulphide bond (Tanaka *et al.*, 1999).

In addition to the *B. mori*, numerous insects produce silk. One example is the larva of the honey bee (*Apis mellifera*). In contrast to *B. mori* the honey bee possesses a silk gene cluster that has probably arisen from gene duplication (Sutherland *et al.*, 2006). Although the two silks have low sequence identity, this implies that the genomic organisation of silk genes can vary substantially between species.

In spite of numerous studies, the secondary structure of the proteins in the silk gland has not been determined, primarily due to difficulties in handling the liquid silk without affecting its structure. Presence of type II β-turns and random coil have been proposed (Asakura *et al.*, 2001a; Asakura *et al.*, 2001b; Ha*, et al.*, 2005). Possibly, the liquid silk forms micelles within the gland, exposing the more hydrophilic terminal regions to the surrounding hydrophilic sericin (Jin & Kaplan, 2003) .

The *B. mori* silk fibres have been reported to consist of bundles of nanofibrils oriented in parallel to the fibre axis (Hakimi *et al.*, 2006). The dominant secondary structure is anti-parallel β-sheet mainly composed of $(GAGAGS)_n$-motifs. Similar to the organisation of spider dragline silk, the β-sheets stack to form crystals surrounded by more flexible structures. In both silks the β-sheets are oriented parallel to the fibre's axis (Rousseau *et al.*, 2004). The *B. mori* silk is generally more crystalline and stiffer than the spider dragline silk. The inter-sheet packing distance is different in the two silks, due to differences in the amino acid sequence (Craig, 1997).

The sericin coat of the *B. mori* silk constitutes a major problem in biomedical applications, since it is highly immunogenic. The coat can be removed (degummed), *e.g* by boiling the silk in water. However, biocompatibility problems in the form of allergic reactions and foreign body response have been reported, also for degummed silk. Most likely, residual sericin is the cause of the allergic reactions observed (Altman*, et al.*, 2003).

Amyloid proteins and fibrils

Amyloid is formed *in vivo* when normally soluble proteins are converted from their native state into insoluble fibrils and plaques. The formation of amyloid is linked to approximately 20 known diseases (*e.g.* Alzheimer's disease, Parkinson's disease, and spongiform encephalopathies). These diseases can be sporadic, inherited or infectious. The proteins that are involved in these diseases have different native configurations (Dobson, 1999; Westermark, 2005). However, the fibrils found in the diseased states are very similar in appearance (Sunde & Blake, 1997). The fibrils are composed of protofilaments (stacks of β-sheets) that may

twist around each other to form a hollow tube (Jimenez *et al.*, 1999; Makin & Serpell, 2005; Makin, Sikorski & Serpell, 2006). The formation of fibrils includes the existence of intermediates causing a lag phase before the rapid development of fibrils occurs. This is known as a nucleation dependent process, *i.e.* a small number of aggregates or seeds are formed, from which larger molecular assemblies grow (Dobson, 1999; Dobson, 2001). The conversion of the dope into a solid fibre in spiders has been suggested to be dependent on a similar process (Kenney, *et al.*, 2002). Interestingly, *B. mori* silk acts as a seeding agent in the development of amyloid protein A (AA) amyloidosis in mice (Lundmark *et al.*, 2005). It has been proposed that all proteins can form amyloid fibrils, though only a handful do so under physiological conditions (Dobson, 1999; Dobson, 2001).

Like amyloids, spider silks form under physiological conditions in a process involving the transition from a soluble to an insoluble state (Vollrath & Knight, 2001). Both types of fibres are composed of protein polymers in β-sheet conformation and factors like low pH or high temperature can induce the conformational change into β-sheets *in vitro* (Dicko, Kenney & Vollrath, 2006). However, there are also some substantial differences. Amyloid fibrils are several magnitudes smaller in size and have a different molecular structure compared to spider silk (Dobson, 2001). In contrast to spider silks, the β-sheets are oriented perpendicular to the fibre axis, generating a cross-β structure (Sunde & Blake, 1997). However, nanofibres formed from a recombinant repetitive part of ADF-4 are of similar size as amyloid fibrils and also contain cross-β structures (Slotta *et al.*, 2007). Furthermore, in contrast to spidroins, when natively folded, proteins forming amyloid have specific functions unrelated to fibre formation (Dobson, 1999).

Present investigation

Aims and scope of thesis

The intention of this thesis was to provide a method that allows large scale production of recombinant spider dragline silk. To achieve this, further understanding of the silk genes and proteins as well as of the polymerisation process was needed.

Our efforts resulted in production of a recombinant silk that holds potential for a number of medical and technical applications. Furthermore, the molecular properties of *E. australis* MaSp proteins and their encoding genes were studied, which revealed the presence of a third MaSp-like gene as well as a potential MaSp1 gene cluster. The characterisation of the N-terminal part of *E. australis* MaSp1 and comparison to published spider silk sequences enabled the identification of a common translational start site, signal peptides and a highly conserved non-repetitive domain consisting of approximately 130 residues. Finally, the secondary structure, stability, and conformational transitions of recombinant N- and C-terminal domains, repetitive regions, and combinations thereof from *E. australis* MaSp1 were studied. Factors previously suggested to influence the fibre formation *in vivo*, like pH and salt concentrations, were examined.

Materials and experimental procedures

Materials and methods used in this thesis that will benefit from some clarification are described in this section. Other methods are presented in the respective papers.

Euprosthenops australis

For a number of reasons, the choice of spider species to study fell on the nursery web spider *E. australis*. Firstly, the dragline silk of *Euprosthenops* sp. is one of the strongest and least extendable (1.5 GPa and 17% respectively) of all dragline silks examined (Blackledge & Hayashi, 2006; Madsen, Shao & Vollrath, 1999). The yielding point is also higher than for most other silks, reaching approximately 1.2 GPa (Madsen, Shao & Vollrath, 1999), meaning that it remains elastic even at high stress. Secondly, it has been found to supercontract less than 20% in aqueous environment (Shao & Vollrath, 1999). These traits are desirable when designing a new biomaterial with supreme mechanical properties. Thirdly, the spiders' relatively large size enabled efficient dissection of the glands.

The *E. australis* lives on the African continent, in bush landscape, and belongs to the fishing spiders (Pisauridae, Figure 1). Despite their name they are found in quite dry habitats and do not hunt prey by fishing. The *E. australis* builds a funnel web and a nursery for its spiderlings, hence it is also called nursery web spider. It is relatively large, the body measures some 3 cm in length. The prosoma (cephalothorax) is covered with white fine hairs, the opisthosoma is yellow and the

legs are striped. The specimens collected for this study were caught in November 2003 in the bush of north-eastern South Africa. Their identity as *E. australis* (Simon, 1898) was confirmed by Astri and John LeRoy of the Spider Club of South Africa. The webs are approximately 2-3 m long and 1 m at the widest part, forming a triangular sheet-like structure. The broadest part of the sheet is attached to a bush or similar about 2 m off the ground. It narrows progressively towards the ground, leading into a funnel that quite often occupies an old animal burrow. The webs are long term, lasting for 2-3 months. The diameter of the dragline silk of *Euprosthenops* sp is approximately 2 μm (Madsen, Shao & Vollrath, 1999). The spiders collected (all females) were found sitting in the funnel waiting for prey to get caught in the web. They all carried an egg case underneath the opisthosoma or had already made a nursery web where the egg case had been placed.

Funnel-web spiders can run quickly over their silken sheet (up to 45 cm/s), whereas insects only can walk clumsily on it. The spider manages to run fast by running on tip toe, putting the tarsi at a quite steep angle to the sheet web. Feathery hairs push against the more densely woven fibres on the sheet web and prevent the feet from sinking in. In contrast, most insects put their legs in a flat manner on the sheet web, which causes considerable adhesion between the leg cuticle and the silk threads. Consequently their movements are severely hampered (Foelix, 1996).

Circular dichroism (CD) spectroscopy

CD spectroscopy is used to determine the secondary structures of peptides and proteins. The technique is based on that chiral molecules interact differently with left- and right-circularly polarised light (Adler, Greenfield & Fasman, 1973; Woody, 1995). In a protein or peptide, the most abundant absorbing species is the peptide bond. Using left- and right-circularly polarised light, CD instruments measure the difference in a sample's absorbance from which the molar ellipticity (θ) can be calculated. Different secondary structures will generate characteristic curves when the residual molar ellipticity is plotted against the wavelength (λ) in the far-UV (~190-250 nm) region. Random coil, *i.e.* disordered, conformation generates a spectrum with minimum around 197 and a maximum at 220 nm. A spectrum with a maximum at 192 nm and a double minimum at 208 and 222 nm is characteristic for α-helices, while β-sheets will give a spectrum with a maximum around 195 nm and a minimum at 217 nm (Figure 7) (Greenfield & Fasman, 1969). The molar ellipticity at 208 or 222 nm is commonly used to estimate the helix content. Furthermore, if the spectrum shows a mixture of different secondary structures, the relative amounts of these can be estimated by curve fitting algorithms.

