Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:56
Faculty of Veterinary Medicine and Animal Science

Molecular blueprint behind spider silk

– a multi-omics study of the major and minor ampullate glands

Sumalata Sonavane

Molecular blueprint behind spider silk

– a multi-omics study of the major and minor ampullate glands

Sumalata Sonavane
Faculty of Veterinary Medicine and Animal Science
Department of Animal Biosciences
Uppsala

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Uppsala 2024

Acta Universitatis Agriculturae Sueciae
2024:56

Cover: Photograph of a female *Larinioides sclopetarius* (bridge spider). Photo by Lena Holm, Department of Animal Biosciences, SLU, 2024.

ISSN 1652-6880
ISBN (print version) 978-91-8046-052-1
ISBN (electronic version) 978-91-8046-053-8
https://doi.org/10.54612/a.1evrucb870

Swedish University of Agricultural Sciences, Department of Animal Biosciences, Uppsala, Sweden

Print: SLU Grafisk service, Uppsala 2024

Molecular blueprint behind spider silk – a multi-omics study of the major and minor ampullate glands

Abstract

Spider silk fibers exhibit exceptional mechanical properties, inspiring efforts to develop artificial analogues. Specifically, major ampullate silk is renowned for its high tensile strength and toughness, while minor ampullate silk exhibits higher extensibility and water resistance. However, replicating these properties remains challenging due to limited knowledge of silk biosynthesis. This thesis investigates the molecular basis of silk production in the bridge spider, *Larinioides sclopetarius*, focusing on major and minor ampullate glands.

Histological analysis revealed cellular compartmentalization and layered secretion in most silk glands, suggesting a complex and multi-layered structure for the corresponding fibers. The presence of active carbonic anhydrase in most glands suggests potential pH gradients. We identified 18 predominant proteins in major ampullate silk, which are produced and secreted from five cell types organized into three epithelial zones (A, B, & C) in the gland. These secretions do not mix but form distinct layers in the lumen and persist in the silk fiber. Based on these findings, we proposed a three-layered model for the silk with major ampullate spidroin 1 (MaSp1), MaSp2, MaSp4 in the inner layer; MaSp3 and ampullate-like spidroins in the middle; and a heterogeneous class of non-spidroin proteins termed spider-silk constituting elements (SpiCE) in the outer layer. Similar cellular organization and layered secretion were observed in minor ampullate glands, with three cell types confined to three epithelial zones (A, B, & C). A comparative analysis revealed shared and distinct cellular compositions and gene expression profiles between the two ampullate glands. We further suggest that the minor ampullate silk has a multi-layered structure, analogous to major ampullate silk, but with distinct protein compositions.

This research provides novel insights into the cellular and molecular mechanisms underlying major and minor ampullate silk production. Our findings provide vital information for future efforts in the development of bioinspired materials with enhanced properties.

Keywords: silk biology, silk fiber, silk bioprocessing, single cell RNA sequencing, spatial transcriptomics

Molecular blueprint behind spider silk – a multi-omics study of the major and minor ampullate glands

Abstract

Spindeltråd har enastående mekaniska egenskaper, vilket inspirerar till att utveckla konstgjorda fibrer med samma kvaliteter. Särskilt känd är tråden som produceras i de stora ampullkörtlarna (major ampullate glands) då den har hög draghållfasthet och seghet, medan tråden från de små ampullkörtlarna (minor ampullate glands) uppvisar högre töjbarhet och vattenbeständighet. Att replikera dessa egenskaper är dock fortfarande utmanande på grund av begränsad kunskapsnivå om de molekylära processer som styr biosyntesen av spindeltråd. Därför undersöker den här avhandlingen den molekylära grunden för silkesproduktion i brospindeln, Larinioides sclopetarius, med fokus på major och minor ampullate körtlar som producerar dem starkaste och segaste spideltrådarna.

Histologisk analys avslöjade cellulär kompartmentalisering och skiktad sekretion i de flesta silkeskörtlar, vilket tyder på en komplex och flerskiktsstruktur för motsvarande fibrer. Närvaron av aktivt kolsyraanhydras i de flesta körtlar antyder potentiella pH-gradienter. Vi identifierade 18 dominerande proteiner i major ampullate silke, som produceras och utsöndras av sex celltyper organiserade i tre epitelzoner (A, B och C) i körteln. Dessa sekret blandas inte utan bildar distinkta lager i lumen som kvarstår i silkesfibern. Baserat på dessa fynd föreslog vi en treskiktsmodell för silke med major ampullate spidroin 1 (MaSp1), MaSp2, MaSp4 i det inre lagret; MaSp3 och ampullate-liknande spidroins i mitten; och en heterogen grupp av icke-spidroinproteiner som kallas spindelsilke-konstituerande element (SpiCE) i det yttre lagret. Liknande cellulär organisation och skiktad utsöndring observerades i minor ampullate körtlen, där tre celltyper var närvarande i tre olika epitelzoner (A, B och C). En jämförande analys visade likadana men också annorlunda cellulära kompositioner och genuttrycksprofiler hos de två körtlarna. Vi föreslår att minor ampullate silke har en flerskiktsstruktur, jämförbar med major ampullate silke, men med olika proteinkompositioner.

Denna forskning ger nya insikter i de cellulära och molekylära mekanismerna bakom major och minor ampullate silkesproduktion. Därför är dessa resultat grundläggande för ytterligare utveckling av bioinspirerade material som har förbättrade egenskaper.

Dedicated to my father.

Contents

List of publications 13

List of tables 15

List of figures 17

Abbreviations 19

1. Introduction 23
   - 1.1 Spider silk and applications 23
     - 1.1.1 Silk types and their functions 24
     - 1.1.2 Spidroin architecture 24
     - 1.1.3 Mechanical properties 25
   - 1.2 Major ampullate silk and glands 27
     - 1.2.1 The major ampullate gland 27
     - 1.2.2 Major ampullate silk composition 29
   - 1.3 Minor ampullate silk and glands 33
     - 1.3.1 The minor ampullate gland 33
     - 1.3.2 Minor ampullate silk composition 33
   - 1.4 Major and minor ampullate fibers 35
     - 1.4.1 Fiber structure 35
     - 1.4.2 Supercontraction 37
     - 1.4.3 Mechanism of spider silk spinning 37
   - 1.5 Other silk types and glands 40
     - 1.5.1 Gland morphology 40
     - 1.5.2 Spidroins 42
   - 1.6 Spider genes and genomes 43
     - 1.6.1 Spider genomes 43
     - 1.6.2 Spidroin genes 48

2. Aims of this thesis 51

3. Methods 53

3.1 Animals 53
3.2 Histology 53
3.3 Sequencing methods 54
3.3.1 Bulk RNA sequencing 54
3.3.2 Single cell RNA sequencing 55
3.3.3 Spatial transcriptomics 55
3.4 Liquid Chromatography – Mass Spectrometry (LC-MS/MS) 57
3.5 Challenges in analysis of bioinformatic data 58

4. Results and discussion 59
4.1 Paper I 59
4.2 Paper II 65
4.3 Paper III 70

5. Conclusions and outlook 75

References 79

Popular science summary 97

Populärvetenskaplig sammanfattning 101

Acknowledgements 105

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Sonavane, S.**, Westermark, P., Rising, A., & Holm, L. (2023). Regionalization of cell types in silk glands of *Larinioides sclopetarius* suggest that spider silk fibers are complex layered structures. *Scientific Reports*, *13*(1), 22273. https://doi.org/10.1038/s41598-023-49587-z.

II. **Sonavane, S.**, Hassan, S., Chatterjee, U., Soler, L., Holm, L., Mollbrink, A., Greco, G., Fereydouni, N., Pettersson, O. V., Bunikis, I., Churcher, A., Lantz, H., Johansson, J., Reimegård, J., Rising, A. (2024). Origin, Structure and Composition of the Spider Major Ampullate Silk Fiber revealed by Genomics, Proteomics, and Single-Cell and Spatial Transcriptomics. *Science Advances*, (in press).

III. **Sonavane, S.***, Chen, W.*, Chatterjee, U., Hassan, S., Holm, L., Reimegård, J., Rising, A. The Spider Minor Ampullate Gland and Silk: Insights from Multi-Omics and Comparative Analyses (manuscript). *Contributed equally.

Papers I-II are reproduced with permission from the publishers.

List of tables

Table 1: Mechanical properties of five types of silk spun by *A. argentata* (Blackledge & Hayashi, 2006a). Values are shown as mean ± std. error of the mean. 27

Table 2: MaSp subtypes, repetitive motifs and their secondary structure in the fiber. 31

Table 3: Details of available spider genome assemblies. Data obtained from NCBI in July 2024. NA - data not available. 44

Table 4: Summary of results obtained from paper I. ALS- anterior lateral spinneret, PMS- posterior median spinneret, and PLS- posterior lateral spinneret. The CA staining is shown as weak (+), intermediate (++), strong (+++) or no staining (-). 63

List of figures

Figure 1: Silk glands of araneid spiders. (A) Photograph (left) of ventral side of *Trichonephila clavipes* indicating the spinnerets. A schematic (right) of silk gland anatomy and the function of the silk type produced is indicated. Only one set of the paired silk glands is shown. The dots in the spinnerets represent the spigots of the different glands, from where the silk fiber emerges. (B) Schematic sketch of a typical spidroin composed of an N-terminal domain (NT), a repetitive region, and a C-terminal domain (CT). Figure reproduced and modified from (Babb et al., 2017), under Creative Commons Attribution 4.0 International (CC BY 4.0) License. 25

Figure 2: Typical stress-strain curve of *Trichonephila* dragline silk. Figure modified with permission from (Bergmann et al., 2022) under CC BY license (4.0). 26

Figure 3: Major ampullate gland. (A) Schematic image of major ampullate gland indicating different anatomical regions (tail, sac, funnel, duct, spinneret) and the three different epithelial zones: A, B, and C. (B) Schematic image indicating pH values along the gland as reported by (Andersson et al., 2014). Illustrations are inspired from (Rising & Johansson, 2015). 28

Figure 4: Number of genes from different spidroin classes identified in selected spider species. The colors correspond to different spidroin classes. 49

Figure 5: Schematic overview of steps involved in spatial transcriptomics. Fresh-frozen sections are placed on capture areas with special barcodes that contain millions of probes. Each probe is attached to

the surface from its 5' end and has an amplification handle, a spatial barcode, a unique molecular identifier (UMI) and a polydT capture region. Created with BioRender.com. 56

Figure 6: Silk glands of L. sclopetarius female spider. (A-G) Photos of the seven silk gland types: major ampullate (A), minor ampullate (B), aggregate (C), tubuliform (D), flagelliform (E), piriform (F), and aciniform (G) glands. (H) Schematic diagram of silk glands and their corresponding spinnerets (ALS, anterior lateral spinneret; PMS, posterior median spinneret; PLS, posterior lateral spinneret). Scale bar = 1 mm. Abbreviations: MA, major ampullate; MiA, minor ampullate; Pi, piriform; Tu, tubuliform; Fl, flagelliform; Ac1, aciniform type I; Ac2, aciniform type II; AgA, anterior aggregate; AgP, posterior aggregate. Note: Only one set of glands is shown for clarity, except for tubuliform glands. Minor ampullate and flagelliform tails were partially lost during dissection. Figure modified with permission from Scientific Reports (Sonavane et al., 2023). 60

Figure 7: Ampullate glands of *L. sclopetarius*. (A) Morphology of cells identified in the three zones of major and minor ampullate glands. The secretion from each zone forms a layer in the lumen (Lu), indicated as I, II and II in the last panel. Scale bar = 20 μm. (B–C) Schematic illustrations of major (B) and minor (C) ampullate glands. The three anatomical parts (tail, sac & duct) and the three epithelial zones in the tail and sac (zone A, B, & C) are indicated. 61

Figure 8: Proposed model with structure and composition for the major ampullate gland and silk. (A) Schematic of the major ampullate gland indicating the spatial distribution of different cell types. (B) Schematic of the three-layered major ampullate fiber. (C) Heatmap shows the expression of 18 silk genes in the different cell types, black boxes indicate marker genes. (D) Heatmap shows the protein levels in three urea extracts (2M, 4M, & 8M) of the major ampullate fiber. Black boxes indicate proteins identified as significantly enriched (p-value <0.05). 69

Abbreviations

AcSp	Aciniform spidroin
AgSp	Aggregate spidroin
ALS	Anterior lateral spinneretes
CA	Carbonic anhydrase
CARS	Coherent anti-stokes Raman scattering
cDNA	Complementary DNA
CRP	Cysteine-rich protein
CT	C-terminal domain
FlSp	Flagelliform spidroin
GC-MS	Gas chromatography-mass spectrometry
H&E	Hematoxylin and Eosin
LC-MS	Liquid chromatography-mass spectrometry
LLPS	Liquid-liquid phase separation
m/z ratio	Mass-to-charge ratio
MaSp	Major ampullate spidroin
MiSp	Minor ampullate spidroin
mRNA	Messenger ribonucleic acid
MS/MS	tandem mass spectrometry
NCBI	National Center for Biotechnology Information

NMR	Nuclear magnetic resonance
NT	N-terminal domain
OCT	Optimal cutting temperature
PAS	Periodic acid-Schiff base
PBS	Phosphate buffer saline
PCA	Principal component analysis
PLS	Partial least square regression
PLS	Posterior lateral spinnerets
PMS	Posterior median spinnerets
PySp	Pyriform spidroin
SAXS	Small-angle X-ray scattering
scRNA-seq	Single cell RNA sequencing
SpiCE	Spider silk constituting elements
TEM	Transmission electron microscopy
TuSp	Tubuliform spidroin
UMAP	Uniform manifold approximation projection
UMI	Unique molecular identifier
WAXS	Wide-angle X-ray scattering

Three-letter codes for the 20 naturally occurring amino acids

Alanine	Ala
Argenine	Arg
Asparagine	Asn
Aspartate	Asp
Cystein	Cys
Glutamate	Glu
Glutamine	Gln
Glycine	Gly
Histidine	His
IsoLeucine	Ile
Leucine	Leu
Lysine	Lys
Methionine	Met
Phenylalanine	Phe
Proline	Pro
Serine	Ser
Threonine	Thr
Tryptophane	Trp
Tyrosine	Tyr
Valine	Val

1. Introduction

1.1 Spider silk and applications

Spider silk has captivated researchers for decades due to its exceptional mechanical properties (Gosline et al., 1999). Notably, one of the fiber types that spiders produce, the major ampullate silk, exhibits a unique combination of high tensile strength and extensibility, resulting in a superior toughness compared to high-performance synthetic materials like Kevlar (Gosline et al., 1999; Vollrath & Knight, 2001). Spider silk fibers are produced using only renewable resources and under ambient conditions, making them environmentally friendly materials. Spider silk is also biocompatible as it induces minimal inflammatory reactions with tolerance comparable to synthetic polymers, making it a promising candidate for medical applications (Bergmann et al., 2022; Leal-Egana & Scheibel, 2010; Vollrath et al., 2002). A recent study has even demonstrated room-temperature ferromagnetism in spider dragline silks with specific structural defects (Ranade, 2024), highlighting their potential for novel functionalities. Spider silk fibers retain their properties across a wide range of temperatures (Yang et al., 2005), which is an important feature for industrial applications. The combination of these exceptional properties positions spider silk as a highly attractive material with potential for applications across diverse fields (Bakhshandeh et al., 2021; Holland et al., 2019; Vendrely & Scheibel, 2007).