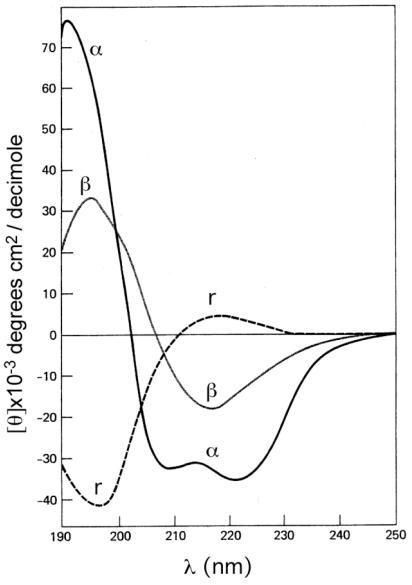

Figure 7. Circular dichroism spectra representative for α- helical (α), β-sheet (β) and random coil (r) conformation.

Phylogenetic analysis

The field of phylogenetics aims to reconstruct the evolutionary past and, in the case of molecular phylogeny, is based on the comparison of DNA or protein sequences. The results from such studies are often presented as trees with branches and nodes. Molecular phylogenetic trees are often drawn with proportional branch lengths, meaning that the length of the branch is proportional to the degree of evolution (roughly, percent sequence difference). The root is the oldest part of the tree and it determines the order of branching. To root a tree it is necessary to introduce an outgroup as an external point of reference. An outgroup is known to have diverged from the group of interest before any members of the ingroup diverged from each other. In the absence of an outgroup the tree is usually left unrooted (Baldauf, 2003).

To construct a tree the sequences of interest must first be aligned by multiple sequence alignment. Alignment programs have separate penalties for introducing gaps (expensive) and extending gaps (cheaper). The resulting alignment can look very different depending on the nature of the sequences and the penalties that are used. Therefore it is generally accepted to adjust alignments by eye to minimise insertion/deletion events (Baldauf, 2003).

There are two main ways to calculate phylogenetic distances. One is to use distance-matrix methods, which is also referred to as clustering or algorithmic methods (*e.g.* neighbour-joining, Fitch-Margoliash). The other is to use discrete data methods, also known as tree-searching methods (*e.g.* parsimony, maximum likelihood, Bayesian methods). The distance-matrix methods calculate the difference (distance) in pairwise combinations for all sequences and assemble them into a tree. Discrete data methods examine each column of the alignment

separately and construct the tree that best fits the data given the model of evolution. In most cases, both methods will result in similar groupings. The discrete data methods are more information-rich and time-consuming than the distance-matrix methods (Baldauf, 2003). The discrete data methods are currently the best to reconstruct relationships between sequences that have been separated for a long time, or are evolving rapidly (Holder & Lewis, 2003).

Bootstrapping and posterior probability are methods used to test the accuracy of the tree, *i.e.* how robustly the dataset supports the tree. Bootstrapping is done by random resampling of the dataset, building trees from each of these cases and calculating the frequency with which the various nodes of the tree are reproduced. Similarly, posterior probablility is an estimate of the robustness of any node of a tree given the tree searching algorithm. Values of 70% or higher are likely to indicate reliable groupings (Baldauf, 2003).

Secondary structure predictions

A protein's function is critically dependent on its tertiary structure. Experimental determination of tertiary structures is much slower than the accumulation of protein sequence data. Even though information needed to fold a protein correctly is present in its amino acid sequence (Anfinsen, 1973), we are not yet able to predict the tertiary structure from amino acid sequence alone. However, the secondary structure of proteins is possible to predict with an overall accuracy of about 70% (Simossis & Heringa, 2004).

The secondary structure prediction programs available can generally be used to assign a protein sequence to one of three different states; α-helix, β-strand or random coil. The programs initially used statistical likelihoods based on that different amino acid residues are more or less prone to adopt these conformations. To improve the accuracy of the prediction, most programs available today use multiple alignments to find homologous proteins of known secondary structure. The underlying assumption is that secondary structure has been more conserved during evolution than amino acid sequence. One example of secondary structure prediction programs is PHD (Rost & Sander, 1993; Rost & Sander, 1994). The system is composed of three levels, two neural networks to predict the secondary structure and one to make a jury decision. The first level, which uses multiple alignments of sequences of high similarity, will predict the secondary structure of each residue in the sequence (*i.e.* it will not consider the secondary structure of surrounding residues). The second level is a structure-structure network that will clean the first output from invalid results (such as *e.g.* α-helices constituted of two residues). The third level, the jury decision, makes the final secondary structure prediction by assembling the predictions from independently trained networks (Kallberg, 2002; Rost, 1996). The PHD system has reached a mean accuracy of prediction of 72% (Rost & Sander, 1994).

Results and discussion

Macroscopic fibres self-assembled from recombinant miniature spider silk proteins (I)

Spider silk is one of the toughest known materials, also exhibiting desirable properties like biocompatibility and biodegradability (Gosline*, et al.*, 1999; Vollrath *et al.*, 2002). Because of these extraordinary properties major efforts and resources have been invested worldwide to use spidroins for design of biomaterials. These attempts have so far failed, mainly because spidroins have unique amino acid sequences that make them very prone to aggregate, which in turn complicates recombinant production of soluble protein and controlled polymerisation. Furthermore, native spidroins are very large which is a complicating factor in existing recombinant expression systems.

A cDNA library made from *E. australis* major ampullate glands, was screened for clones encoding MaSp1. One clone, clone 2 with an insert of approximately 3.8 kb, was chosen for further analysis. To obtain the full length sequence of the cDNA from this clone, nested deletions were made since it was impossible to design specific primers in the repetitive region. By sequencing more than 25 such deletion clones, the complete sequence of clone 2 could be assembled by hand. The deduced amino acid sequence contained a non-repetitive C-terminal domain (~100 amino acid residues long) and an extensive repetitive part (~1100 residues). The repetitive part was extremely uniform, displaying the presence 34 Ala blocks (12-15 residues), interspersed with three different subtypes of Gly-rich repeats (14, 18 or 23 residues).

The repetitive part of MaSp1 from *E. australis* was analysed for frequently occurring motifs. One such motif (four poly-Ala/Gly-rich repeats with a central turn) together with the C-terminal domain was selected for expression. To enhance the solubility and enable efficient purification, this miniature spidroin was expressed as a fusion protein comprised of thioredoxin/His-tag/S-tag/thrombin-cleavage-site/miniature-spidroin.

The fusion protein remained stable for weeks and could be readily purified by immobilised metal ion affinity chromatography. Upon enzymatic release of the fusion tag macroscopic fibres were spontaneously formed, provided that an air-water interface was present and that the protein solution was gently wagged from side to side. In this manner metre-long fibres could be produced under physiological conditions. The structure of these fibres resembled native dragline silk as judged by CD and X-ray diffraction. By using SEM the fibres were found to be composed of tightly aligned fibrils. The mechanical properties were similar to those of regenerated spider silks, corresponding to approximately 10-20% of the tensile strength of native dragline silk. Moreover, the fibres could easily be sterilised by autoclaving and was also shown to support cell adherence with no sign of cytotoxicity. These are imperative traits for a material destined for biomedical applications.

A species hybrid construct composed of a corresponding *Euprosthenops* repetitive part and a *Nephila* C-terminal domain could also form fibres in the same manner, indicating the robustness of this motif. Shortening of the repetitive segment resulted in a dramatic decline of the fibre forming ability. The crucial role of the C-terminal domain in fibre formation was shown as only constructs containing this part formed fibres.

From the results of this study, it can be concluded that the information needed to dictate fibre formation is present in the amino acid sequence of spidroins. Moreover, a miniature spidroin, consisting of four poly-Ala/Gly-rich repeats and the non-repetitive C-terminal domain is sufficient to form macroscopic fibres without the use of spinning.

Major ampullate spidroins from Euprosthenops australis: multiplicity at protein, mRNA and gene levels (II)

In an effort to characterise the MaSp encoding sequences, cDNA libraries were constructed from major ampullate glands of approximately 100 *E. australis* spiders. From these libraries a number of clones were isolated; 14 MaSp1 clones (considered to originate from gene MaSp1a, *cf.* below), two clones encoding the previously uncharacterised *E. australis* MaSp2, as well as a clone encoding a third hitherto unknown spidroin (named MaSpL for MaSp-like). Genomic PCR revealed the presence of two MaSp1 genes (MaSp1a and MaSp1b).

The C-terminal domain of MaSp1, MaSp2 and MaSpL were very similar; 79-89% identity in pairwise comparisons of the amino acid sequences. The most distinctive features were found in the repetitive regions. Apart from the expected differences between MaSp1 and MaSp2 (*i.e.* the presence of GPGXX and QQ motifs in MaSp2), the most striking difference was the exclusive presence of Phe instead of Tyr in MaSp2. This disparity between MaSp sequences has not been previously reported for any other spidroin. The functional significance remains unclear. However, the tyrosyl side chain is able to form hydrogen bonds, a feature not shared by the phenylalanyl side chain. Consequently the inter- and/or intramolecular forces generated by *E. australis* MaSp1 and MaSp2 are likely different. Possibly, this could result in features like reduced tendency to supercontract in aqueous solvents. MaSpL conforms well to repeats characteristic for MaSp1, however large differences in the Gly-rich repeats between MaSpL and MaSp1a suggest that it represents a third member of the MaSp-family. In MaSpL, Tyr and Phe residues alternate at the positions corresponding to the exclusive Tyr residues of MaSp1 and Phe residues of MaSp2. Furthermore, the acidic residue Glu is present in several repeats in MaSpL but is missing in the two other spidroins.

A group of 14 cDNA clones coding for the C-terminal part of MaSp1a showed an unexpectedly high level of heterogeneity in their repetitive parts, while the C-terminal domain was identical in all clones. Apparently, large segments of DNA had been repositioned, deleted or inserted. The cause of the observed pattern remains unknown, but differences in genetic sequences between individuals or the

presence of several MaSp1 genes within one individual were found to be the most likely explanations. Unequal crossing-over and gene conversion may account for such differences in homologous genes within a species. In support of this hypothesis, unequal crossing-over, possibly at Chi-like sequences, and duplication of repetitive segments are events that have been suggested to occur in the *B. mori* silk fibroin gene as well as in the Ap-fibroin gene of the Chinese oak silk moth (*Anthereae pernyi*) (Manning & Gage, 1980; Mita, Ichimura & James, 1994; Sezutsu & Yukuhiro, 2000). Other repetitive regions of DNA, such as the mouse major histocompatibility complex and the minisatellite regions in humans are also believed be subjected to recombination at specific hotspots (Jeffreys, Wilson & Thein, 1985; Steinmetz, Stephan & Fischer Lindahl, 1986).

To pursue this issue further genomic PCR of the C-terminal part of MaSp1a from individual spiders was performed. This revealed the presence of a highly similar MaSp1 gene (MaSp1b) as well as several variants of both MaSp1 genes. However, whether the observed pattern of the cDNA clones is caused by the occurrence of several MaSp1a genes in each individual or by large variation of the MaSp1a gene between individuals was not possible to determine.

Further complicating the picture, arachnids have been suggested to multiply their entire genome in somatic cells to enable high protein synthesis in specific tissues (Rasch & Connelly, 2005). This phenomenon is known as endopolyploidy and it has been documented in every animal species studied to date (including humans) and is especially common among arthropods (Gregory & Hebert, 1999; Nagl, 1976; Rasch & Connelly, 2005). Endopolyploidy arises when mitotic DNA replication is not followed by cell division (Gregory & Hebert, 1999). In spiders, endopolyploidy have been reported to occur *e.g.* in cells from the digestive tract, that can have up to 64 times the haploid DNA level (Rasch & Connelly, 2005). The presence of multiple copies of certain genes or families of genes would be beneficial when intense protein synthesis is required (Rasch & Connelly, 2005). Interestingly, the highest level of endopolyploidy observed, exceeding over one million-ploid, occurs in the silk-producing glands of *B. mori* (Perdix-Gillot, 1979). The amount of DNA in the cells producing the dope in spiders has not been determined, though endopolyploid nuclei have been observed (Gregory & Shorthouse, 2003). This fact has been largely overlooked in previous studies of the spider silk genes. Mitotic gene conversion and increased levels of recombination in polyploid cells have been documented (Shibata, 2001; Storchova *et al.*, 2006). Provided that these events are ongoing in silk gland cells, they could possibly result in unique sets of multiple variants of silk genes within individual cells. Such a scenario could perhaps also contribute to the diversity of the sequences in this study.

N-terminal nonrepetitive domain common to dragline, flagelliform and cylindriform spider silk proteins (III)

Only a few sequences encoding the N-terminal region of flagelliform (Hayashi & Lewis, 1998; Hayashi & Lewis, 2000), cylindriform (Hu*, et al.*, 2006; Zhao*, et al.*, 2005; Zhao*, et al.*, 2006) and dragline silk proteins (Motriuk-Smith*, et al.*, 2005)

have previously been described Surprisingly, these have been found to be quite heterogeneous in length. Furthermore, since spider silk proteins are secretory proteins one would expect the presence of signal peptides that direct proteins to the secretory pathway. In the published N-terminal regions of silk proteins, signal peptides could not be unequivocally determined, leaving the mechanism whereby they enter the secretory pathway open for speculation.

From a cDNA library constructed from *E. australis* major ampullate glands, a clone containing an insert of 1.2 kb was isolated. The translated sequence revealed a MaSp1 protein including a non-repetitive N-terminal region of 154 amino acid residues. In total, the non-repetitive part contained 11 Met codons, all of which represented potential translational start sites. Since the MaSp:s are secretory proteins, the sites were all scanned for possible signal peptides by using SignalP (Bendtsen*, et al.*, 2004). A maximum probability score (1.000) was obtained for the sequence corresponding to a translational start site at Met1, while the others gave relatively low probability scores or were predicted to be non-secretory proteins. Hence, Met1 was suggested to be the true translational start site. This was further corroborated by the presence of Met residues at the corresponding position in the N-terminal regions of *Nephila inaurata madagascariensis* and *Argiope trifasciata* MaSp2 proteins, the flagelliform protein of *N. inaurata madagascariensis*, as well as in cylindriform silk proteins of *Nephila clavata* (CySp1), *Latrodectus hesperus* (TuSp1) and *Argiope bruennichi* (CySp1). When analysed using SignalP, these translational start sites were all predicted to be followed by signal peptides with probability scores of 0.989-0.999. Furthermore, the non-repetitive sequences following the predicted cleavage sites were well conserved. However, three additional sequences had been reported where the translational start site was suggested to be located some 80 codons downstream in the non-repetitive region (CySp2 of *A. bruennichi*, MaSp1 of *Latrodectus geometricus* and the flagelliform protein of *N. clavipes*). This was surprising since these sequences consequently would lack a substantial part of the conserved domain shared by other spidroins. By carefully examining the coding DNA sequences, possible sequencing errors and deletions were revealed. When these were corrected, all published N-terminal regions of spider silk proteins were found to share a positionally conserved start codon and strongly predicted signal peptides followed by a 130 residue non-repetitive domain.

The major ampullate spidroins of *E. australis* contain repetitive blocks with up to 15 consecutive Ala residues, forming significant hydrophobic stretches. Transmembrane proteins have an N-terminal signal peptide that initiates the translocation just as for secretory proteins, but also contain one or several internal hydrophobic stretch(es) in the polypeptide chain that act as stop-transfer signals. The stop-transfer region forms an α-helix that is integrated into the lipid membrane. Experimental data have shown that the hydrophobic stretch needs to be about 16 residues long to anchor the protein in the membrane (von Heijne, 1988), *i.e.* the Ala stretches in *E. australis* MaSp proteins could be candidates for acting as stop-transfer signals. However, alanine stretches of 17 residues do not act as such signals, in contrast to equally long stretches of leucine or valine (Mingarro *et al.*, 2000).

Phylogenetic tree analysis showed that the N-terminal domain of different species and silk types groups according to their glandular origin. This indicates that the N-terminal domain of major ampullate, cylindriform and flagelliform silk proteins share a common origin dated before the divergence of *Nephila, Argiope, Latrodectus* and *Euprosthenops* some 125-240 million years ago.

The sequence similarity of the N-terminal domains in different silks was reflected in secondary structure predictions. By using the PHD algorithm, five positionally conserved α-helices with short connecting loops were predicted in all analysed sequences. This prediction was supported by experimental data using far-UV CD spectroscopy to evaluate the secondary structure of a recombinantly expressed N-terminal domain of *E. australis*. The spectrum showed a typical α-helical appearance and the helical content was estimated to 60%. This is in good agreement with the secondary structure prediction where the five α-helices cover approximately 70% of the domain (*cf.* IV). Helical wheel analysis showed that conserved residues, a majority of which are hydrophobic, in general tend to cluster on one surface of the helices suggesting these to be buried in a folded structure. Possibly, the tertiary structure of this domain could be a bundle of five α-helices, as has been observed in some other proteins (*e.g.* the SWIRM domain from the human transcriptional adaptor, ADA2α) (Qian *et al.*, 2005).

An intriguing observation is the presence of a strictly conserved TTGXXN motif connecting helix 4 and 5. When compared to known functional and structural motifs using the Expasy tools, no hits are found.