Native and recombinant spider silk fibers have already been utilized for several applications, for example in the medical field, they have been successfully employed as sutures (Hennecke et al., 2013), for wound healing (Öksüz et al., 2021), as biosensors (Xu et al., 2019), for tissue engineering

(MacIntosh et al., 2008; Salehi et al., 2020; Strauss et al., 2024), and for nerve regeneration (Kornfeld et al., 2021; Millesi et al., 2021; Stadlmayr et al., 2024). Spider silk also has potential for industrial applications, such as textiles, construction materials and high-performance ropes (Bourzac, 2015).

1.1.1 Silk types and their functions

Individual spiders of the orb-weaving *Araneidae* family produce up to seven distinct silk types within specialized silk glands in their opisthosoma (abdomen) (Foelix, 1996; Kovoor, 1987; Peakall, 1969) (**Figure 1**). Five of these glands, namely the major ampullate, minor (or small) ampullate, flagelliform (or coronate), aciniform, and tubuliform (or cylindriform) glands, are responsible for spinning silk fibers (Foelix, 1996; Peakall, 1969). In addition, the aggregate glands produce a glue-like substance, while the piriform (or pyriform) glands produce a substance that is a combination of both fibers and glue (Greco et al., 2020; Wang et al., 2020). Spiders utilize these different silk types for various purposes, including building webs to catch flying insects. The major ampullate silk provides structural support for the webs and serves as a safety line. This fiber is sometimes referred to as dragline silk, but this is partially misleading as the dragline typically comprises major ampullate, minor ampullate, and potentially other silk types (Boutry & Blackledge, 2010; Young et al., 2021). The minor ampullate silk is also used for temporary scaffolding of the webs, while the flagelliform silk creates the capture spiral of the web which is coated with sticky glue droplets produced by aggregate glands (Kovoor, 1987). Piriform glue is used to cement the web junctions and to attach silk to surfaces (Wolff et al., 2015). Aciniform silk is used for both swathing the captured prey and for the soft inner lining of the egg sacs. Finally, tubuliform silk, produced only by females, is used as a protective outer coating of egg sacs (Kovoor, 1987).

1.1.2 Spidroin architecture

Spider silk fibers are primarily composed of silk proteins, known as spidroins, which are generally large (250–350 kDa) and present a specific molecular architecture featuring non-repetitive N-terminal and C-terminal domains flanking a highly repetitive central region (Ayoub et al., 2007; Challis et al., 2006; Hayashi et al., 1999; Rising et al., 2006) (**Figure 1B**). The terminal domains are around 110-160 amino acid residues long, are unique to spidroins, evolutionary conserved and present in almost all

spidroins (Babb et al., 2022; Babb et al., 2017; Garb et al., 2010; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021; Rising et al., 2006). The terminal domains are responsible for regulating spidroin solubility and polymerization, while the presence of specific amino acid motifs in the repetitive region are associated with fiber mechanical properties (Andersson et al., 2014; Askarieh et al., 2010; Hagn et al., 2010; Hayashi et al., 1999; Rising, 2014). The spidroins are named based on the silk gland in which they are predominantly found or expressed, such as Major ampullate Spidroins (MaSp), Minor ampullate Spidroins (MiSp), et cetera.

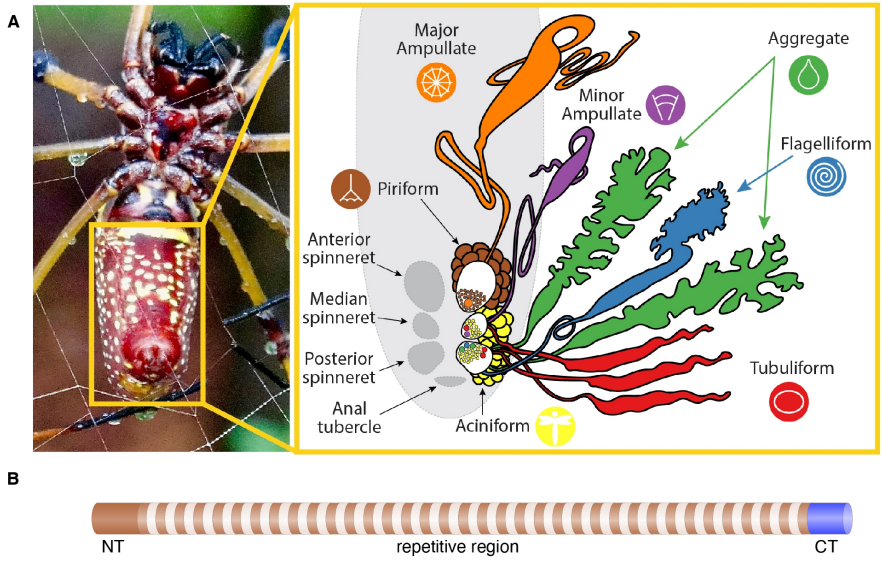

Figure 1: Silk glands of araneid spiders. (A) Photograph (left) of ventral side of *Trichonephila clavipes* indicating the spinnerets. A schematic (right) of silk gland anatomy and the function of the silk type produced is indicated. Only one set of the paired silk glands is shown. The dots in the spinnerets represent the spigots of the different glands, from where the silk fiber emerges. (B) Schematic sketch of a typical spidroin composed of an N-terminal domain (NT), a repetitive region, and a C-terminal domain (CT). Figure reproduced and modified from (Babb et al., 2017), under Creative Commons Attribution 4.0 International (CC BY 4.0) License.

1.1.3 Mechanical properties

The mechanical properties of spider silk fibers are typically measured and evaluated using stress-strain curves generated through tensile tests (Gosline

et al., 1999). The stress (σ) is determined by the ratio of the applied load or force (F) to the cross-sectional area of the fiber (A). The strain (ε), on the other hand, is given by the ratio of displacement in length (ΔL) to the original fiber length (L). A typical stress-strain curve for a major ampullate silk fiber (as shown in **Figure 2**) displays an initial, linear part called the elastic region where the fiber deformation is reversible. The slope of this linear part corresponds to Young's modulus (E), a measure of the fiber's stiffness. As the stress and strain increase, the curve reaches the yield point, marking the transition from elastic to plastic region. In the plastic region, the fiber deforms irreversibly with a non-linear increase in the stress. The maximum stress the fiber can withstand before breaking is defined as the ultimate tensile strength of the fiber. The strain at break is defined as the ultimate extensibility. The fiber toughness is the amount of energy required to break the fiber and is calculated by the area under the curve (Gosline et al., 1999).

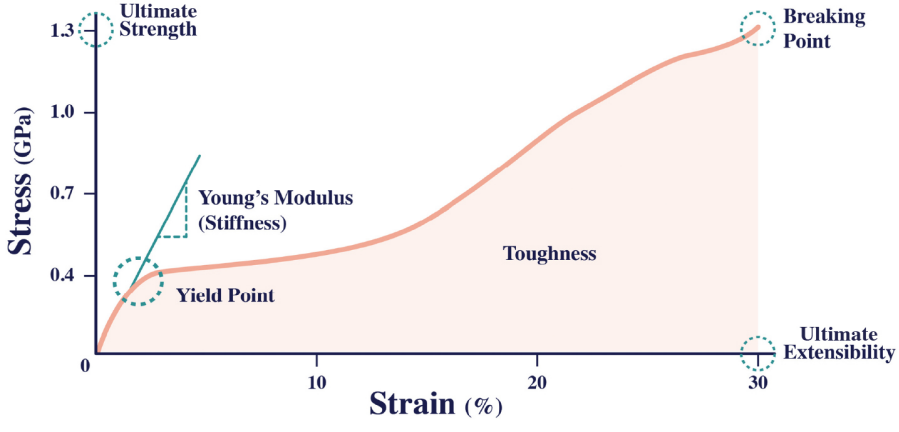

Figure 2: Typical stress-strain curve of *Trichonephila* dragline silk. Figure modified with permission from (Bergmann et al., 2022) under CC BY license (4.0).

Each spider silk type displays unique mechanical properties that appear to be specifically tailored for its function (Blackledge & Hayashi, 2006a; Gosline et al., 1999). For instance, the major ampullate silk, known for its strength and extensibility, is ideal for the main structural elements of the web (Gosline et al., 1999). Conversely, minor ampullate silk is less strong but offers higher extensibility, which is required for temporary scaffolding of the web (Blackledge & Hayashi, 2006a). The capture silk is the most extensible (Gosline et al., 1999), while aciniform silk is the toughest silk among all silk

types (Blackledge & Hayashi, 2006a). The mechanical properties of a specific silk type can vary significantly between different species (Arakawa et al., 2022), but as an example the mechanical properties of the silk fiber types from the orb-weaver spider, *Argiope argentata,* are summarized in **Table 1**.

Table 1: Mechanical properties of five types of silk spun by *A. argentata* (Blackledge & Hayashi, 2006a). Values are shown as mean ± std. error of the mean.

Silk type	Young's modulus (GPa)	Ultimate strength (GPa)	Ultimate extensibility (%)	Toughness (J cm^{-3})
Major ampullate	8.0±0.8	1.495±0.065	20.5±0.5	136±7
Minor ampullate	10.6±1.2	0.923±0.154	33±3.3	137±22
Capture spiral	0.001±0.0001	0.534±0.04	172±5	75±6
Tubuliform	11.6±2.1	0.476±0.09	28.57±1.5	95±17
Aciniform	10.4±1.4	1.052±0.12	40.4±2.4	230±31

1.2 Major ampullate silk and glands

1.2.1 The major ampullate gland

The major ampullate gland comprises three distinct anatomical regions: a long and convoluted tail, a wider sac (or ampulla), and a narrow, tapering duct (**Figure 3A**). The tail and sac are comprised of three types of single-layered columnar epithelial cells with basally located nuclei and numerous secretory vesicles (Andersson et al., 2013). These cells are confined to three sharply demarcated zones: A, B, and C (Andersson et al., 2013) (**Figure 3A**). According to results obtained using immuno-transmission electron microscopy (TEM), the cells in zone A and B produce spidroins, which are subsequently secreted and stored as separate layers within the gland lumen (Andersson et al., 2013). In contrast, the cells in zone C do not appear to produce spidroins, but their secretion forms a third layer in the lumen (Andersson et al., 2013).

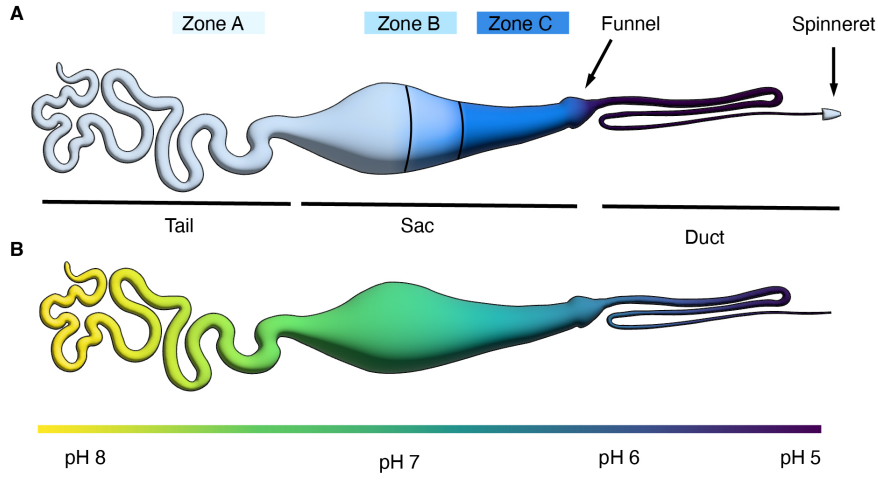

Figure 3: Major ampullate gland. (A) Schematic image of major ampullate gland indicating different anatomical regions (tail, sac, funnel, duct, spinneret) and the three different epithelial zones: A, B, and C. (B) Schematic image indicating pH values along the gland as reported by (Andersson et al., 2014). Illustrations are inspired from (Rising & Johansson, 2015).

The duct is a long structure with three limbs folded into an S-shape (**Figure 3A**). Its diameter tapers towards the spigots. The duct is lined by a thick cuticular intima, the functions of which are not fully understood, but it is suggested to provide structural support to the duct and protect the underlying epithelial cells from damage during fiber pulling (Rising & Johansson, 2015). Additionally, the cuticle is thought to function as a dialysis membrane, facilitating the exchange of ions and dehydration of the liquid silk feedstock (dope) as it transforms into a solid fiber (Vollrath & Knight, 1999). The sac and the duct are connected via the funnel (Andersson et al., 2013) whose function remains unknown.