Structural properties of non-repetitive and repetitive parts of major ampullate spidroin 1 from Euprosthenops australis. Implications for fibre formation (IV)

To understand how the spider can maintain the spidroins in liquid form at high concentrations and the process by which the dope is converted into a fibre, stuctural studies on soluble spidroins are required. However, analysis of the dope without causing a change of the spidroins' secondary structure is a major obstacle. In this paper, structural studies of recombinantly expressed parts of spidroins, that can be purified and maintained in aqueous solution, is provided. Since full length spidroins are unlikely to be recombinantly produced in the near future, studies of representative parts are relevant for optimising the polymerisation conditions for recombinant spidroins *in vitro*.

The *E. australis* MaSp1 N- and C-terminal domain, a representative part of the repetitive region (4Rep, 4 tandem poly-Ala/Gly-rich repeats) and the repetitive part covalently linked to the C-terminal domain (4RepCT) were chosen for recombinant production and characterisation using CD spectroscopy. The 4RepCT protein corresponds to the miniature spidroin described in I. The N-and C-terminal domains were found to be folded with an α-helical content of 50-60%, in agreement with secondary structure predictions. They remained stable upon thermal heating to approximately 50°C, whereafter they unfolded and underwent

an irreversible formation of β-sheet structure. The repetitive part showed a flexible helical conformation at 20°C and a more gradual unfolding starting already at 30°C, before irreversible transition to β-sheet. The 4RepCT displayed a more abrupt change to β-sheet conformation at 50°C.

As expected the C-terminal domain formed dimers under non-reducing conditions, as judged by SDS-PAGE. CD spectra of the C-terminal domain under reduced and non-reduced conditions suggested that disruption of the disulphide bridge results in a lowered thermal stability.

The effects of factors proposed to induce fibre formation *in vivo* (*e.g.* acidification and changes in ion concentration) were examined. None of the terminal domains, the 4Rep or the 4RepCT was found to change conformation as a result of changes in pH or salt concentrations. Thus, it is possible that the change in pH and ion concentration along the spinning apparatus observed *in vivo* has limited effect on the fibre formation. However, although it appears unlikely, it remains to be determined if the MaSp:s have other, hitherto unknown, regions that may respond to these factors. Moreover, it is also possible that the full length protein is more sensitive to pH and ion concentration than the part now studied.

Concluding remarks and future perspectives

Spider silk has many excellent properties which make it suitable as an implantable biomaterial. It obtains an extreme toughness from combining strength and extensibility (Gosline, *et al.*, 1999). Furthermore, its biocompatibility, degradability and possible antimicrobial properties have attracted a lot of focus (Foelix, 1996; Vollrath, *et al.*, 2002). In spite of vast efforts, it has not been possible to fully reproduce the spiders' production of silk in a large-scale fashion. The success of hitherto published efforts has been limited. This thesis presents for the first time methods to produce recombinant miniature spidroins and macroscopic fibres under physiological conditions without the use of denaturing agents. The assembly into fibres can be controlled and do not require spinning procedures. Moreover, the material can probably be produced at low costs, since the bacterial expression system is suitable for industrial large-scale production in fermentors. This represents a major breakthrough with a serious bearing on future biomaterial development.

The miniature spidroins polymerise very efficiently and the polymerisation process can be executed in a controlled fashion. The recombinant fibres presented here show a number of favourable characteristics, but there are probably several ways to improve their properties. The process by which the fibres are formed should be studied in greater detail, since it likely affects the characteristics of the fibres. Different spinning techniques such as electrospinning and wet spinning could improve and customise the mechanical properties of the artificial silk. Spinning could also yield homogenous fibres that allow weaving or knitting of the material, thus rendering three-dimensional structures with designed mechanical properties. For applications like tissue-engineering, scaffolds with interconnected pores are desirable to provide sufficient opportunity for cell migration and expansion while maintaining transport of nutrients and metabolic wastes. By modifying the polymerisation conditions, films, foams or gels of the recombinant miniature spider silk protein can be produced instead of fibres. Furthermore, it has *e.g.* been proposed that hydrogels from natural polymers such as hyaluronate have numerous applications in tissue engineering as well as drug delivery although several disadvantages, including weak structural integrity, have been observed. Hence, there could be a role for spider silk, that exhibits impressive mechanical properties, for the formation of stable hydrogels. Since the formation of the fibres described herein can be controlled and since they form under physiological conditions, it is also possible that fibres could form *in situ*.

A major advantage of protein-based scaffolds is the possibility to tailor the properties of the fibres according to the needs in a specific tissue. It appears possible to incorporate motifs, *e.g.* RGD for promoting cell attachment or IKVAV for neurite sprouting and nerve regeneration, into the miniature spidroin amino acid sequence. Similarly, by introduction of amino acid residues with reactive side-chains, sites for attachment of functional additives, *e.g.* growth factors, can be engineered into the recombinant miniature spider silk's primary structure.

Preliminary studies have shown that both native spider silk, as well as the recombinant fibres produced in this thesis are biocompatible. To get a full picture of how recombinant spider silk may influence cell behaviour, parameters like proliferation, growth, locomotion, differentiation and survival should be further studied *in vitro* using a variety of cells including normal keratinocytes, fibroblasts, endothelial cells, chondrocytes, osteoblasts and mesenchymal stem cells. These *in vitro* studies should be followed by testing of biocompatibility and determination of degradation kinetics of the recombinantly produced fibres in animal studies.

In this thesis, a first major step towards a large-scale production of recombinant spider silk has been taken. The properties of our fibres make them attractive for biomedical and other applications. There is also an important environmental advantage of our approach since - at least in the long term perspective - recombinantly produced spider silk could develop into an attractive alternative to synthetic fibres.

The genetics behind nature's high performance fibres still remains enigmatic. Despite considerable efforts to isolate complete MaSp genes and to determine the number of silk genes present, there is only fragmental information available. The task of isolating and sequencing a region of ~10 kb that is as repetitive as the one in the MaSp genes, is a difficult challenge. To elucidate the number of genes present, techniques like fluorescent in situ hybridisation (FISH) might provide some answers although cross reactivity of the probes will probably be a major obstacle. Nevertheless, spiders are ancient organisms and a more comprehensive study of their genetics may reveal interesting and perhaps unknown mechanisms of evolution. Furthermore, increased understanding of the relationship between protein sequence motifs and mechanical properties may provide important insights to the molecular mechanisms acting in the fibre.

The N-terminal domain of different spider silks and species has been conserved for more than 100 million years, suggesting a crucial role for this part of the spidroins. The determination of its function might provide information that increases the understanding of the formation of the fibre and/or molecular mechanisms behind its mechanical properties. It is noteworthy that the fibres described in this thesis form spontaneously without an N-terminal domain present. The C-terminal domain is also conserved, however not to the extent of its N-terminal counterpart. This part of the MaSp:s forms a disulfide dependent dimer and is most likely needed for fibre formation, but its function remains to be revealed.

Recombinant production enables structural studies of partial spidroins, which might provide some answers to their function. In IV, recombinantly produced MaSp1 N- and C-terminal domains are studied by CD spectroscopy and found to be folded, stable and soluble. Further studies using *e.g.* NMR and X-ray diffraction, will probably result in increased comprehension of their structure and function. Understanding of how the soluble spidroins are converted into a solid fibre will most likely provide methods for obtaining fibres *in vitro* with improved properties. Moreover, increased knowledge of the mechanism of spidroin

polymerisation could reveal important comparative aspects to the understanding of *e.g.* amyloid fibril formation.

References

Adler, A.J., Greenfield, N.J. & Fasman, G.D. 1973. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol. 27*, 675-735.

Altman, G.H., Diaz, F., Jakuba, C., Calabro, T., Horan, R.L., Chen, J., Lu, H., Richmond, J. & Kaplan, D.L. 2003. Silk-based biomaterials. *Biomaterials 24*, 401-16.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science 181*, 223-30.

Arcidiacono, S., Mello, C., Kaplan, D., Cheley, S. & Bayley, H. 1998. Purification and characterization of recombinant spider silk expressed in Escherichia coli. *Appl Microbiol Biotechnol 49*, 31-38.

Arcidiacono, S., Mello, C.M., Butler, M., Welsh, E., Soares, J.W., Allen, A., Ziegler, D., Laue, T. & Chase, S. 2002. Aqueous processing and fiber spinning of recombinant spider silks. *Macromolecules 35*, 1262-1266.

Asakura, T., Ashida, J., Yamane, T., Kameda, T., Nakazawa, Y., Ohgo, K. & Komatsu, K. 2001a. A repeated beta-turn structure in poly(Ala-Gly) as a model for silk I of Bombyx mori silk fibroin studied with two-dimensional spin-diffusion NMR under off magic angle spinning and rotational echo double resonance. *J Mol Biol 306*, 291-305.

Asakura, T., Umemura, K., Nakazawa, Y., Hirose, H., Higham, J. & Knight, D. 2007. Some observations on the structure and function of the spinning apparatus in the silkworm Bombyx mori. *Biomacromolecules 8*, 175-81.