Along the gland, there is a pH gradient, from slightly basic pH (7.6) in the proximal tail region to an acidic pH (5.7) halfway through the spinning duct, and it likely decreases further in the last part of the duct (Andersson et al., 2014) (**Figure 3B**). Enzyme activity staining of histological sections has demonstrated that carbonic anhydrase (CA) plays a vital role in generating and maintaining this remarkable pH gradient, which was also confirmed by measurements of the intraluminal pH in the presence and absence of methazolamide, a membrane-permeable CA inhibitor (Andersson et al.,

2014). CAs are a diverse group of metalloenzymes that catalyze the reversible conversion of carbon dioxide (CO_2) and water (H_2O) into bicarbonate (HCO_3^-) and a proton (H^+). These enzymes are found across all kingdoms of life and categorized into eight distinct families: α-, β-, γ-, δ-, ζ-, η-, θ-, and ι-CAs (Nocentini et al., 2021; Supuran, 2018). Among these, α-CAs, the most extensively studied group, are present in various prokaryotic and eukaryotic organisms. CAs play diverse physiological roles, including acid secretion, regulation of acid-base balance, and water homeostasis (Davenport, 1939; Purkerson & Schwartz, 2007).

In the major ampullate gland, active CA is present from zone C to the end of the duct (Andersson et al., 2014). As the pH decreases, both the HCO_3^- concentration and CO_2 pressure increase along the length of the gland (Andersson et al., 2014) (**Figure 3B**). Within the sac, the concentrations of Na^+, K^+, and Cl^- ions are 199 mM, 6 mM, and 164 mM, respectively (Andersson et al., 2014). Along the duct, the concentrations of Na^+ and Cl^- decrease while K^+, P, and S increase (Knight & Vollrath, 2001). However, the specific ion concentrations in the duct remain unknown.

1.2.2 Major ampullate silk composition

Major ampullate spidroins

Semi-quantification of proteins in major ampullate fibers using mass spectrometry has shown that the MaSps are the most abundant proteins, although the relative abundance of different MaSp types can differ between species (Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono, Ohtoshi, et al., 2021; Larracas et al., 2016; Nakamura et al., 2023). In major ampullate silk extracted from four *Nephilinae* spider species, MaSps collectively constitute over 90% (w/w) of the fiber in all four species (Kono, Nakamura, et al., 2021; Kono, Ohtoshi, et al., 2021).

The first full-length MaSp genes (MaSp1 and MaSp2) were described from the black widow spider, *Latrodectus hesperus (Theridiidae),* in 2007 (Ayoub et al., 2007) and revealed that the encoded proteins contain extended repetitive regions capped by terminal domains. Since then, many more MaSp genes have been characterized and the MaSp family is currently categorized into five subfamilies (MaSp1–5) based on the specific motifs within their repetitive regions and phylogenetic clustering of the terminal domains. The repetitive regions of the MaSps are characterized by a high abundance of

alanine and glycine, forming distinct motifs, as shown in **Table 2**. MaSp1 contains poly-Ala, Gly-Gly-X (where X = Ala, Leu, Gln, Tyr, Arg), and Gly-X-Gly motifs (where X = Leu, Arg, Gln, Trp, Pro) (Arguelles et al., 2023). MaSp2 contains poly-Ala and Gly-Gly-X found in MaSp1, as well as Gly-Pro-Gly and Gln-Gln motifs. The repetitive regions of MaSp3 vary between species, with some species featuring poly-Ala (Arguelles et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021) and Gly-Gly-X motifs (Kono, Nakamura, et al., 2021; Kono et al., 2019). While in other species, such as *L. Hesperus*, the MaSp3 lacks the poly-Ala motifs and instead contains more polar and more acidic residues than the MaSp1 and MaSp2 from the same species (Collin et al., 2018). Interestingly, MaSp4 and MaSp5, so far identified only in the *Caerostris* family, lack poly-Ala motifs (Garb et al., 2019; Kono, Ohtoshi, et al., 2021). MaSp4 is rich in Pro, with 52% of its repetitive region comprised of Gly-Pro-Gly-Pro-Gln motifs, while MaSp5 is abundant in Gly-Gly-X motifs (Garb et al., 2019; Kono, Ohtoshi, et al., 2021; Rui Wen et al., 2022b).

These repeat motifs adopt distinct secondary structures in the major ampullate fiber which are suggested to contribute to the mechanical properties (Hayashi et al., 1999). The poly-Ala motifs form antiparallel β-sheets while the Gly-rich region forms β-turns and 3_1-helices (Ayoub et al., 2007; Dicko et al., 2004; Hayashi et al., 1999; Hijirida et al., 1996; Lefèvre et al., 2011; Parkhe et al., 1997; van Beek et al., 2002). The specific contribution of different MaSp types to the mechanical properties of the silk is a topic of ongoing research, but there is some correlation between the presence of distinct MaSps and the mechanical properties of the fiber. For instance, MaSp1 likely contributes to the fiber strength, while MaSp2 is correlated with fiber extensibility and supercontraction (Arakawa et al., 2022; Hayashi & Lewis, 1998a). MaSp3 and MaSp4 are suggested to be correlated with higher fiber toughness (Arakawa et al., 2022; Babb et al., 2022; Garb et al., 2019).

Table 2: MaSp subtypes, repetitive motifs and their secondary structure in the fiber.

MaSp type	Characteristic motifs	Potential role in (correlation with) mechanical properties	Secondary structure in the fiber	References
MaSp1	poly-Ala, Gly-Gly-X, Gly-X-Gly	Fiber strength	poly-Ala: β-sheets, Gly-rich: β-turns and 3_1-helices	(Arakawa et al., 2022; Arguelles et al., 2023; Ayoub et al., 2007; Babb et al., 2022; Dicko et al., 2004; Garb et al., 2019; Hayashi & Lewis, 1998b; Hayashi et al., 1999; Hijirida et al., 1996; J. E. Jenkins et al., 2010; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021; Lefèvre et al., 2011; Parkhe et al., 1997; van Beek et al., 2002)
MaSp2	poly-Ala, Gly-Gly-X, Gly-Pro-Gly, Gln-Gln	Fiber extensibility, supercontraction		
MaSp3	Vary between species, often contain poly-Ala and Gly-Gly-X	Fiber toughness		
MaSp4	Gly-Pro-Gly-Pro-Gln	Fiber toughness		
MaSp5	Gly-Gly-X			

Other proteins

In addition to MaSps, multiple other spidroin types have been identified in the major ampullate glands as mRNA transcripts (Babb et al., 2022; Babb et al., 2017; Hu et al., 2023; Lane et al., 2013; Whaite et al., 2018), and in the major ampullate silk using proteomic studies, albeit to much lower levels than MaSps (Chaw et al., 2015; Hu et al., 2023; Jorge et al., 2022; Larracas et al., 2016; Nakamura et al., 2023). The identification of these additional spidroins may be attributed to the complex composition of dragline silk, which harbors various silk types, as mentioned earlier. A novel class of low-molecular-weight non-spidroins, collectively termed spider silk-constituting elements (SpiCEs), has also been discovered in the fiber using proteomic analyses and account for 1 – 5 % of the fiber's protein content (Kono, Nakamura, et al., 2021; Kono, Ohtoshi, et al., 2021). The SpiCEs are a diverse group of proteins that differ in molecular weight and exhibit low sequence similarity (Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono,

Ohtoshi, et al., 2021; Pham et al., 2014), suggesting a lack of a shared function. A subset of these proteins are Cys-rich (sometimes referred to as Cys-rich proteins (CRPs)) (Larracas et al., 2016; Pham et al., 2014), but the sequence similarity between these proteins is generally low (Pham et al., 2014). Using Western blot analysis of crude bacterial lysates containing recombinant CRP1 under reducing and non-reducing conditions, Pham et al. showed that CRPs can assemble into high-molecular-weight complexes by formation of disulfide bonds (Pham et al., 2014). The authors further performed molecular modeling of CRP1 and suggested that these proteins form macromolecular complexes via cysteine slipknot-like structures and disulfide bonding, which may contribute to the fiber's toughness (Pham et al., 2014). Only one study has investigated the mechanical properties of silk composites produced from recombinant spidroins and a SpiCE protein (Kono, Nakamura, et al., 2021). Composite films made of recombinant MaSp2 and 0-5% (by weight) of SpiCE-NMa1 (found in the *Nephilinae* family) result in a two-fold increase in the tensile strength compared to films made of MaSp2 alone. However, corresponding composite silk fibers with SpiCE-NMa1 exhibit inferior properties compared to pure spidroin-based counterparts (Kono, Nakamura, et al., 2021).

Other components

Several studies have identified lipids in silk fibers, composed mainly of long-chain (C_{28}-C_{34}), highly hydrophobic lipids (Schulz, 2001; Sponner et al., 2007; Vollrath & Knight, 1999). Analysis of the *T. clavipes* web using gas chromatography coupled with mass spectrometry (GC-MS) has revealed alkyl methyl ethers as the major lipid group (50-80%), with other identified components including fatty acids, wax esters, and glyceryl ethers (Schulz, 2001). Sponner et al. (2007) identified an outermost lipid layer in the *T. clavipes* dragline, using a lipophilic dye oil red. They suggested that this layer (10-20 nm) is formed exclusively by lipids.

Coherent Anti-Stokes Raman Scattering (CARS) microscopic analysis, that images CH_2-stretches, has demonstrated that both major and minor ampullate silks in *T. madagascariensis* have lipid layers with a thickness of (0.93 ± 0.1) μm and (0.65 ± 0.1) μm, respectively (Iachina et al., 2023). The authors propose that the observed discrepancy in layer thickness for major ampullate silks (compared to previous reports) might be due to species-specific variations or sample preparation methods. Furthermore, Laurdan Generalized Polarization measurements have indicated that the lipids in the

major ampullate fiber are in a liquid or gel-phase, while the minor ampullate silk exhibits a more liquid-ordered phase. This distinction is attributed to the functional differences between the two silk types (Iachina et al., 2023).

In addition to lipids, dragline silk has been reported to contain an outer layer rich in glycoproteins (Augsten et al., 2000; Sponner et al., 2007).

1.3 Minor ampullate silk and glands

1.3.1 The minor ampullate gland

The minor ampullate glands share the gross morphology of the major ampullate glands, with the only notable difference being their smaller size (Kovoor, 1987; Mullen, 1969). The tail and sac of minor ampullate glands are composed of two types of tall, single-layered columnar epithelial cells with basally located nuclei (Moon & Kim, 1989a; Mullen, 1969). Active CA is present in intracellular vesicles and at the apical cell membrane of the epithelium in the distal part of the minor ampullate sac and duct (Andersson et al., 2014). The gland's duct connects to a spigot on the median spinnerets (Moon & Kim, 1989b; Mullen, 1969). The duct has a thick cuticular intima and a thin layer of outer connective tissue (Moon & Kim, 1989b). The duct can be divided into proximal (near the sac) and distal regions (near the spinnerets). The distal duct region has columnar epithelial cells, whereas the proximal region has cuboidal or squamous cells (Moon & Kim, 1989b).

1.3.2 Minor ampullate silk composition

Minor ampullate spidroins

The minor ampullate spidroins (MiSps) are the main constituents of the minor ampullate silk. The first documented full-length MiSp gene sequence was that of *Araneus ventricosus* (Chen et al., 2012). Subsequent genomic/transcriptomic studies have identified MiSp genes in several species (Arakawa et al., 2022; Babb et al., 2017; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021). There are no reports detailing the distinguishing features of the different MiSp genes or if they can be further classified as MiSp1, MiSp2, etc., as the MaSps. One study reports that in *T. clavipes*, the MiSp1 and MiSp2 share similar repetitive motifs, although the MiSp2 motifs are less conserved and

that MiSp1 and MiSp2 can be distinguished by a conserved serine residue in the Gly-Ala-rich repeat motifs of MiSp2 (Colgin & Lewis, 1998). If this is applicable to other species' MiSps is not known.

The repetitive regions of MiSps primarily encompass Gly-Ala, Gly-Gly-X, and poly-Ala amino acid motifs (Chen et al., 2012). The Gly-Gly-X and poly-Ala motifs are also found in some MaSps (see section 1.2), however, the poly-Ala motifs in MiSps are shorter and less abundant compared to those in MaSps (Chen et al., 2012; Colgin & Lewis, 1998; Hayashi et al., 1999). The MiSps feature small non-repetitive regions, called "spacers", that interrupt the repetitive regions. These regions are about 130 amino acid residues long and rich in serine (Colgin & Lewis, 1998; Qi et al., 2024). They are highly conserved and fold into a globular domain (Qi et al., 2024). Integration of the spacer domain into a recombinant chimeric spidroin promotes self-assembly into silk-like fibers and leads to enhanced tensile strength of the fibers (Qi et al., 2024). The effects of the spacer are suggested to be attributed to the presence of amyloidogenic motifs in the spacer domain (Qi et al., 2024).

Other proteins and components

Using nano liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS), Nakamura et al. (2023) found that the minor ampullate silks of *A. ventricosus* and *Trichonephila clavata* primarily consist of spidroins, but also SpiCE proteins and other small components similar to those in the major ampullate glands. Interestingly, the *A. ventricosus* minor ampullate silk is primarily made up of MiSp proteins with a small fraction of MaSp1 and SpiCE proteins, while *T. clavata* minor ampullate silk contains larger fractions of MaSp1 than MiSp. These compositional differences are suggested to arise from the distinct functional roles of minor ampullate silk in the two species. The authors further revealed that certain SpiCE proteins are shared between the major and minor ampullate silks in both species. Notably, SpiCE-NMa1, previously identified in the major ampullate silk from the *Nephilinae* family (Kono, Nakamura, et al., 2021), was also found in the minor ampullate silk of *T. clavata* (Nakamura et al., 2023).