Asakura, T., Yamane, T., Nakazawa, Y., Kameda, T. & Ando, K. 2001b. Structure of Bombyx mori silk fibroin before spinning in solid state studied with wide angle x-ray scattering and (13)C cross-polarization/magic angle spinning NMR. *Biopolymers 58*, 521-5.

Augsten, K., Muhlig, P. & Herrmann, C. 2000. Glycoproteins and skin-core structure in Nephila clavipes spider silk observed by light and electron microscopy. *Scanning 22*, 12-5.

Baldauf, S.L. 2003. Phylogeny for the faint of heart: a tutorial. *Trends Genet 19*, 345-51.

Beckwitt, R. & Arcidiacono, S. 1994. Sequence Conservation in the C-Terminal Region of Spider Silk Proteins (Spidroin) from Nephila-Clavipes (Tetragnathidae) and Araneus-Bicentenarius (Araneidae). *J Biol Chem 269*, 6661-6663.

Beckwitt, R., Arcidiacono, S. & Stote, R. 1998. Evolution of repetitive proteins: spider silks from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae). *Insect Biochem Mol Biol 28*, 121-30.

Bell, A.L. & Peakall, D.B. 1969. Changes in fine structure during silk protein production in the ampullate gland of the spider Araneus sericatus. *J Cell Biol 42*, 284-95.

Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol 340*, 783-95.

Bini, E., Foo, C.W., Huang, J., Karageorgiou, V., Kitchel, B. & Kaplan, D.L. 2006. RGD-functionalized bioengineered spider dragline silk biomaterial. *Biomacromolecules 7*, 3139-45.

Bini, E., Knight, D.P. & Kaplan, D.L. 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. *J Mol Biol 335*, 27-40.

Blackledge, T.A. & Hayashi, C.Y. 2006. Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *J Exp Biol 209*, 2452-61.

Bon, M. 1710-1712. A Discourse upon the Usefulness of the Silk of Spiders. By Monsieur Bon, President of the Court of Accounts, Aydes and Finances, and President of the Royal Society of Sciences at Montpellier. Communicated by the Author. *Philosophical Transactions (1683-1775) 27*, 2-16.

Bram, A., Branden, C.I., Craig, C., Snigireva, I. & Riekel, C. 1997. X-ray diffraction from single fibres of spider silk. *J Appl Cryst 30*, 390-392.

Candelas, G.C. & Cintron, J. 1981. A spider fibroin and its synthesis. *J. Exp. Zool. 216*, 1-6.

Candelas, G.C., Ortiz, A. & Molina, C. 1986. The cylindrical or tubiliform glands of Nephila clavipes. *J Exp Zool 237*, 281-5.

Candelas, G.C.A., Geradu; Carrasco, Clara; Dompenciel, Raquel 1990. Spider Silkglands contain a tissue-specific Alanine tRNA that Accumulates in vitro in Response to the Stimulus for Silk Protein Synthesis. *Dev Biol 140*, 215 - 220.

Casem, M.L., Turner, D. & Houchin, K. 1999. Protein and amino acid composition of silks from the cob weaver, Latrodectus hesperus (black widow). *Int J Biol Macromol 24*, 103-8.

Challis, R.J., Goodacre, S.L. & Hewitt, G.M. 2006. Evolution of spider silks: conservation and diversification of the C-terminus. *Insect Mol Biol 15*, 45-56.

Chen, X., Knight, D.P. & Vollrath, F. 2002. Rheological characterization of nephila spidroin solution. *Biomacromolecules 3*, 644-8.

Craig, C. 1997. Evolution of arthropod silks. *Annu Rev Entomol 42*, 231-267.

Cunniff, P.M., Fossey, S.A., Auerbach, M.A., Song, J.W., Kaplan, D.L., Adams, W.W., Eby, R.K., David, M. & Vezie, D.L. 1994. Mechanical properties and thermal properties of dragline silk from the spider *Nephila clavipes*. *Polym adv technol 5*, 401 - 410.

Dicko, C., Kenney, J.M., Knight, D. & Vollrath, F. 2004. Transition to a beta-sheet-rich structure in spidroin in vitro: the effects of pH and cations. *Biochemistry 43*, 14080-7.

Dicko, C., Kenney, J.M. & Vollrath, F. 2006. Beta-silks: enhancing and controlling aggregation. *Adv Protein Chem 73*, 17-53.

Dicko, C., Vollrath, F. & Kenney, J.M. 2004. Spider silk protein refolding is controlled by changing pH. *Biomacromolecules 5*, 704-10.

Dobson, C.M. 1999. Protein misfolding, evolution and disease. *Trends Biochem Sci 24*, 329-32.

Dobson, C.M. 2001. Protein folding and its links with human disease. *Biochem Soc Symp*, 1-26.

Dong, Z.Y., Lewis, R.V. & Middaugh, C.R. 1991. Molecular Mechanism of Spider Silk Elasticity. *Arch Biochem Biophys 284*, 53-57.

Eles, P.T. & Michal, C.A. 2004. Strain dependent local phase transitions observed during controlled supercontraction reveal mechanisms in spider silk. *Macromolecules 37*, 1342-1345.

Fahnestock, S.R. & Bedzyk, L.A. 1997. Production of synthetic spider dragline silk protein in Pichia pastoris. *Appl Microbiol Biotechnol 47*, 33-9.

Fahnestock, S.R. & Irwin, S.L. 1997. Synthetic spider dragline silk proteins and their production in Escherichia coli. *Appl Microbiol Biotechnol 47*, 23-32.

Foelix, F. 1996. *Biology of spiders*. second edition second edition. Oxford University press Oxford pp.

Fornes, R.E., Work, R.W. & Morosoff, N. 1983. Molecular Orientation of Spider Silks in the Natural and Supercontracted States. *J Polym Sci Part B-Polymer Physics 21*, 1163-1172.

Frische, S., Maunsbach, A.B. & Vollrath, F. 1998. Elongate cavities and skin-core structure in Nephila spider silk observed by electron microscopy. *J Microsc 189*, 64-70.

Garb, J.E., Dimauro, T., Vo, V. & Hayashi, C.Y. 2006. Silk genes support the single origin of orb webs. *Science 312*, 1762.

Gatesy, J., Hayashi, C., Motriuk, D., Woods, J. & Lewis, R. 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science 291*, 2603-5.

Gillespie, D.B., Viney, C. & Yager, P. 1994. Raman-Spectroscopic Analysis of the Secondary Structure of Spider Silk Fiber. In *Silk Polymers. Materials Science and Biotechnology*. Edited by D. Kaplan, W.W. Adams, B. Farmer & C. Viney155-167. pp.

Gordon, J.E. 1988. *The science of structures and materials*. Scientific American Books, HPHLP. New York

Gosline, J.M., DeMont, E.M. & Denny, M.W. 1986. The structure and properties of spider silk. *Endeavour 10*, 37-43.

Gosline, J.M., Denny, M.W. & Demont, M.E. 1984. Spider Silk as Rubber. *Nature 309*, 551-552.

Gosline, J.M., Guerette, P.A., Ortlepp, C.S. & Savage, K.N. 1999. The mechanical design of spider silks: From fibroin sequence to mechanical function. *J Exp Biol 202*, 3295-3303.

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C. & Savage, K. 2002. Elastic proteins: biological roles and mechanical properties. *Philos Trans R Soc Lond B Biol Sci 357*, 121-132.

Greenfield, N. & Fasman, G.D. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry 8*, 4108-16.

Gregory, T.R. & Hebert, P.D. 1999. The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res 9*, 317-24.

Gregory, T.R. & Shorthouse, D.P. 2003. Genome sizes of spiders. *J Hered 94*, 285-90.

Grubb, D.T. & Ji, G.D. 1999. Molecular chain orientation in supercontracted and re-extended spider silk. *Int J Biol Macromol 24*, 203-210.

Guerette, P.A., Ginzinger, D.G., Weber, B.H.F. & Gosline, J.M. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science 272*, 112-115.

Ha, S.W., Gracz, H.S., Tonelli, A.E. & Hudson, S.M. 2005. Structural study of irregular amino acid sequences in the heavy chain of Bombyx mori silk fibroin. *Biomacromolecules 6*, 2563-9.

Hakimi, O., Knight, D.P., Knight, M.M., Grahn, M.F. & Vadgama, P. 2006. Ultrastructure of insect and spider cocoon silks. *Biomacromolecules 7*, 2901-8.

Hayashi, C.Y. 2002. Evolution of spider silk proteins: insight from phylogenetic analyses. *Exs*, 209-23.

Hayashi, C.Y. & Lewis, R.V. 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J Mol Biol 275*, 773-784.

Hayashi, C.Y. & Lewis, R.V. 2000. Molecular architecture and evolution of a modular spider silk protein gene. *Science 287*, 1477-1479.