1.4 Major and minor ampullate fibers

1.4.1 Fiber structure

The major ampullate silk fibers can significantly vary in diameter depending on the species (1– 14 μm) (Arakawa et al., 2022; Augsten et al., 2000; Iachina et al., 2023; Sampath et al., 2012). Various techniques have been employed to elucidate the ultrastructure of major ampullate silk, including light microscopy (Vollrath et al., 1996), TEM (Augsten et al., 2000; Frische et al., 1998; Sponner et al., 2007; Vollrath & Knight, 1999), scanning electron microscopy (Augsten et al., 2000; Du et al., 2006; Frische et al., 1998; Sponner et al., 2007; Wang & Schniepp, 2018), atomic force microscopy (Du et al., 2006; Li et al., 1994; Miller et al., 1999; Silva & Rech, 2013; Wang & Schniepp, 2018), small- and wide-angle X-ray scattering (SAXS & WAXS) (Miller et al., 1999; Riekel et al., 2017; Riekel & Vollrath, 2001), confocal microscopy (Augsten et al., 2000; Iachina et al., 2023), and nuclear magnetic resonance (NMR) spectroscopy (Hronska et al., 2004). There is an emerging consensus that the fiber has a skin-core structure (Andersson et al., 2013; Augsten et al., 2000; Frische et al., 1998; Iachina et al., 2023; Li et al., 1994; Poza et al., 2002; Riekel et al., 2017; Sponner et al., 2007; Vollrath et al., 1996) with spidroins in the central core (Sponner et al., 2005; Sponner et al., 2007) which are encased by unknown proteins, carbohydrates and lipids in the skin layer (Augsten et al., 2000; Iachina et al., 2023; Sponner et al., 2007). However, the exact number of layers is debated, and models with three (Iachina et al., 2023; Li et al., 1994), four (Vollrath et al., 1996), and five (Sponner et al., 2007) layers have been proposed.

The major ampullate fiber core is composed of so called nanofibrils (about 10 nm in diameter) running parallel to the fiber axis (Iachina et al., 2023; Li et al., 1994; Sapede et al., 2005; Stehling et al., 2019; Vollrath et al., 1996; Wang & Schniepp, 2018; Yazawa et al., 2019), which is thought to be important for the tensional integrity of the fiber (Fraternali et al., 2020). These nanofibrils are semi-crystalline with β-crystals distributed in an amorphous matrix (Du et al., 2006; Sapede et al., 2005; Silva & Rech, 2013; Simmons et al., 1996). The β-crystals are composed of the poly-Ala motifs of the MaSps that form stacked β-sheets, while the amorphous regions comprise the less-ordered glycine-rich motifs (Parkhe et al., 1997; Simmons et al., 1996). Although often referred to as amorphous, the glycine-rich regions in the fiber contain 3_1-helical Gly-Gly-X motifs (Kümmerlen et al.,

1996; van Beek et al., 2002) and Gly-Pro-Gly-X-X motifs that form type II β-turns (Janelle E. Jenkins et al., 2010).

The poly-Ala β-crystals contribute to the fiber's strength, while the Gly-rich repeats are more important for the fiber's flexibility and extensibility (Guerette et al., 1996; Holland et al., 2008; Janelle E. Jenkins et al., 2010; Simmons et al., 1996; van Beek et al., 2002). When the major ampullate fiber is stretched, the fiber undergoes several structural changes before it is fractured (Keten et al., 2010; Nova et al., 2010). Upon initial stretching, the fiber exhibits elastic deformation (**Figure 2**). This is a reversible process attributed to the stretching of the amorphous regions (Keten et al., 2010; Nova et al., 2010). After the yield point, irreversible plastic deformation occurs possibly due to the disruption of hydrogen bonds within the amorphous matrix. If the fiber continues to be stretched, the forces are transferred to the β-sheet crystals. Here, a "stick-slip" mechanism (dynamic breaking and reformation of hydrogen bonds) involving the β-strands within the crystals temporarily hinders the fiber fracture. Continued stretching beyond this point leads to that the crystals eventually fail and the fiber fractures (Cranford et al., 2012; Keten et al., 2010; Nova et al., 2010).

The minor ampullate silk exhibits a skin-core architecture similar to that of the major ampullate silk but the diameter of these fibers is smaller (Guinea et al., 2012; Iachina et al., 2023; Nakamura et al., 2023). A recent study on the major and minor ampullate silks from *Trichonephila inaurata madagascariensis* revealed that both types of silk contain an autofluorescent protein core encased in an outer lipid-rich layer (Iachina et al., 2023). The protein core in both fibers is comprised of two layers, with the inner layer consisting of nanofibrils aligned parallel to the long fiber axis. However, the fibrils in the minor ampullate silk are smaller (116 nm ± 12 nm) compared to those found in the major ampullate silk (145 nm ± 18 nm). The hierarchical nanostructure of both ampullate silks is similar in that they consist of highly oriented nanocrystals interconnected by amorphous chains (Gosline et al., 1999; Hayashi et al., 1999; Papadopoulos et al., 2009; Parkhe et al., 1997; Riekel & Vollrath, 2001), but in the minor ampullate silk, the crystals are formed by both poly-Ala and Gly-Ala motifs (Parkhe et al., 1997). Although the size of crystals in minor ampullate silks varies between species, minor ampullate silk is generally more crystalline and has larger β-sheet crystals compared to major ampullate silk (Nakamura et al., 2023; Sampath et al., 2012).

1.4.2 Supercontraction

Upon exposure to high humidity or water immersion, the major ampullate silk fibers can dramatically decrease in length by up to ~60% (Work, 1977a). This phenomenon significantly alters the mechanical properties of the silk, leading to increased strain at break (i.e., greater extensibility) and decreased tensile strength (Liu et al., 2005). WAXS and Raman spectroscopy studies of major ampullate silk fibers have shown that exposure to water disrupts the orientation of both the β-sheet crystals and the amorphous components relative to the fiber axis (Grubb & Ji, 1999; Sampath & Yarger, 2015; Shao et al., 1999). The exact reason for supercontraction is not fully understood but presence of specific motifs within MaSp proteins, including Tyr-Gly-Gly-Leu-Gly-Ser-(Asn)-Gln-Gly-Ala-Gly-Arg in MaSp1 (Yang et al., 2000), Gly-Pro-Gly-X-X in MaSp2 (Liu et al., 2008), and Gly-X-Gly in MaSp1 (Guan et al., 2011), are suggested to play a role in swelling and contraction of major ampullate silk.

The minor ampullate silk demonstrates notable water resistance compared to the major ampullate silk. When in contact with water, the minor ampullate silk undergoes minimal supercontraction (<14%) and maintains its mechanical properties to a higher degree compared to major ampullate silk (Guinea et al., 2012; Nakamura et al., 2023; Work, 1977b). Studies have shown that the amorphous and crystalline blocks in the major ampullate silk experience greater loss of orientation during supercontraction compared to those in the minor ampullate silk (Nakamura et al., 2023; Sampath et al., 2012), resulting in differing levels of supercontraction between the two types of silk. Possibly, the increased resistance to supercontraction of the minor ampullate silk could be related to the absence of the MaSp specific motifs mentioned above. However, the degree of supercontraction appears to vary across species; for instance, *Trichonephila* minor ampullate silk exhibits a higher level of contraction than *Araneus* and *Argiope* (Nakamura et al., 2023). The reason for this inter-species difference is not known.

1.4.3 Mechanism of spider silk spinning

Spidroin storage and assembly

The spidroins are stored at high concentrations in a soluble form in the lumen of the silk glands before they are converted to fibers in the secretory ducts

that lead to the spinnerets (**Figure 3A**). Most of the knowledge related to the storage and polymerization of spidroins stems from studies of the major ampullate gland and biochemical and biophysical characterization of recombinant versions of the MaSps and MiSps. The polymerization of the major ampullate silk fiber is proposed to be a precisely controlled molecular event mediated by the spidroin terminal domains via a "lock and trigger" mechanism (Rising & Johansson, 2015). In the lumen of the gland, the spidroins are dimeric, interconnected by their CT domains which form a constitutive dimer (Andersson et al., 2014; Gao et al., 2013; Hagn et al., 2010). As the dope travels from the sac and through the duct, the terminal domains undergo significant structural changes in response to decreased pH and shear forces (Hagn et al., 2010; Landreh et al., 2010). The NT undergoes sequential protonation of specific amino acid residue side chains, and forms a stable antiparallel homodimer (Gaines et al., 2010; Kronqvist et al., 2014; Landreh et al., 2010; Otikovs et al., 2015; Schwarze et al., 2013). This leads to the interconnection of the protein chains into large molecular networks ("the lock") since the spidroins are already interconnected via their CT domains (Andersson et al., 2014; Kronqvist et al., 2014). Concurrent with the NT dimerization, the CT domain starts to unfold and transitions into a β-sheet conformation in response to the reduced pH and increased shear forces along the duct (Andersson et al., 2014; De Oliveira et al., 2024; Hagn et al., 2010). This, in combination with the fiber pulling forces and elevated pCO_2, is thought to trigger the structural transition of the repetitive regions ("the trigger") (Andersson et al., 2014). When the fiber is pulled from the spider, the pulling forces will propagate via the protein chains, since they are firmly interconnected via the stable NT dimers and constitutive CT dimers, leading to the alignment of poly-Ala blocks of the repetitive regions and the formation of β-sheet crystals (Rising & Harrington, 2023; Rising & Johansson, 2015). Additionally, factors such as extensional flow, changes in salt concentration, and water removal are also important for the spinning process (Knight & Vollrath, 1999; Vollrath & Knight, 1999; Vollrath & Knight, 2001).

Both spidroin terminal domains are evolutionary conserved, unique to spiders and serve as hallmarks for the classification of a protein as a spidroin. This suggests that the polymerization process is similar for different spider silk fiber types. In support of this, the mechanism of NT dimerization is highly conserved between different spidroin types, although different sets of

carboxylate side chains can be involved (Otikovs et al., 2015; Sarr et al., 2022). Likewise, the monomer subunit of the CT dimer is stabilized by two intramolecular salt bridges, of which at least one is evolutionarily conserved across different spidroin types and across spider species (Andersson et al., 2014; Collin et al., 2018; Gao et al., 2013; Hagn et al., 2010; Strickland et al., 2018). Disruption of this salt bridge could occur during spinning due to the changes in ion concentrations along the duct which would destabilize the CT (Andersson et al., 2014).

The spidroins are remarkably soluble when stored in the sac, forming an aqueous liquid feedstock with concentrations reaching 30–50 % (w/v) (Chen et al., 2002; Hijirida et al., 1996). Three main models of how soluble spidroins are organized within the gland have emerged: as micelles (Jin & Kaplan, 2003; Parent et al., 2018), as a liquid crystalline feedstock (Knight & Vollrath, 1999; Lin et al., 2017; Vollrath & Knight, 2001), and as phase-separated liquid droplets (Malay et al., 2020).

Micelle theory: This theory proposes that spidroins self-assemble into micelles. The highly soluble terminal domains form a hydrophilic outer shell, shielding the aggregation-prone repetitive regions in the core (Hagn et al., 2010; Jin & Kaplan, 2003; Kronqvist et al., 2017; Parent et al., 2018). As the dope travels through the narrow duct of the gland, it experiences shear forces which are thought to elongate the micelles. Simultaneously, the decreased pH will lead to antiparallel NT dimerization which could interconnect individual micelles and initiate the fiber formation process (Eisoldt et al., 2012; Hagn et al., 2010; Kronqvist et al., 2017). In support of this theory, cryo-TEM and NMR spectroscopy studies of diluted dope suggest the presence of small micelles (~50 nm in diameter) that further form hierarchical assemblies of several hundred nanometers in size (Parent et al., 2018). Additionally, micelle formation has been observed in recombinant mini-spidroins containing both terminal domains (Andersson et al., 2017).

Liquid crystalline theory: This theory states that the dope adopts a liquid crystalline state (Kerkam et al., 1991; Knight & Vollrath, 1999; Vollrath & Knight, 2001) allowing the proteins to flow through narrow ducts as a liquid while also maintaining an ordered structure (Knight & Vollrath, 1999). While different studies have reported liquid crystals of varying sizes in the sac (1–2.2 μm in diameter) (Knight & Vollrath, 1999; Lin et al., 2017), the exact organization of spidroins within these structures remains unclear.

Liquid-liquid phase separation (LLPS) theory: Proteins that undergo phase separation form two distinct aqueous phases, a concentrated phase and a more dilute phase. Recombinant spidroins and native dope can both undergo LLPS as they form large droplets (1-10 μm diameter) when incubated at relatively high concentrations of potassium phosphate (> 300 mM) and other kosmotropic anions (Exler et al., 2007; Malay et al., 2020; Mohammadi et al., 2019). Interestingly, acidification of the phase-separated spidroins containing both terminal domains triggers the formation of nanofibrils, suggesting a potential role for LLPS in silk formation (Malay et al., 2020). A recent study further demonstrated that the LLPS is driven by salting out effects and that the CT domain is essential for the phase separation of the spidroins and the formation of fibers (Leppert et al., 2023).

These three models are not mutually exclusive and may coexist. It should be noted that structural studies of native soluble dope are challenging as the experimental procedures often include drying, dilution, or fixation of the dope, which tend to alter the proteins' conformation and may introduce artifacts.

1.5 Other silk types and glands

In this section, the following silk types and glands are described: flagelliform, aggregate, tubuliform, aciniform, and piriform.

1.5.1 Gland morphology

Flagelliform glands

Flagelliform glands have been referred to as "coronata" glands (Rodríguez & Candelas, 1995) and median ampullate glands (Mullen, 1969), in the latter case, due to their resemblance to the ampullate glands. However, flagelliform glands have a smaller tail, a less bulged sac, and a much shorter duct compared to ampullate glands (Mullen, 1969). They also lack the funnel region found in the ampullate glands. In *T. clavipes*, the tail of the flagelliform gland has a branched structure (Rodríguez & Candelas, 1995). The tail and sac of the gland are composed of two types of tall columnar epithelial cells and the ducts connect to a spigot on the posterior lateral spinneret (Mullen, 1969; Peters & Kovoor, 1991).

Tubuliform glands

The tubuliform glands are only found in females, and the development of these glands closely parallels the maturation of the ovaries during vitellogenesis (Kovoor, 1987; Moon, 1998). Individual spiders can have three pairs of tubuliform glands, two of which connect to the posterior lateral spinnerets, while the third pair connects to the posterior median spinnerets (Mullen, 1969; Peters & Kovoor, 1991). The gland body is a tubular structure connected to the duct via a structure that resembles the funnel of the ampullate glands (Mullen, 1969). The gland wall is lined by columnar epithelial cells (Mullen, 1969). Studies of the *L. hesperus*, have shown that both the tubuliform glands and the underlying epithelial cells undergo significant morphological changes throughout the female's reproductive cycle (Herrera et al., 2015). During the vitellogenic phase, the glands are bigger and exhibit shorter epithelial cells filled with vesicles, likely associated with increased secretory activity. In contrast, the glands are smaller and have taller cells when the spider is in the avitellogenic phase (Herrera et al., 2015).

Aciniform glands

Two types of aciniform glands are described: type A and type B (Kovoor, 1987; Mullen, 1969). The type A glands occur in groups and are somewhat spherical, with a small nob at the anterior end. Each gland has a long duct that connects to the posterior lateral spinneret. The type B glands, on the other hand, occur singly and are fewer in number. These glands have an elongated anterior end, and their ducts connect to the posterior median spinnerets (Mullen 1969). In some species, two cell types can be distinguished in the secretory portion of these glands (Kovoor, 1987) but in *L. sclopetarius* only one cell type has been described (Mullen, 1969).

Aggregate glands

Two pairs of aggregate glands, anterior and posterior, are reported (Kovoor, 1987; Mullen, 1969; Peters & Kovoor, 1991). In araneid spiders, both pairs are similar in their anatomy and histochemistry, however, they differ in theridiid spiders (Kovoor, 1987). The glands are large and characterized by an irregularly folded body made of single-layered thin columnar epithelial cells with a spacious lumen (Kovoor, 1987; Moon, 2018; Mullen, 1969). The

gland connects to convoluted ducts that have cuticular intima surrounded by single-layered epithelial cells with basal nuclei. The epithelium of the duct is further surrounded by "nodules" which are made of a group of cells with large prominent nuclei and indistinct cell boundaries (Mullen, 1969). The function of the nodules of aggregate glands is unclear but they are reported to be rich in glycogen and mitochondria (Moon, 2018; Zylberberg & Kovoor, 1979).

Piriform glands

These glands occur in large groups. Depending on the species, they are either gourd-shaped (Mullen, 1969) or pear-shaped (Kovoor, 1987; Peters & Kovoor, 1991), and their ducts exit through the anterior spinnerets (Mullen, 1969; Peters & Kovoor, 1991). Two types of columnar epithelial cells have been reported in these glands (Kovoor, 1987; Kovoor & Zylberberg, 1980; Mullen, 1969; Peters & Kovoor, 1991).

1.5.2 Spidroins

The repetitive regions of flagelliform spidroins (FlSp) contain Gly-Pro-Gly-Gly-X and Gly-Gly-X motifs (where X is usually Ala, Ser, Tyr, or Val), which occur >40 times and ~8 times, respectively, before being interrupted by a "spacer" region (Hayashi & Lewis, 2001). The FlSp spacer is ~28 amino acid residues long and highly conserved non-repetitive region that shows very little sequence similarity to other spidroins (Hayashi & Lewis, 1998a, 2001). Aciniform (AcSp) and tubuliform spidroins (TuSp) have less defined motif compositions, characterized by long, complex repeat units with high intragenic homogenization (Ayoub et al., 2013; Wen et al., 2017; Wen et al., 2018a). Common motifs in TuSp repeats include poly-Ala, poly-Ser, poly-(Ser-Ala), Ala-X, and Ser-Gln (Wen et al., 2017; Zhao et al., 2006). Aggregate spidroins (AgSp) have repetitive regions containing Gly-Gly-Gln, Pro-Gly-Gly, Gly-Pro-Gly, Gln-Gly-Pro, and Gln-Gln motifs (Stellwagen & Renberg, 2019). Similarly, piriform spidroins (PySp) have proline-rich motifs (Pro-X-Pro-X-Pro) and Gln-Gln-containing motifs within their repetitive regions.

1.6 Spider genes and genomes

1.6.1 Spider genomes

Despite the remarkable diversity of over 52,000 identified spider species classified within 134 families (World Spider Catalog, https://wsc.nmbe.ch/), the availability of high-quality annotated spider genome assemblies remains limited. However, due to the accessibility of low-cost and high-throughput next-generation sequencing methods, the last decade has seen a growing number of spider assemblies. As of July 2024, the National Center for Biotechnology Information (NCBI) contains 67 spider genome assemblies representing 46 unique species (including the *L. sclopetarius* assembly) (**Table 3**).

The assembly level varies considerably, with 23 of them reaching the chromosomal level, offering the most contiguous and complete representation. However, a substantial number remain at the scaffold (38 assemblies) and contig levels (6 assemblies) (**Table 3**). The average spider genome size is approximately 2.53 Gb, with the longest spider genome reported for *Maratus speciosus* (9.11 Gb). The number of chromosomes ranges from 9 to 46 (**Table 3**). The GC content varies from 27% to 40%. Interestingly, the average GC content in spiders (32%) falls on the lower end of the typical range observed in eukaryotes (30–60%) (Bohlin & Pettersson, 2019). The exact reason for this, despite the prevalence of Gly and Ala in spidroin genes, which are encoded by GC-containing codons, remains unclear.

Out of the 67 assemblies deposited on NCBI, 16 are annotated (including *L. sclopetarius*) (**Table 3**). Of these, 4 have NCBI RefSeq annotation, a repository that contains a curated collection of well-annotated genomes. Notably, one of these belongs to *Argiope bruennichi,* a spider from the orb-weaving araneid family (Sheffer et al., 2021). The reported total number of genes in spiders varies greatly, from 19348 to 528426, with a median of 33692 genes. The median number of protein-coding genes is 28426. The spider genomes have short exons and long introns (Sanggaard et al., 2014).

Table 3: Details of available spider genome assemblies. Data obtained from NCBI in July 2024. NA - data not available.

	Accession	Name	Assembly length (Gb)	Assembly level	Chromosome count	GC percent	Genes	Protein coding genes
1	GCA_032207245.1	*Pardosa pseudoannulata*	2.417	Chromosome	15	29	NA	NA
2	GCA_037975125.1	*Latrodectus hesperus*	1.644	Chromosome	13	28.5	NA	NA
3	GCA_951213105.1	*Amaurobius ferox*	3.565	Chromosome	23	33	NA	NA
4	GCA_026930045.1	*Uloborus diversus*	2.152	Chromosome	10	34	19348	15750
5	GCA_039090855.1	*Macrothele yani*	6.793	Chromosome	46	40	NA	NA
6	GCA_947563725.1	*Argiope bruennichi*	1.778	Chromosome	13	29.5	37666	18361
7	GCA_933210815.1	*Meta bourneti*	1.383	Chromosome	13	35.5	NA	NA
8	GCA_963680715.1	*Tetragnatha montana*	0.785	Chromosome	13	33.5	NA	NA
9	GCA_949128135.1	*Parasteatoda lunata*	1.411	Chromosome	12	29.5	NA	NA
10	GCA_947359465.1	*Metellina segmentata*	1.665	Chromosome	13	35.5	NA	NA
11	GCA_907164885.2	*Dolomedes plantarius*	2.78	Chromosome	14	33.5	NA	NA
12	GCA_023701765.1	*Hylyphantes graminicola*	0.936	Chromosome	13	32.5	NA	NA
13	GCA_038502835.1	*Octonoba sinensis*	1.344	Chromosome	9	32.5	NA	NA
14	GCA_964059485.1	*Gibbaranea gibbosa*	2.817	Chromosome	13	32	NA	NA
15	GCA_963931845.1	*Maratus michaelseni*	5.956	Chromosome	15	31	NA	NA
16	GCA_964019185.1	*Maratus speculifer*	5.888	Chromosome	15	30	NA	NA

17	GCA_963978545.1	*Linyphia triangularis*	1.349	Chromosome	13	31.5	NA	NA
18	GCA_963932465.1	*Maratus speciosus*	9.11	Chromosome	15	31	NA	NA
19	GCA_963930645.1	*Salticus scenicus*	5.16	Chromosome	15	30	NA	NA
20	GCA_030067965.1	*Latrodectus elegans*	1.559	Chromosome	14	28	NA	NA
21	GCA_963932185.1	*Troglohyphantes excavatus*	1.028	Chromosome	13	32	NA	NA
22	GCA_019343175.1	*Oedothorax gibbosus*	0.821	Chromosome	13	32	29717	29717
23	GCA_000365465.3	*Parasteatoda tepidariorum*	1.229	Scaffold	NA	29.5	23655	19750
24	GCA_010614865.2	*Stegodyphus dumicola*	2.551	Scaffold	NA	33.5	26749	20729
25	GCA_024610705.1	*Tetragnatha versicolor*	1.068	Scaffold	NA	33.5	NA	NA
26	GCA_021605075.1	*Caerostris darwini*	1.502	Scaffold	NA	30	67936	56046
27	GCA_019974015.1	*Nephila pilipes*	2.695	Scaffold	NA	30.5	74147	72265
28	GCA_026289955.1	*Argiope argentata*	1.901	Scaffold	NA	29.5	NA	NA
29	GCA_013235015.1	*Araneus ventricosus*	3.657	Scaffold	NA	32.5	283660	278918
30	GCA_000661875.1	*Acanthoscurria geniculata*	7.178	Scaffold	NA	40	NA	NA
31	GCA_019973935.1	*Trichonephila clavipes*	2.874	Scaffold	NA	32	528426	517366
32	GCA_019973975.1	*Trichonephila clavata*	2.498	Scaffold	NA	31.5	78379	70418
33	GCA_000611955.2	*Stegodyphus mimosarum*	2.739	Scaffold	NA	33.5	27252	27135
34	GCA_006491805.2	*Dysdera silvatica*	1.366	Scaffold	NA	35	NA	NA

35	GCA_026290005.1	*Latrodectus geometricus*	1.685	Scaffold	NA	27.5	NA	NA
36	GCA_001188405.1	*Loxosceles reclusa*	3.262	Scaffold	NA	39.5	NA	NA
37	GCA_038380435.1	*Ryuthela nishihirai*	3.111	Scaffold	NA	38.5	NA	NA
38	GCA_964023285.1	*Larinioides sclopetarius*	2.274	Scaffold	NA	30.5	22953	22843
39	GCA_038373885.1	*Cheiracanthium punctorium*	2.568	Contig	NA	34	NA	NA
40	GCA_021605095.1	*Caerostris extrusa*	1.421	Scaffold	NA	29.5	86115	82821
41	GCA_038373865.1	*Uloborus plumipes*	1.463	Contig	NA	33.5	NA	NA
42	GCA_019973955.1	*Trichonephila inaurata madagascariensis*	2.507	Scaffold	NA	31	59979	51576
43	GCA_026543055.1	*Argiope trifasciata*	1.88	Scaffold	NA	29.5	NA	NA
44	GCA_026543865.1	*Argiope aurantia*	1.838	Scaffold	NA	29.5	NA	NA
45	GCA_947070885.1	*Tetragnatha kauaiensis*	1.086	Scaffold	NA	33.5	NA	NA
46	GCA_015342795.1	*Argiope bruennichi*	1.67	Chromosome	13	29.5	23259	23259
47	GCA_008065355.1	*Pardosa pseudoannulata*	4.208	Scaffold	NA	31.5	NA	NA
48	GCA_000697925.2	*Latrodectus hesperus*	1.234	Scaffold	NA	27	NA	NA
49	GCA_024610695.1	*Tetragnatha versicolor*	1.039	Scaffold	NA	33.5	NA	NA
50	GCA_024450205.1	*Caerostris darwini*	1.453	Scaffold	NA	30	NA	NA
51	GCA_951215355.1	*Amaurobius ferox*	2.368	Scaffold	NA	33	NA	NA
52	GCA_024586335.1	*Uloborus diversus*	2.276	Scaffold	NA	33.5	NA	NA
53	GCA_947563775.1	*Argiope bruennichi*	1.154	Scaffold	NA	29.5	NA	NA

54	GCA_963680735.1	*Tetragnatha montana*	0.778	Scaffold	NA	33.5	NA	NA
55	GCA_949128125.1	*Parasteatoda lunata*	1.27	Contig	NA	29.5	NA	NA
56	GCA_947359365.1	*Metellina segmentata*	1.372	Contig	NA	35.5	NA	NA
57	GCA_907164875.2	*Dolomedes plantarius*	2.381	Scaffold	NA	33.5	NA	NA
58	GCA_933210875.1	*Meta bourneti*	0.448	Contig	NA	36	NA	NA
59	GCA_963931785.1	*Maratus michaelseni*	5.027	Scaffold	NA	30.5	NA	NA
60	GCA_964059525.1	*Gibbaranea gibbosa*	2.864	Scaffold	NA	32	NA	NA
61	GCA_964019115.1	*Maratus speculifer*	4.593	Scaffold	NA	30	NA	NA
62	GCA_963930615.1	*Salticus scenicus*	4.776	Scaffold	NA	30	NA	NA
63	GCA_963932175.1	*Maratus speciosus*	7.46	Scaffold	NA	31	NA	NA
64	GCA_963978535.1	*Linyphia triangularis*	1.153	Scaffold	NA	31.5	NA	NA
65	GCA_963932215.1	*Troglohyphantes excavatus*	0.916	Contig	NA	32	NA	NA
66	GCA_002102615.1	*Trichonephila clavipes*	2.439	Scaffold	NA	31.5	22738	22738
67	GCA_008297655.1	*Anelosimus studiosus*	2.033	Scaffold	NA	28	NA	NA

While these annotated assemblies provide a valuable foundation for assembling and annotating novel spider species, they also present some issues. For example, two scaffold-level assemblies of the same spider species (*T. clavipes*) (Babb et al., 2017; Kono, Nakamura, et al., 2021), show slightly different genome lengths (2.44 vs. 2.87 Gb) and significantly different numbers of predicted protein-coding genes (22738 vs. 517366) (Table 3). This inconsistency is also observed in genomes assembled at the chromosomal level. For instance, a comparison between two chromosomal-level assemblies reported for *A. bruennichi*, despite having the same number of chromosomes (13), reveals disparity in total number of genes (23259 in GCA_015342795.1 and 37666 in GCA_947563725.1) and protein-coding genes (23259 in GCA_015342795.1 and 18361 in GCA_947563725.1). These issues can be attributed to inherent challenges associated with both sequence assembly and annotation of the spider genomes (as mentioned in section 3.5).