Hayashi, C.Y. & Lewis, R.V. 2001. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. *Bioessays 23*, 750-6.

Hayashi, C.Y., Shipley, N.H. & Lewis, R.V. 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int J Biol Macromol 24*, 271-275.

Higgins, L.E., White, S., Nunez-Farfan, J. & Vargas, J. 2007. Patterns of variation among distinct alleles of the Flag silk gene from Nephila clavipes. *Int J Biol Macromol* 40, 201-16.

Hijirida, D.H., Do, K.G., Michal, C., Wong, S., Zax, D. & Jelinski, L.W. 1996. 13C NMR of Nephila clavipes major ampullate silk gland. *Biophys J 71*, 3442-7.

Hinman, M.B. & Lewis, R.V. 1992. Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes dragline silk is a two-protein fiber. *J Biol Chem 267*, 19320-4.

Holder, M. & Lewis, P.O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet 4*, 275-84.

Hronska, M., van Beek, J.D., Williamson, P.T., Vollrath, F. & Meier, B.H. 2004. NMR characterization of native liquid spider dragline silk from Nephila edulis. *Biomacromolecules 5*, 834-9.

Hu, X., Kohler, K., Falick, A.M., Moore, A.M., Jones, P.R. & Vierra, C. 2006. Spider Egg Case Core Fibers: Trimeric Complexes Assembled from TuSp1, ECP-1, and ECP-2. *Biochemistry 45*, 3506-3516.

Hu, X., Yuan, J., Wang, X., Vasanthavada, K., Falick, A.M., Jones, P.R., Mattina, C.L., Vierra C.A. 2007. Analysis of aqueous glue coating proteins on the silk fibers of the cob weaver, Latrodectus hesperus. *Biochemistry 46*, 3294-303.

Huemmerich, D., Helsen, C.W., Quedzuweit, S., Oschmann, J., Rudolph, R. & Scheibel, T. 2004a. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry 43*, 13604-12.

Huemmerich, D., Scheibel, T., Vollrath, F., Cohen, S., Gat, U. & Ittah, S. 2004b. Novel assembly properties of recombinant spider dragline silk proteins. *Curr Biol 14*, 2070-4.

Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K. & Mizuno, S. 2000. Silk fibroin of Bombyx mori is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio. *J Biol Chem 275*, 40517-28.

Ittah, S., Cohen, S., Garty, S., Cohn, D. & Gat, U. 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. *Biomacromolecules 7*, 1790-5.
Jackson, C.O.B., John P. 1995. Molecular weight distribution of Nephila clavipes dragline silk. *Macromolecules 28*, 5975 - 5977.
Jeffreys, A.J., Wilson, V. & Thein, S.L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature 314*, 67-73.
Jimenez, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M. & Saibil, H.R. 1999. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J 18*, 815-21.
Jin, H.J. & Kaplan, D.L. 2003. Mechanism of silk processing in insects and spiders. *Nature 424*, 1057-61.
Kallberg, Y. (2002). *Bioinformatic methods in protein characterization,* PhD thesis, Karolinska Institutet.
Kenney, J.M., Knight, D., Wise, M.J. & Vollrath, F. 2002. Amyloidogenic nature of spider silk. *Eur J Biochem 269*, 4159-4163.
Kerkam, K., Viney, C., Kaplan, D. & Lombardi, S. 1991. Liquid crystallinity of natural silk secretions. *Nature 349*, 596-598.
Knight, D.P., Knight, M.M. & Vollrath, F. 2000. Beta transition and stress-induced phase separation in the spinning of spider dragline silk. *Int J Biol Macromol 27*, 205-10.
Knight, D.P. & Vollrath, F. 1999. Liquid crystals and flow elongation in a spider´s silk production line. *Proc R Soc Lond B 266*, 519-523.
Knight, D.P. & Vollrath, F. 2001. Changes in element composition along the spinning duct in a Nephila spider. *Naturwissenschaften 88*, 179-82.
Koonin, E.V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet 39*, 309-38.
Kovoor, J. 1987. *Comparative structure and histochemistry of silk-producing organs in arachnids*. Springer-Verlag. Berlin. 160-186 pp.
Kovoor, J. & Lopez, A. 1983. Composition et histologie des l'appareil séricigene des *Argyrodes*, relations avec le comportement de ces araignés (Theridiidae). *Rev. Arachnol. 5*, 29-43.
Kummerlen, J., vanBeek, J.D., Vollrath, F. & Meier, B.H. 1996. Local structure in spider dragline silk investigated by two- dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules 29*, 2920-2928.
Lao, P.J. & Forsdyke, D.R. 2000. Crossover hot-spot instigator (Chi) sequences in Escherichia coli occupy distinct recombination/transcription islands. *Gene 243*, 47-57.
Lawrence, B.A., Vierra, C.A. & Moore, A.M. 2004. Molecular and mechanical properties of major ampullate silk of the black widow spider, Latrodectus hesperus. *Biomacromolecules 5*, 689-95.
Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J.-F., Duguay, F., Chretien, N., Welsh, E.A., Soares, J.W. & Karatzas, C.N. 2002. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science 259*, 472-476.
Lewis, R.V., Hinman, M., Kothakota, S. & Fournier, M.J. 1996. Expression and purification of a spider silk protein: A new strategy for producing repetitive proteins. *Prot Expr Purif 7*, 400-406.
Li, S.F.Y., McGhie, A.J. & Tang, S.L. 1994. New Internal Structure of Spider Dragline Silk Revealed by Atomic-Force Microscopy. *Biophys J 66*, 1209-1212.
Liao, D., Pavelitz, T., Kidd, J.R., Kidd, K.K. & Weiner, A.M. 1997. Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J 16*, 588-98.
Liu, Y., Shao, Z. & Vollrath, F. 2005. Relationships between supercontraction and mechanical properties of spider silk. *Nat Mater 4*, 901-5.
Lundmark, K., Westermark, G.T., Olsen, A. & Westermark, P. 2005. Protein fibrils in nature can enhance amyloid protein A amyloidosis in mice: Cross-seeding as a disease mechanism. *Proc Natl Acad Sci U S A 102*, 6098-102.

Madsen, B., Shao, Z.Z. & Vollrath, F. 1999. Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int J Biol Macromol 24*, 301-306.
Madsen, B. & Vollrath, F. 2000. Mechanics and morphology of silk drawn from anesthetized spiders (vol 87, pg 148, 2000). *Naturwissenschaften 87*, 288-288.
Magoshi, J., Magoshi, Y. & Nakamura, S. 1994. Mechanism of Fiber Formation of Silkworm. In *Silk Polymers*292-310. pp.
Makin, O.S. & Serpell, L.C. 2005. Structures for amyloid fibrils. *FEBS J 272*, 5950-61.
Makin, O.S., Sikorski, P. & Serpell, L.C. 2006. Diffraction to study protein and peptide assemblies. *Curr Opin Chem Biol 10*, 417-22.
Manning, R.F. & Gage, L.P. 1980. Internal structure of the silk fibroin gene of Bombyx mori. II. Remarkable polymorphism of the organization of crystalline and amorphous coding sequences. *J Biol Chem 255*, 9451-7.
Matsunami, K., Kokubo, H., Ohno, K. & Suzuki, Y. 1998. Expression pattern analysis of SGF-3/POU-M1 in relation to sericin-1 gene expression in the silk gland. *Dev Growth Differ 40*, 591-7.
Mello, C.M., Senecal, K., Yeung, B., Vouros, P., Kaplan, D. 1994. Initial characterisation of Nephila Clavipes dragline silk. In *Silk polymers. Materials science and Biotechnology*. Edited by D. Kaplan, Wade, W. W., Farmer, B. and Viney, C. American Chemical Society. Washington. 67-79. pp.
Menassa, R., Zhu, H., Karatzas, C.N., Lazaris, A., Richman, A. & Brandle, J. 2004. Spider dragline silk proteins in transgenic tobacco leaves: accumulation and field production. *Plant Biotechnol J 2*, 431-8.
Michal, C.A. & Jelinski, L.W. 1998. Rotational-echo double-resonance in complex biopolymers: a study of Nephila clavipes dragline silk. *J Biomol NMR 12*, 231-41.
Mingarro, I., Nilsson, I., Whitley, P. & von Heijne, G. 2000. Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol 1*, 3.
Mita, K., Ichimura, S. & James, T.C. 1994. Highly repetitive structure and its organization of the silk fibroin gene. *J Mol Evol 38*, 583-92.
Mita, K., Ichimura, S., Zama, M. & James, T.C. 1988. Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA. *J Mol Biol 203*, 917-25.
Motriuk-Smith, D., Smith, A., Hayashi, C.Y. & Lewis, R.V. 2005. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules 6*, 3152-9.
Nagl, W. 1976. DNA endoreduplication and polyteny understood as evolutionary strategies. *Nature 261*, 614-5.
O'Brien, J.P., Fahnestock, S.R., Termonia, Y. & Gardner, K.C.H. 1998. Nylons from nature: Synthetic analogs to spider silk. *Adv Mat 10*, 1185-1195.
Ochi, A., Hossain, K.S., Magoshi, J. & Nemoto, N. 2002. Rheology and dynamic light scattering of silk fibroin solution extracted from the middle division of Bombyx mori silkworm. *Biomacromolecules 3*, 1187-96.
Parkhe, A.D., Seeley, S.K., Gardner, K., Thompson, L. & Lewis, R.V. 1997. Structural studies of spider silk proteins in the fiber. *J Mol Recog 10*, 1-6.
Peakall, D.B. 1969. Synthesis of silk, mechanism and location. *Am. Zoologist 9*, 71-79.
Perdix-Gillot, S. 1979. DNA synthesis and endomitosis in the giant nuclei of the silk gland of Bombyx mori. *Biochemie 61*, 171-204.
Plazaola, A. & Candelas, G.C. 1991. Stimulation of fibroin synthesis elicits ultrastructural modifications in spider silk secretory cells. *Tissue Cell 23*, 277-84.
Pollock, C.M. & Shadwick, R.E. 1994. Relationship between body mass and biomechanical properties of limb tendons in adult mammals. *Am J Physiol 266*, R1016-21.
Prince, J., Mcgrath, K., Digirolamo, C. & Kaplan, D. 1995. Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry 34*, 10879-10885.
Qian, C., Zhang, Q., Li, S., Zeng, L., Walsh, M.J. & Zhou, M.M. 2005. Structure and chromosomal DNA binding of the SWIRM domain. *Nat Struct Mol Biol 12*, 1078-85.
Rasch, E.M. & Connelly, B.A. 2005. Genome size and endonuclear DN A replication in spiders. *J Morphol 265*, 209-14.