1.6.2 Spidroin genes

A limited number of reported genomic studies have described spidroins (Babb et al., 2022; Babb et al., 2017; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021; Sanggaard et al., 2014)(**Figure 4**). The spidroin catalogues reported in these studies contain some incomplete genes. The number of identified spidroins within a species varies from 11 in *A. ventricosus* (Kono et al., 2019) to 33 in *T. clavipes* (Kono, Nakamura, et al., 2021) (**Figure 4**). Notably, there is a disparity in the number of reported spidroins for the same species from different studies (compare *C. darwini, T. clavata* and *T. clavipes* in **Figure 4**). However, at least one spidroin from each of the seven classes is reported in all except *Stegodyphus mimosarum* (Sanggaard et al., 2014), which lacks the AgSp and FlSp and is not an orb-weaving spider.

Spidroin genes tend to cluster on a subset of chromosomes. For instance, 26 spidroin genes are located across nine of the 13 chromosomes in *T. clavata* (Hu et al., 2023). Furthermore, genes belonging to specific MaSp/MiSp groups are localized in distinct chromosomes. In *T. clavata*, MaSp1a–c & MaSp2e, MaSp2a–d, and MiSp-a–e genes are independently found on chromosomes 4, 7, and 6, respectively (Hu et al., 2023). In *Trichonephila antipodiana*, MaSp group 1, 2 and MiSp group are located on chromosomes 9, 1 and 4, respectively (Fan et al., 2021; Hu et al., 2023).

Spidroins have variable gene architectures. Among the different classes of spidroins, MaSp, MiSp, AgSp and FlSp are shown to have introns. The introns in these spidroin types differ in terms of sequence and length. MaSp, MiSp and AgSp genes tend to have longer but fewer introns compared to FlSp genes, with several reported MaSp, MiSp, and AgSp genes containing only single exons (Ayoub et al., 2007; Babb et al., 2017; Chen et al., 2012; Kono et al., 2019; Sanggaard et al., 2014; Stellwagen & Burns, 2021; Stellwagen & Renberg, 2019; Rui Wen et al., 2022a; R. Wen et al., 2022). In contrast, FlSp genes have evenly distributed exonic and intronic regions (Babb et al., 2017; Hayashi & Lewis, 2000; Kono et al., 2019). For instance, *A. ventricosus* FlSp gene has 23 evenly distributed intronic regions (Kono et al., 2019) and partial sequences cloned from FlSp genes of *T. clavpies* and *T. madagascarienses* contain 3 to 12 introns, each about 1420 bp long (Hayashi & Lewis, 2000). Introns within these FlSp genes share high similarity, particularly within species (average 87%) compared to between the species (about 75%). Interestingly, within a species, the introns are more similar (87%) than the exons (73%) (Hayashi & Lewis, 2000). TuSp, AcSp, and PySp generally contain single exons (Ayoub et al., 2013; Babb et al., 2017; Chaw et al., 2017; Kono et al., 2019; Sanggaard et al., 2014; Wen et al., 2018b).

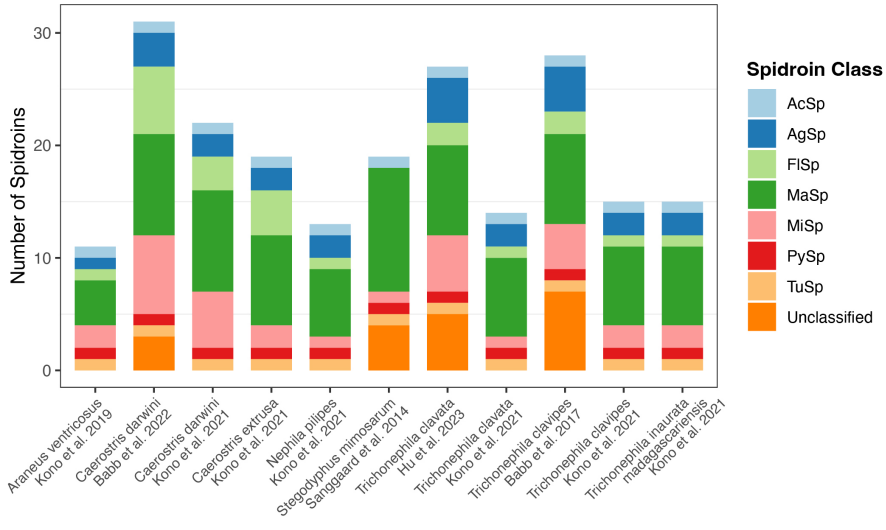

Figure 4: Number of genes from different spidroin classes identified in selected spider species. The colors correspond to different spidroin classes.

2. Aims of this thesis

The overall aim of this thesis was to characterize all the silk glands present in *L. sclopetarius* and provide in-depth knowledge of the major and minor ampullate glands and fibers of this spider species. More specific aims were:

1. To study the anatomy and morphology of silk glands in *L. sclopetarius* using various histological methods (paper I)
2. To provide a high-quality spider genome, uncover the composition and structure of the major ampullate silk fiber and to establish a spatially resolved transcriptional map of the major ampullate gland (Paper II)
3. To describe the cellular and transcriptional profile of the minor ampullate gland, identify genes important in the spinning of minor ampullate silk, and to provide a comparative analysis to the major ampullate gland and silk (Paper III)

3. Methods

Detailed methodologies corresponding to specific techniques can be found in the Materials and Methods sections of the respective papers. This chapter provides an overview of the methods used in this thesis, focusing particularly on areas requiring additional explanation.

3.1 Animals

For all the studies (Paper I–III), adult female Swedish bridge spiders, *L. sclopetarius,* were used. The spiders were collected from Skarholmen, Uppsala during the months of June to November. Taxonomic identity of the spiders was verified by the Museum of Natural History, Stockholm, Sweden. If the spiders were housed, they were kept in large containers conducive to web-spinning and were provided with meal worms or *Drosophila* flies and water.

3.2 Histology

For all studies, spiders were anesthetized (either with carbon dioxide or exposure to -20 °C) and dissected on ice-cold wax plates using phosphate-buffered saline (PBS, pH 7) or Spider Ringer solution at pH 7 (Schartau & Leidescher, 1983). Dissections involved a cut at the pedicel, followed by pinning the abdomen and removing the exoskeleton. Subsequently, the abdomen was either moved to a fixing solution (papers I-III), snap-frozen for spatial transcriptomics (in papers II & III) or further processed to isolate silk glands (papers I-III). Tissues were embedded in resin (high-resolution histology, Papers I-III), paraffin (routine stains, Paper I), or optimal cutting temperature medium (spatial transcriptomics, Papers II, III). Hematoxylin

and Eosin (H&E) staining was used in all papers for general tissue morphology. Additional stains were used in Paper I: PAS for glycoproteins (Sponner et al., 2007) and Thioflavin-S/Congo Red for detecting amyloid-like fibrils (Elghetany & Saleem, 1988; Kelenyi, 1967).

Localization of Carbonic Anhydrase (CA) activity

Due to the presence of multiple CA families, isoforms, and the incomplete characterization of spider CAs, immunohistochemical methods are not suitable for identifying them in spider tissue sections. To address this challenge, in Paper I, a specialized histochemical technique devised by Ridderstråle, was employed to localize the CA activity (Ridderstrale, 1976; Ridderstrale, 1991). This method effectively localizes CA activity on tissue sections regardless of the specific isoform present. The CA enzyme catalyzes the following reversible reaction:

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$$

In the CA staining method, resin sections are incubated floating on a solution with $NaHCO_3$, $CoSO_4$, H_2SO_4, and KH_2PO_4. In locations with CA activity, the chemical reaction shown above occurs whereby CO_2 leaves, causing a localized increase in pH and formation of a cobalt-phosphate-carbonate complex. When these sections are moved to a solution containing $CoSO_4$ and $(NH_4)_2S$, this complex converts into a black cobalt sulphide precipitate. The sections can be counterstained with dyes such as Azur or toluidine blue to visualize morphology. Control sections treated with the CA inhibitor acetazolamide can be used to eliminate nonspecific staining.

3.3 Sequencing methods

This section describes the high-throughput sequencing methods used in papers II and III, along with the challenges encountered and solutions adopted during data analysis.

3.3.1 Bulk RNA sequencing

In Paper II, bulk RNA-seq data was instrumental in identifying genes that were overexpressed in silk glands compared to other non-silk tissues in spiders. Additionally, in Papers II and III, it facilitated the identification of

differentially expressed genes in the major and minor ampullate glands, as well as in different regions within these glands.

3.3.2 Single cell RNA sequencing

Unlike bulk RNA-seq, which provides the global gene expression profiles, single-cell RNA sequencing (scRNA-seq) offers insights into the gene expression patterns of individual cells within a heterogeneous population. In papers II and III, the 10X Chromium single cell 3' platform was utilized for identifying cell populations in major and minor ampullate glands, respectively. This method involves generation of gel bead-in-emulsions (GEMs) by mixing the 10X barcoded gel beads with cells, enzyme, and partitioning oil. Within these GEMs, single cells undergo reverse transcription to generate barcoded cDNA. All cDNA generated from individual cells share a common 10X barcode. The resulting barcoded library can be directly used for single-cell whole transcriptome sequencing or targeted sequencing workflows. Utilizing standard short-read sequencing of 10X barcoded libraries facilitates extensive transcriptional profiling of thousands of individual cells, allowing for the identification of cellular subtypes within the samples. Additionally, scRNA-seq enables sample multiplexing, providing the flexibility to analyze multiple samples simultaneously. The single cell data from the major ampullate glands was further used to improve the 3' untranslated region (UTR) annotation of the genes.

3.3.3 Spatial transcriptomics

In papers II and III, we utilized spatial transcriptomics, an unbiased technique that resolves the spatial profile of a transcriptomic pool from a tissue section while retaining histological information (Stahl et al., 2016). Initially, fresh-frozen tissue is sectioned and placed onto a specialized glass slide with capture areas (**Figure 5**). Each capture area, measuring 6.5x6.5 mm, contains ~5000 spots. These spots, each roughly 55 μm in diameter and spaced 100 μm apart, effectively capture a few cells within their area. A single spot has millions of barcoded probes containing a unique molecular identifier (UMI) and a spatial barcode unique to that spot, allowing for precise localization of the captured mRNA. Additionally, they feature a poly(dT) that binds to the polyA tail of mRNA molecules and an amplification handle. Following tissue staining with dyes like H&E to

identify cell types, the tissue is permeabilized, allowing mRNA molecules to bind to the capture oligos. A reverse transcription reaction then generates cDNA libraries, which are subsequently sequenced to obtain gene expression data. This data is superimposed onto a high-resolution image of the tissue section, enabling visualization of mRNA expression within the tissue in a spatially resolved manner.

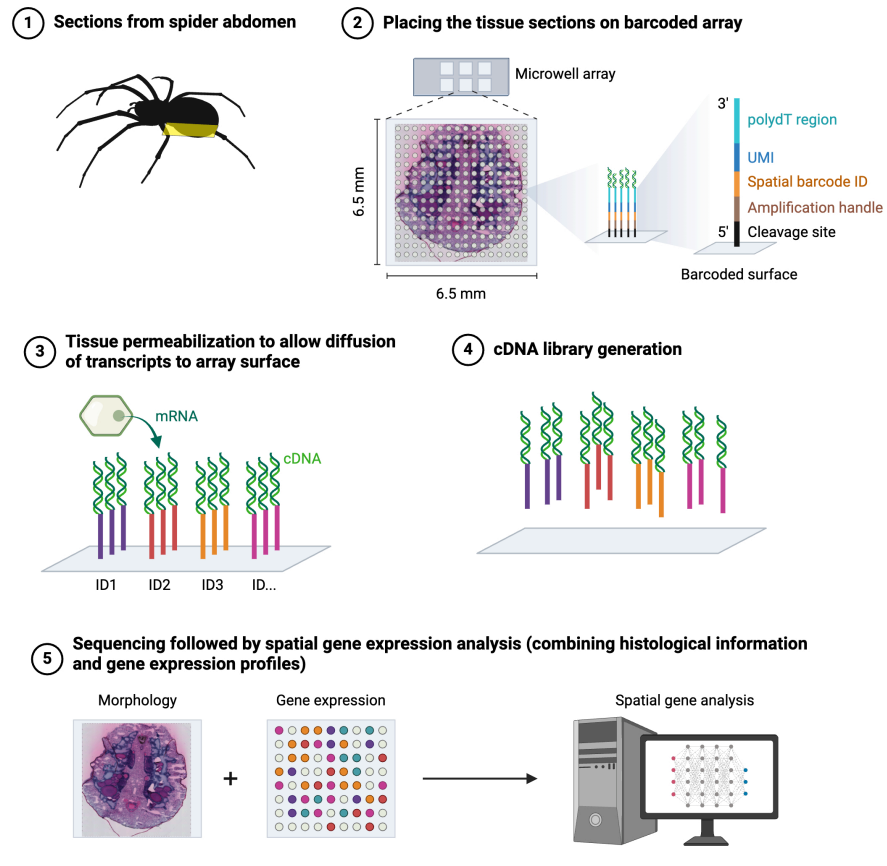

Figure 5: Schematic overview of steps involved in spatial transcriptomics. Fresh-frozen sections are placed on capture areas with special barcodes that contain millions of probes. Each probe is attached to the surface from its 5' end and has an amplification handle, a spatial barcode, a unique molecular identifier (UMI) and a polydT capture region. Created with BioRender.com.

Eight sections were obtained and sequenced using this method. The spots were categorized as covering the silk glands and other tissues based on the morphology in the H&E-stained sections. For some sections, the spots covering the major and minor ampullate glands were further categorized according to if they covered the epithelium of the zone A, B or C.

3.4 Liquid Chromatography – Mass Spectrometry (LC-MS/MS)

In order to identify the protein composition of the silk fibers and the silk glands, we employed LC-MS/MS in paper II. Briefly, the proteins in a sample are digested by a protease to generate peptides. The digested protein mixture is injected onto a liquid chromatographic column that separates the peptides based on their hydrophobicity. The eluted peptides flow into the mass spectrometer which records their mass-to-charge ratio (*m/z*) and generates a mass spectrum with precursor intensities. Tandem mass spectrometry (MS/MS) is employed to provide fragmentation spectra with amino acid sequence data used for peptide identification. The *m/z* of the product ions are searched against a database of proteins with known sequences to find matches between experimental and theoretical fragmentations.

Challenges in proteomics analysis of the silk

There are several challenges when it comes to the identification and quantification of proteins within the silk fibers using LC-MS/MS: (1) It is difficult to isolate pure major ampullate silk since the fibers are easily contaminated by other silk types during spinning, and while handling the samples. (2) Identification of the proteins requires the fiber to be dissolved and different chemical treatments can extract different proteins from the fibers (Larracas et al., 2016; Sponner et al., 2007). (3) During the analysis of proteomic data, relative quantification of the proteins within the silk fiber using label-free quantification is challenging. Some quantification methods rely on the molecular weight of the proteins, which can lead to biased results due to the large difference between the spidroins (up to 350 kDa) and the SpiCE proteins (<100 kDa) (Kono, Nakamura, et al., 2021). The extreme repetitiveness of the spidroins can also give skewed results due to multi-

mapping if using quantification based on spectral counts. Consequently, the proteomic studies of the silk fibers should be interpreted with caution.

3.5 Challenges in analysis of bioinformatic data

De-novo genome assembly is a hierarchical process where partially or fully overlapping short DNA sequences (reads) are first assembled to obtain longer contiguous assembly components, called contigs. These contigs are then linked together into larger scaffolds and ultimately assembled into chromosomes depending on the availability of mapping information. Accurate annotation of genes requires a highly contiguous genome assembly. However, short-read sequencing methods generate fragments with only up to 200 base pairs in length, posing difficulties in accurately assembling the repetitive regions of the spidroins. This limitation can lead to gaps and fragmented assemblies. Since the spidroins can have up to a hundred repeat motifs arranged in tandem, assembling these genes necessitates sequencing techniques that can generate longer reads such as PacBio sequencing (read length >10 kb) (Kumar et al., 2024; Rhoads & Au, 2015), as well as hybrid assembling approaches that combine short- and long-read data. Additionally, spider genes/proteins are not well characterized, and many proteins lack functional annotations. Since most annotation tools are automated and rely on homology to available genes and proteins, the annotation can result in automatic assignment of gene/protein names even with low sequence similarity, sometimes assigning misleading gene names.

Assembling genomes of non-model organisms, such as spiders, can be further complicated when closely related high-quality reference genomes are unavailable. When we sequenced the *L. sclopetarius* genome in 2019, only two araneid spider assemblies were available in NCBI – *A. ventricosus* (3.66 Gb) and *T. clavipes* (2.44 Gb) – both at the scaffold level (Babb et al., 2017; Kono et al., 2019). To obtain a high-quality assembly and a good coverage of spidroin genes, we employed PacBio long-read sequencing and further polished the assembly using chromium 10X-linked (short) reads.

4. Results and discussion

For all the projects in this thesis, the bridge spider (*Larinioides sclopetarius*), member of the *Araneidae* family, was selected as model organism due to several reasons. These spiders are abundant in Europe and readily accessible in Uppsala. They construct orb-webs utilizing a variety of silk types, and notably, they produce major ampullate fibers with strength ~1.2 GPa and toughness ~220 MJ/m^3, as demonstrated in Paper II.

4.1 Paper I

In this paper, we identified and characterized seven types of silk glands from *L. sclopetarius*: Major ampullate, minor ampullate, flagelliform, aggregate, tubuliform, aciniform, and piriform glands (**Figure 6**). The anatomy and location of the glands in the abdomen were determined by dissection under a microscope. Using thin (1-2 μm) resin sections for histological evaluation allowed visualization of silk glands at a higher resolution than previously described (Mullen, 1969), enabling identification of some novel cell types and zones. A summary of the results is presented in **Table 4**.

The major and minor ampullate glands exhibited similar morphology (with tail, sac, funnel and a three-limbed duct) differing primarily in size (**Figure 7**). Both gland types displayed three distinct epithelial zones (A, B, and C) within their tail and sac. While these zones were previously described for the major ampullate glands (Andersson et al., 2013), their presence in the minor ampullate glands is a novel finding. The secretion in the lumen near the epithelium of each zone of both glands stained similarly to the granules in the corresponding zones by H&E, although distinguishing zone A and B secretions in minor ampullate glands was challenging due to their similar

staining. These findings laid the foundation for the analysis of the respective glands in papers II and III.

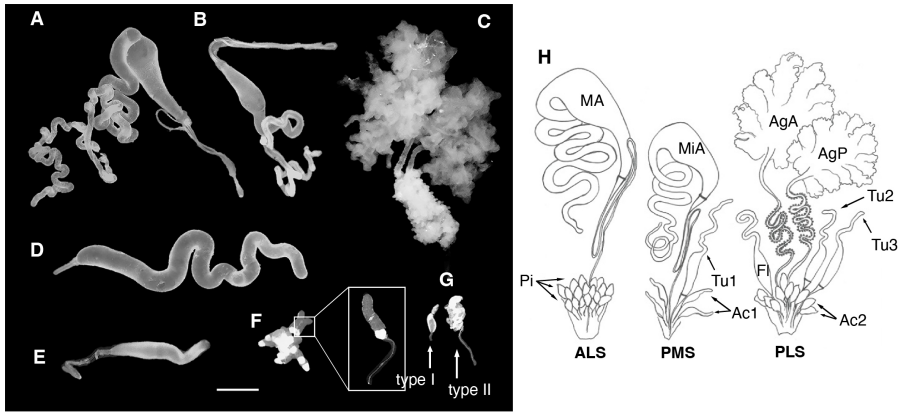

Figure 6: Silk glands of L. sclopetarius female spider. (A-G) Photos of the seven silk gland types: major ampullate (A), minor ampullate (B), aggregate (C), tubuliform (D), flagelliform (E), piriform (F), and aciniform (G) glands. (H) Schematic diagram of silk glands and their corresponding spinnerets (ALS, anterior lateral spinneret; PMS, posterior median spinneret; PLS, posterior lateral spinneret). Scale bar = 1 mm. Abbreviations: MA, major ampullate; MiA, minor ampullate; Pi, piriform; Tu, tubuliform; Fl, flagelliform; Ac1, aciniform type I; Ac2, aciniform type II; AgA, anterior aggregate; AgP, posterior aggregate. Note: Only one set of glands is shown for clarity, except for tubuliform glands. Minor ampullate and flagelliform tails were partially lost during dissection. Figure modified with permission from Scientific Reports (Sonavane et al., 2023).

Flagelliform glands, referred to as median ampullate glands by Mullen (Mullen, 1969), resembled the ampullate glands but lacked a funnel and had smaller tails. The gland epithelium was comprised of two distinct zones (A and B). Aggregate glands were characterized by branched bodies with convoluted ducts. The gland epithelium harbored two short columnar epithelial cell types (I and II), identified for the first time, which were distinct but were not confined to specific zones. The ducts were surrounded by a layer of cuboidal epithelial cells, which was further surrounded by large nodules consisting of a group of cells. Tubuliform glands were long and tubular and connected to their short ducts with a funnel. The gland epithelium consisted of only one cell type, but its morphology varied in individual spiders, as described previously (Herrera et al., 2015; Moon, 1998). Two

types of aciniform glands were identified: type I, occurring individually, and type II, found in groups. The aciniform and piriform glands were pear-shaped, except that the type I aciniform glands had a small tail proximally. In line with previous reports (Mullen, 1969), all three gland types were composed of two columnar epithelial zones, A and B.

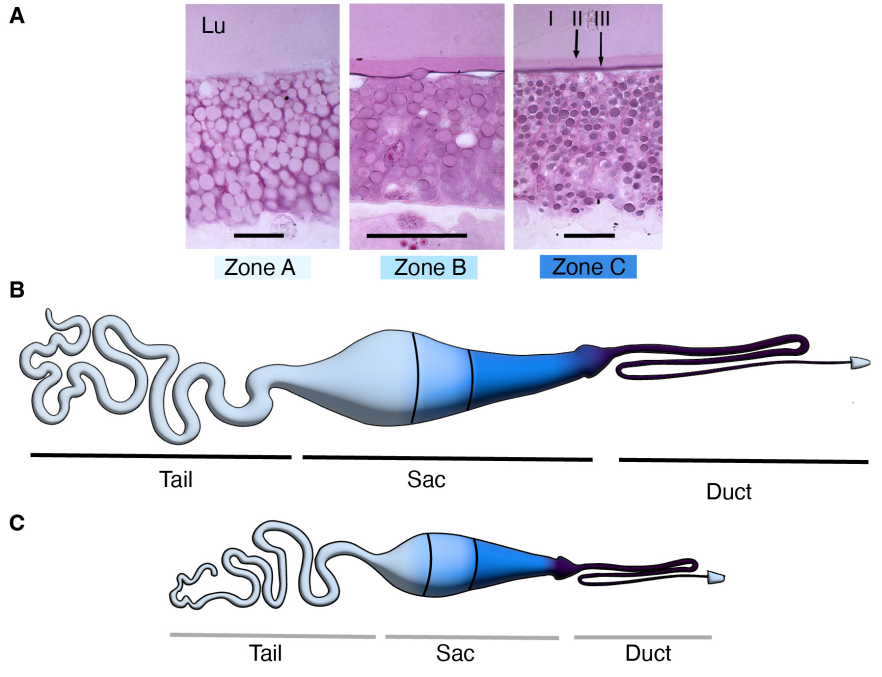

Figure 7: Ampullate glands of *L. sclopetarius*. (A) Morphology of cells identified in the three zones of major and minor ampullate glands. The secretion from each zone forms a layer in the lumen (Lu), indicated as I, II and II in the last panel. Scale bar = 20 µm. (B–C) Schematic illustrations of major (B) and minor (C) ampullate glands. The three anatomical parts (tail, sac & duct) and the three epithelial zones in the tail and sac (zone A, B, & C) are indicated.

As mentioned earlier, the pH gradient established by CA within the major ampullate gland, ranging from 7.6 in the sac to < 5.7 in the duct, is critical for spidroin assembly and fiber formation (Andersson et al., 2014). Given that the spidroin terminal domains, which are affected by environmental pH, are pivotal for fiber formation (Andersson et al., 2014; Askarieh et al., 2010; Hagn et al., 2010; Kronqvist et al., 2014; Landreh et al., 2010; Schwarze et

al., 2013) and are present in all spidroin types (Arakawa et al., 2022; Babb et al., 2022; Babb et al., 2017; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021), we hypothesized that CA is present in all other silk glands which could mean that there are pH gradients also in these glands. Using a histochemical method on resin sections, we detected CA activity in all silk glands, except type I aciniform and tubuliform glands, with the strongest staining in aggregate glands and their duct nodules.

Although it remains to be determined if the other silk glands also have pH gradients, the presence of CA activity suggests that this could be the case and that the polymerization mechanisms are similar to those in major ampullate glands. Despite that the aggregate glands produce a glue-like substance and not a fiber, they displayed the strongest CA activity. The active CA in these glands could perform other physiological functions as these are versatile enzymes (Supuran, 2018) but may also regulate pH as in the major ampullate gland since the AgSps also have NT and CT (Babb et al., 2022; Babb et al., 2017; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021). The lack of CA activity in the tubuliform glands could be a result of that the gland was inactive at the time of sacrifice. The dual function of aciniform silk in prey wrapping and soft inner casing of the egg cocoons (Kovoor, 1987), combined with the presence of two aciniform gland types, suggests the possibility of different silk types produced by each gland. However, this aspect remains unclear and needs further investigation. If different silk types are indeed produced by the two glands, it is plausible that aciniform type I glands produce silk for the cocoon's inner lining. If the glands are inactive, it could explain the absence of CA activity, similar to the tubuliform glands.

We detected positive periodic acid Schiff (PAS) staining in all glands except for type I aciniform glands, suggesting that carbohydrates and/or glycoproteins are components of silk fibers.

Table 4: Summary of results obtained from paper I. ALS- anterior lateral spinneret, PMS- posterior median spinneret, and PLS- posterior lateral spinneret. The CA staining is shown as weak (+), intermediate (++), strong (+++) or no staining (-).

Gland	Anatomical features	Regionalization	CA Activity
Major ampullate	• One pair, connected to ALS • Long tail, bulged sac, funnel, S-shaped duct • Two secondary ampullate glands	Three zones in the sac with sharp borders- • Zone A: tail and proximal third of the sac • Zone B: central region of the sac • Zone C: distal end of the sac (connects to duct)	+ zone C cells, funnel region, cuticle of the duct
Minor ampullate	• One pair, connected to PMS • Long tail, bulged sac, funnel, S-shaped duct • Smaller in size than major ampullate	3 zones similar to major ampullate glands	+++ funnel and cuticle of the duct
Flagelliform	• One pair, connected to PMS • Short tail, narrow sac, long duct	Two zones with sharp borders • Zone A: tail and about half of the sac • Zone B: distal part of the gland (connects to the duct)	+ Basal layer of the gland +++ small region in beginning of zone B and secretion near these cells
Tubuliform	• Three pairs (One pair connected to PMS, two others connected to PLS) • Long and tubular body, funnel, thin and short ducts	• One zone • Slightly varying cell morphology based on reproductive stage	- No staining

Aciniform	Type 1 • Many glands occur singly, connect to PMS • Small sac, small tail	Two zones with sharp borders • Zone A – tail • Zone B – distal part (connects to duct)	- No staining
	Type 2 • Many (more in number than type 1), present in clusters, connected to PLS • Small and pear-shaped	Two zones with sharp borders. • Zone A: proximal end • Zone B: distal end (connects to duct)	++ big granules in zone B cells, secretion near these cells ++ cuticle of the duct +++ small granules in zone B cells
Aggregate	• Two pairs (anterior & posterior), connected to PLS • Branched body, spacious lumen, wide & convoluted ducts with numerous nodules	• Two types of cells (type A and B) with sharp borders • Duct is surrounded by single layer of cuboidal epithelium that is further surrounded by nodules • Nodules have big cells with irregular nuclei & no borders	(+ to +++) Variable staining in the duct +++ cytoplasm and basal membrane of nodules +++ basal cell membrane of the gland and small region at the distal end
Piriform	• Many glands, present as cluster, connected to ALS • Pear-shaped	Two zones with sharp borders • Zone A: proximal end • Zone B: distal end (connects to duct)	+ zone B and secretion near zone B

4.2 Paper II

Building upon the detailed anatomical and morphological characterization of silk glands in *L. sclopetarius* presented in Paper I, this study aimed to address some additional questions regarding biology of the major ampullate glands. Specifically, we elucidated the cellular composition and established a transcriptional map of the major ampullate gland, and associated these with the structure and composition of major ampullate silk fibers.