Rost, B. 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol 266*, 525-39.
Rost, B. & Sander, C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*, 584-99.
Rost, B. & Sander, C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins 19*, 55-72.
Rousseau, M.E., Lefevre, T., Beaulieu, L., Asakura, T. & Pezolet, M. 2004. Study of protein conformation and orientation in silkworm and spider silk fibers using Raman microspectroscopy. *Biomacromolecules 5*, 2247-57.
Scheller, J., Gührs, K.-H., Grosse, F. & Conrad, U. 2001. Production of spider silk proteins in tobacco and potato. *Nature Biotechnology 19*, 573-577.
Sezutsu, H. & Yukuhiro, K. 2000. Dynamic rearrangement within the Antheraea pernyi silk fibroin gene is associated with four types of repetitive units. *J Mol Evol 51*, 329-38.
Shao, Z.Z. & Vollrath, F. 1999. The effect of solvents on the contraction and mechanical properties of spider silk. *Polymer 40*, 1799-1806.
Shibata, T. 2001. Functions of homologous DNA recombination. *RIKEN Review 41*, 21-23.
Simmons, A., Ray, E. & Jelinski, L.W. 1994. Solid-State C-13 Nmr of Nephila-Clavipes Dragline Silk Establishes Structure and Identity of Crystalline Regions. *Macromolecules 27*, 5235-5237.
Simmons, A.H., Michal, C.A. & Jelinski, L.W. 1996. Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science 271*, 84-87.
Simossis, V.A. & Heringa, J. 2004. Integrating protein secondary structure prediction and multiple sequence alignment. *Curr Protein Pept Sci 5*, 249-66.
Sirichaisit, J., Brookes, V.L., Young, R.J. & Vollrath, F. 2003. Analysis of structure/property relationships in silkworm (Bombyx mori) and spider dragline (Nephila edulis) silks using Raman Spectroscopy. *Biomacromolecules 4*, 387-394.
Slotta, U., Hess, S., Spiess, K., Stromer, T., Serpell, L. & Scheibel, T. 2007. Spider Silk and Amyloid Fibrils: A Structural Comparison. *Macromol Biosci 7*, 183-188.
Sponner, A., Schlott, B., Vollrath, F., Unger, E., Grosse, F. & Weisshart, K. 2005a. Characterization of the protein components of Nephila clavipes dragline silk. *Biochemistry 44*, 4727-36.
Sponner, A., Unger, E., Grosse, F. & Weisshart, K. 2005b. Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat Mater 4*, 772-5.
Sponner, A., Vater, W., Rommerskirch, W., Vollrath, F., Unger, E., Grosse, F. & Weisshart, K. 2005c. The conserved C-termini contribute to the properties of spider silk fibroins. *Biochem Biophys Res Commun 338*, 897-902.
Stauffer, S.L., Coguill, S.L. & Lewis, R.V. 1994. Comparison of physical properties of three silks from Nephila clavipes and Araneus gemmoides. *J Arachn 22*, 5-11.
Steinmetz, M., Stephan, D. & Fischer Lindahl, K. 1986. Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell 44*, 895-904.
Storchova, Z., Breneman, A., Cande, J., Dunn, J., Burbank, K., O'Toole, E. & Pellman, D. 2006. Genome-wide genetic analysis of polyploidy in yeast. *Nature 443*, 541-7.
Sunde, M. & Blake, C. 1997. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem 50*, 123-59.
Sutherland, T.D., Campbell, P.M., Weisman, S., Trueman, H.E., Sriskantha, A., Wanjura, W.J. & Haritos, V.S. 2006. A highly divergent gene cluster in honey bees encodes a novel silk family. *Genome Res 16*, 1414-21.
Swanson, B.O., Blackledge, T.A., Summers, A.P. & Hayashi, C.Y. 2006. Spider dragline silk: correlated and mosaic evolution in high-performance biological materials. *Evolution Int J Org Evolution 60*, 2539-51.
Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kikuchi, A., Ohtomo, K., Takagi, T. & Mizuno, S. 1999. Determination of the site of disulfide linkage between heavy and light chains of silk fibroin produced by Bombyx mori. *Biochim Biophys Acta 1432*, 92-103.

Terry, A.E., Knight, D.P., Porter, D. & Vollrath, F. 2004. pH induced changes in the rheology of silk fibroin solution from the middle division of Bombyx mori silkworm. *Biomacromolecules 5*, 768-72.
Thiel, B.L., Guess, K.B. & Viney, C. 1997. Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. *Biopolymers 41*, 703-19.
Thiel, B.L., Kunkel, D.D. & Viney, C. 1994. Physical and Chemical Microstructure of Spider Dragline - a Study by Analytical Transmission Electron-Microscopy. *Biopolymers 34*, 1089-1097.
Tillinghast, E.K., Chase, S.F. & Townley, M.A. 1984. Water Extraction by the Major Ampullate Duct During Silk Formation in the Spider, Argiope-Aurantia Lucas. *J Insect Physiol 30*, 591-&.
Tracy, R.B., Baumohl, J.K. & Kowalczykowski, S.C. 1997. The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences. *Genes Dev 11*, 3423-31.
van Beek, J.D., Hess, S., Vollrath, F. & Meier, B.H. 2002. The molecular structure of spider dragline silk: folding and orientation of the protein backbone. *Proc Natl Acad Sci U S A 99*, 10266-71.
Vollrath, F. 2000. Strength and structure of spiders' silks. *Rev Mol Biotechnol 74*, 67-83.
Vollrath, F., Barth, P., Basedow, A., Engström, W. & List, H. 2002. Local tolerance to spider silks and protein polymers in vivo. *In Vivo 16*, 229-34.
Vollrath, F., Holtet, T., Thogersen, H.C. & Frische, S. 1996. Structural organization of spider silk. *Proc R Soc Lond Ser B-Biological Sciences 263*, 147-151.
Vollrath, F. & Knight, D.P. 1999. Structure and function of the silk production pathway in the spider Nephila edulis. *Int J Biol Macromol 24*, 243-9.
Vollrath, F. & Knight, D.P. 2001. Liquid crystalline spinning of spider silk. *Nature 410*, 541-8.
Vollrath, F., Knight, D.P. & Hu, X.W. 1998. Silk production in a spider involves acid bath treatment. *Proc R Soc Lond B 265*, 817-820.
Vollrath, F., Madsen, B. & Shao, Z.Z. 2001. The effect of spinning conditions on the mechanics of a spider's dragline silk. *Proc R Soc Lond Ser B-Biological Sciences 268*, 2339-2346.
von Heijne, G. 1988. Transcending the impenetrable: how proteins come to terms with membranes. *Biochim Biophys Acta. 2*, 307-33.
von Heijne, G. 1990. The signal peptide. *J Membr Biol 115*, 195-201.
Wang, S.P., Guo, T.Q., Guo, X.Y., Huang, J.T. & Lu, C.D. 2006. Structural analysis of fibroin heavy chain signal peptide of silkworm Bombyx mori. *Acta Biochim Biophys Sin (Shanghai) 38*, 507-13.
Warwicker, J. 1960. Comparative studies of fibroins II. The crystal structures of various fibroins. *J Mol Biol 2*, 350-362.
Westermark, P. 2005. Aspects on human amyloid forms and their fibril polypeptides. *FEBS J 272*, 5942-9.
Williams, D. 2003. Sows' ears, silk purses and goats' milk: new production methods and medical applications for silk. *Med Device Technol 14*, 9-11.
Wilson, R.S. 1962. The structure of the dragline control valves in the garden spider. *Q. J. Microsc. Sci. 103*, 549-555.
Wilson, R.S. 1969. Control of drag-line spinning in certain spiders. *Am Zool 9*, 103-111.
Woody, R.W. 1995. Circular dichroism. *Methods Enzymol.*, 34-71.
Work, R.W. 1977. Mechanisms of major ampullate silk fiber formation by orb-web-spinning spiders. *Trans Am Microsc Soc 96*, 170-189.
Work, R.W. 1984. Duality in major ampullate silk and precursive material from orb-web-building spiders (araneae). *Trans Am Microsc Soc 103*, 113-121.
Work, R.W. & Morosoff, N. 1982. A physico-chemical study of the supercontraction of spider major ampullate silk fibers. *Text Res J 52*, 349-356.
Xu, M. & Lewis, R.V. 1990. Structure of a Protein Superfiber - Spider Dragline Silk. *Proc Natl Acad Sci U S A 87*, 7120-7124.

Zhao, A., Zhao, T., Sima, Y., Zhang, Y., Nakagaki, K., Miao, Y., Shiomi, K., Kajiura, Z., Nagata, Y. & Nakagaki, M. 2005. Unique molecular architecture of egg case silk protein in a spider, Nephila clavata. *J Biochem (Tokyo) 138*, 593-604.

Zhao, A.C., Zhao, T.F., Nakagaki, K., Zhang, Y.S., Sima, Y.H., Miao, Y.G., Shiomi, K., Kajiura, Z., Nagata, Y., Takadera, M. & Nakagaki, M. 2006. Novel Molecular and Mechanical Properties of Egg Case Silk from Wasp Spider, Argiope bruennichi. *Biochemistry 45*, 3348-3356.

Zhou, C.Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.G. & Janin, J. 2001. Silk fibroin: structural implications of a remarkable amino acid sequence. *Proteins 44*, 119-22.

Zhou, C.Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R. & Li, Z.G. 2000. Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic Acids Res 28*, 2413-9.

Zhou, L., Chen, X., Shao, Z., Huang, Y. & Knight, D.P. 2005. Effect of metallic ions on silk formation in the Mulberry silkworm, Bombyx mori. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys 109*, 16937-45.

Acknowledgements

The work presented in this thesis was carried out at the Department of Biomedical Sciences and Veterinary Public Health and the Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences. It was generously supported by a grant (G5RD-CT-2002-00738) to the SPIDERMAN consortium from the European Commission.

First, I would like to express my sincere gratitude to my supervisors:

Wilhelm Engström, my supervisor, for your outstanding entrepreneurship and for introducing me to the exciting world of spider silk. I could never have imagined that that crazy boat trip to Finland would be the initiation to this thesis. I truly appreciate your concerns and friendship. Thanks also for all the laughter and entertaining stories during many trips and parties!

Jan Johansson, my first co-supervisor, for your expertise, your ”positiva väst” and excellent mentorship. I am grateful that you dared to take on a lost veterinarian with minimal knowledge of proteins. It has been a true pleasure working with you and I will surely miss it!

Göran Hjälm, my second co-supervisor, for interesting discussions and for your expertise in the lab. Thanks for always helping out and for being a great tutor with lots of patience!

Thanks are due to my friends and colleagues from the SPIDERMAN consortium; the vice co-ordinator Prof **Paul Hatton** and his colleagues Dr **Ana MacIntosh** and Dr **Aileen Crawford** from Sheffield University; Prof **Pankaj Vadgama** and his colleague Dr **Osnat Hakimi** of IRC materials at Queen Mary University, London; Prof **Gerard Wagemaker** and Dr **Fatima Kaya-Aerts** from Erasmus University, Rotterdam; Dr **Sonja von Aulock** from INPUT, University of Konstanz; Dr **David Knight** and Dr **Nick Skaer** from Oxford Biomaterials in Newbury; Dr **Florence Barrere** from University of Twente; Prof **Fritz Vollrath** from University of Oxford; Prof **Bror Morein** from Isconova, Uppsala; and last but definitely not least Dr **Agneta Larhed** and Dr **Cecilia Sturesson** from Q-Med, Uppsala. The progress of this project has been greatly aided by the enthusiastic help and assistance from my university´s EU-administrator Mrs **Ingrid Ragnarsdotter** and the scientific officer of the European Commission Dr **Astrid Christina Koch**.

Astri and **John Leroy** for your hospitality and for your invaluable expertise in wild life, especially in spiders. Thanks for making our trip to South Africa so successful, exciting and fun!

Torbjörn Mathisen, Björn Attoff and **Henrik Magnusson** at Radi Medial Systems AB and **Roland Hjelm** at Q-med AB for fruitful collaborations.

Past and present members of the group:
Stefan; my team-mate, for your kindness, support and friendship. It has been so much fun working together and our trip to South Africa is a memory I will cherish as long as I live. Thanks for all the parties and laughter, but also for deep discussions about life in general.
Charlotte; my roommate, for being a great friend and for always being eager to discuss private as well as scientific matters. I appreciate that you have put up with the stressed, hungry, pregnant monster I've turned into the last few months.
My; for your energy, deep knowledge and for always helping out. I am so glad you decided to join us! Thanks also for critical reading of this thesis.
Kerstin; the lab-guru, for lifting the spirit of everyone around you, always sharing your knowledge, and being patient - no matter how many stupid questions I ask.
Margareta; for your interest, scientific input and for nice chats.
Hanna; for all the good times both at work and parties, and for sharing invaluable expertise in the lab. I have missed you this last year!
Anna; for your creative suggestions and for sharing your scientific knowledge. Thanks also for being considerate and willing to share the mysteries of pregnancy and babies.
Siwei; for your knowledge, interesting discussions, and for sharing some of the Chinese cuisine with us.
Magnus; for being that nice happy guy, always eager to help. I really miss our chit-chats at the coffee breaks.
Erica; for dissecting the spiders, good laughs and fun trips.
Special thanks also to your wonderful family members; **Hanna** and **Ida, Petter, Jocke, Fredrik** and **Alva, Linus** and **Emilie, Ingela** and **Theo**.

Other past and present members of the AFB Department: Thanks for making this lab such a nice place to work! **Staffan E** for nice (but a bit crazy) skiing/conference trips, and fruitful scientific discussions. SE's group; **Cissi** for great companionship and expertise help with technicalities regarding the thesis. **Louise, Rahma, Jay, Elena, Hanan, Gunnar M,** and **Liya** for nice chats about science and life in general during lunches and coffee breaks. **Gunnar F** for lending your pipettes and bench. **Gunnar P** for being such an entertainer at dissertation parties. GP´s group; **Maria, Ida, Elin, Anette, Sara, Tiago** and **Osama** for always cheering me up. **Magnus Å**, for help in the lab as well as with theoretical issues. **Jenny, Anders** and **Frida**, you certainly made me feel welcome at the department. Also – **Tommy, Kersti** and **Maud**, for always standing up for me and for fixing the numerous economical and administrative problems I had in the beginning of this project. **Piotr** – thanks for your patience and for fixing my computer over and over again!

Members of the BVF Department; **Martin**, for making a great effort to help out when things were a bit messy. **Jonas** and **Jens** for valuable scientific input, and **Ann-Sofie** for putting up with the unorthodox arrangements and for help with economical and administrative matters.

My friends, especially; **Anna**, for being there no matter what. Thanks also to your family, **Fredrik** and my adorable godson **Alfons**, for just being yourselves and always making me feel welcome in your home. **Björn** and **Fredrik**, **Jocke** and **Anne, Noel, Mathilda** and **Åsa** for crazy wild parties back in the days and calmer (but equally nice) dinners nowadays. Thanks for fun skiing trips and for great friendship. **The "vets"** for good times and fiendship. People at BMC; **Micke, Lotta** and **Fredrik, Anh-Tri, Chris** and **Camilla**, thanks for all the fun parties and interesting discussions.

Mostly I want to thank my loving family; **Mum** and **Claes**, for being so supportive and for always being there. **Dad** and **Karin**, for nice gatherings and for letting me know that you are only a phone call away. **Lina** for being the best sister ever, I am so happy that I have you! **Kristofer**, I couldn't have wished for a better brother-in-law. I am really looking forward to becoming an aunt and spend some time with you guys!
And finally, the greatest thanks of all goes to **Åsa**, my gorgeous wife (!) for all your love, care, support and humour. I love you.