To accomplish this, we needed a high-quality genome with 3' untranslated regions (UTRs) annotated. Therefore, we first performed a *de-novo* assembly of *L. sclopetarius* genome sequenced using a combination of short- (10X genomics) and long-read (PacBio) sequencing methods. This approach yielded a high-quality genome assembly (Table 1, paper II), comprising 1602 contigs and spanning 2.27 Gb with a GC content of 30.5%. The completeness of the genome assembly, assessed using Benchmarking Universal Single Copy Orthologs (BUSCO) for the Arachnida_odb10 lineage, was determined to be 97.3% and no potential contamination was found in the final assembly.

The assembled genome was annotated using a combination of methods to identify protein-coding genes, repetitive elements, and non-coding RNA. The annotation quality was improved by utilizing bulk transcriptomic data from the following tissues isolated from female spiders: major ampullate, minor ampullate, flagelliform, aggregate, and piriform/aciniform, abdomen, and head. The annotation process yielded 22,860 protein-coding genes, with 78% (17,886 genes) assigned functional annotations and 59% (13,524 genes) found in the UniProt database. A BUSCO analysis using the Arachnida_odb10 lineage indicated a gene set completeness of 96.8%.

Given the important role of spidroins, we comprehensively curated these genes. Employing a combined approach of homology-based searches for NT and CT, identification of repetitive motifs, and manual inspection of the transcriptome, we identified 35 complete spidroin genes, higher than previously reported (Babb et al., 2022; Babb et al., 2017; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono et al., 2019; Kono, Ohtoshi, et al., 2021). All identified spidroin genes exhibited canonical features, including signal peptides, the N-terminal domain, and repetitive regions. Although three spidroins lacked the CT domain, manual inspection revealed stop codons in

all spidroins and no signs of sequencing errors, indicating that our spidroin catalogue contains exclusively full-length genes. Based on homology to known spidroins, phylogenetic analyses of terminal domains, and distinct repetitive motifs, we classified the spidroins into distinct types. Notably, at least one paralogue from each of the seven spidroin orthologous groups was identified. Specifically, 12 MaSps were identified: MaSp1a–c, MaSp2a–f, MaSp3a,b, and MaSp4. Additionally, two proteins lacking CT were identified and named ampullate-like (AmSp-like1 and 2), based on clustering of their NT domain with other ampullate spidroins, and similar expression profiles to these. RNA-seq analysis revealed tissue-specific expression patterns of the different spidroins. Most spidroin types were highly expressed in their corresponding glands, with some exhibiting broader expression, in line with previous reports (Babb et al., 2017; Chaw et al., 2021; Clarke et al., 2017).

Next, we identified the constituent proteins of major ampullate silk using LC-MS/MS. By only considering the proteins found both in major ampullate silk fibers and isolated major ampullate glands, we minimized potential contaminants from other glands/fibers. Assuming that the most abundant proteins contribute most to the properties of the fiber, we considered only proteins with peptide counts > 0.1 % of total spectra. This analysis revealed 18 proteins as the predominant constituents in the major ampullate silk, referred to as "the 18 silk proteins," and the corresponding genes as "the 18 silk genes". Ten of these were from MaSp1, MaSp2, MaSp3 and MaSp4 groups, collectively constituting 96% of fiber's protein content. These findings align with the previous reports of dominance of MaSps in major ampullate silk (Chaw et al., 2015; Hu et al., 2023; Kono, Nakamura, et al., 2021; Kono, Ohtoshi, et al., 2021; Larracas et al., 2016), although the relative abundance MaSp subtypes can vary across species (Kono, Nakamura, et al., 2021; Kono et al., 2019). In addition to MaSps, AmSp-like1 & 2 and six proteins designated as SpiCE-LMa1 to SpiCE-LMa6, were also identified as constituents of the fiber. Despite their common naming as SpiCE-LMa, they had diverse properties in terms of molecular weight, predicted tertiary structure, and amino acid composition.

To find out which cell types the silk proteins are expressed in and the spatial distribution of these, we integrated bulk RNA sequencing, scRNAseq and spatial transcriptomic analyses. Bulk RNA sequencing of the whole gland revealed that the 18 silk genes were among the most highly expressed

transcripts in the gland (Table 2, paper II). To localize gene expression within the gland, we cut the gland into tail, sac, and duct regions and sequenced the bulk RNA. Consequently, tail samples primarily contained zone A transcripts, while sac samples contained transcripts from all three zones (**Figure 7**). Using principal component analysis (PCA) of enriched genes in the gland, a major ampullate gene set was identified, which was further used for partial least squares discriminant analysis (PLS-DA) on tail, sac, and duct samples. This analysis clearly distinguished the transcriptomic profiles of these regions. Overlaying the 18 silk genes onto the PLS-DA plot revealed that all MaSp genes were expressed in the tail, except MaSp3a and MaSp3b, which were expressed in the sac along with AmSp-like1 and AmSp-like2. Additionally, SpiCE-LMa1, SpiCE-LMa2, and SpiCE-LMa4 were expressed in the sac. Hierarchical clustering of expression profiles of the 18 genes revealed six clusters. The MaSp2 genes were almost exclusively expressed in the tail (cluster 1) and could be assigned to zone A. Genes predominantly expressed in the sac (clusters 5 and 6) were assigned to the sac. Genes in clusters 2, 3, and 4 exhibited expression in both tail and sac. However, since the sac samples contained all three zones, precise localization for these genes' expression was challenging.

To improve the resolution of silk gene expression, we used spatial transcriptomics (10X Visium) on six abdominal sections. UMAP analysis of manually annotated spots effectively clustered silk gland spots, distinguishing them from other tissues. Within the silk gland cluster, spots from different glands grouped accordingly and the spidroin expression was specifically localized to corresponding glands, consistent with bulk RNA-seq findings. The major ampullate spots were annotated further as zone A, B or C and the UMAP analysis revealed distinct expression profiles for each zone. Expression patterns of the 18 silk genes were visualized as a heatmap, with 15 genes showing differential expression across zones (Table 2, paper II). MaSp1, MaSp2, and SpiCE-LMa6 were predominantly expressed in zone A, MaSp3, AmSp-like1, and SpiCE-LMa4 in zone B, and SpiCE-LMa1 in zone C. SpiCE-LMa2 and AmSp-like2 were expressed in both zones B and C.

Even though MaSp1 and MaSp2 genes were both expressed in zone A, they showed different levels of expression in different zone A spots, suggesting the potential presence of multiple cell types in zone A. To investigate this further, we used scRNA-seq which identified eight cell types

within the major ampullate gland, with three cell types specific to the tail, three to the sac, and two to the duct. Comparing these data with spatial transcriptomics allowed the assignment of six cell types to specific zones, named based on their location and marker gene expression: three in zone A (ZoneA_MaSp1, ZoneA_MaSp2, ZoneA_SpiCE-LMa), one in zone B (ZoneB_MaSp3), one in zone C (ZoneC_SpiCE-LMa), and one distributed across distal part of zone A, zones B and C (ZoneABC). Notably, all 18 silk genes were among the top 100 marker genes for at least one cell type, with most genes specifically expressed in a single cell type (Table 2, paper II). The expression of the 18 silk genes was significantly higher in five of the eight cell types identified.

To determine the spatial distribution of cell types within the gland, we integrated scRNA-seq and spatial transcriptomic data. Additionally, image analysis was used on H&E-stained spatial sections to determine perimeters and area of all tail cross-sections. A correlation was observed between cross-sectional area and gene expression, revealing that ZoneA_MaSp2 cells were predominantly located in the proximal tail region, while ZoneA_MaSp1 cells were more abundant distally. ZoneA_SpiCE-LMa cells were found in all parts of the tail. Collectively, our findings demonstrated that the production of major ampullate silk proteins is confined to five distinct cell types with specific spatial localization within the tail and sac regions of the gland.

To verify the zone-specific origin of the secretions as suggested previously (Andersson et al., 2013), we used digital image analysis which revealed a correlation between the H&E staining pattern of the vesicles in the epithelium and the layers of the silk dope. This indicated that the secretion from zone A forms the bulk of the secretion in the lumen, zone B secretion forms a middle layer, and the zone C forms the outermost layer.

To investigate if the three layers persist in the major ampullate silk fiber, we subjected major ampullate silk fibers to solubilization using three different urea concentrations (2M, 4M, and 8M) (Vollrath et al., 1996). Lower urea concentrations (2M) will primarily solubilize the outer layer while the higher concentration (8M) will also dissolve the whole fiber. LC-MS/MS analysis of the solubilized fractions revealed differential protein enrichment in the different urea extracts. Proteins enriched in the 2M fraction (SpiCE-LMa1-6, AmSp-like1, AmSp-like2) aligned with proteins expressed in zones B and C, while proteins enriched in the 8M fraction (MaSp1 and MaSp2) corresponded to zone A proteins. MaSp3 proteins, primarily

expressed in zone B, were most abundant in the 4M fraction. To validate these findings and exclude potential solubility-related artifacts, we repeated the solubilization experiment on formic acid-dissolved and dried silk fibers. The obtained protein enrichment patterns supported the presence of distinct protein layers within the intact silk fiber.

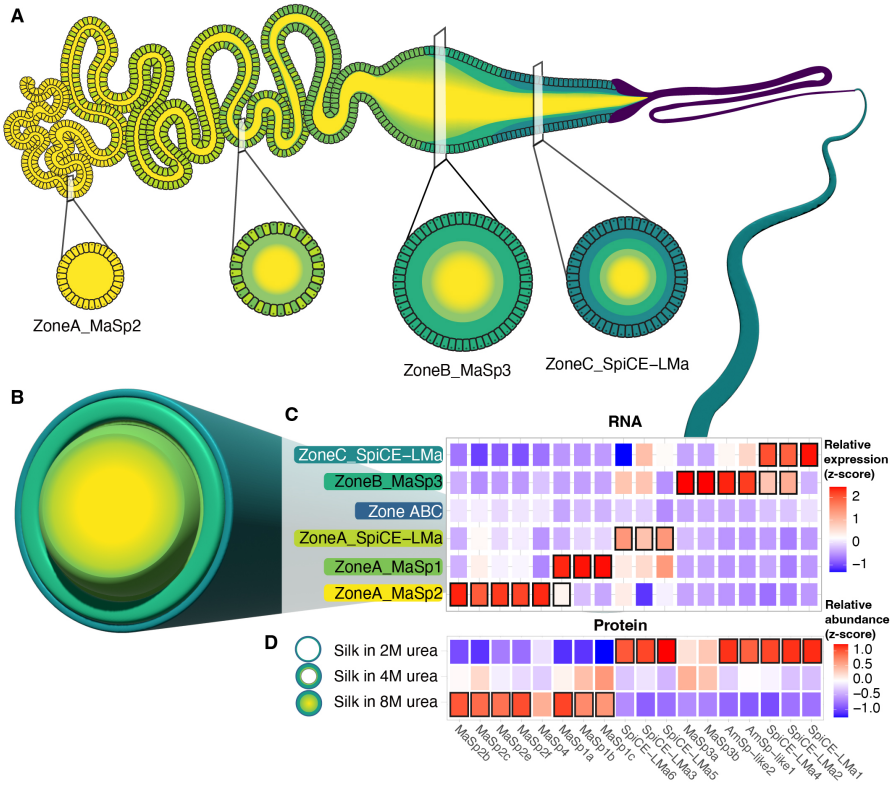

Figure 8: Proposed model with structure and composition for the major ampullate gland and silk. (A) Schematic of the major ampullate gland indicating the spatial distribution of different cell types. (B) Schematic of the three-layered major ampullate fiber. (C) Heatmap shows the expression of 18 silk genes in the different cell types, black boxes indicate marker genes. (D) Heatmap shows the protein levels in three urea extracts (2M, 4M, & 8M) of the major ampullate fiber. Black boxes indicate proteins identified as significantly enriched (p-value <0.05).

These results, combined with spatial gene expression data, strongly suggest that the major ampullate silk fiber comprises three layers with distinct protein compositions (**Figure 8**). The inner core is enriched in

MaSp1, MaSp2, and MaSp4 proteins produced by zone A cells, while the middle layer predominantly contains MaSp3 and AmSp-like proteins from zone B cells. The outer layer is composed of non-spidroin proteins, including SpiCE-LMa proteins, expressed in zone C cells.

4.3 Paper III

This paper aimed to characterize the morphology and gene expression profile of the *L. sclopetarius* minor ampullate gland, as well as the protein composition of its silk. A comparative analysis with the major ampullate gland was conducted to identify similarities and differences between the two ampullate glands.

Spacers in MiSps

While MiSpc lacked a spacer, the presence of spacers in MiSpa, MiSpb, and MiSpd aligns with previous findings in other MiSps (Babb et al., 2022; Babb et al., 2017; Chen et al., 2012; Colgin & Lewis, 1998; Kono et al., 2019; Yang et al., 2023). Notably, the sequence and predicted fold of the spacers shows impressive conservation with corresponding spacer domain from *T. antipodiana* with an experimentally determined structure (Yang et al., 2023) (suppl. Fig. 2 & 3 in paper III). Incorporation of spacers in recombinant spidroins has shown to improve the mechanical properties of the resulting fibers (Qi et al., 2024). It is intriguing that despite this, the minor ampullate silk is less strong compared to the major ampullate silk. Another interesting observation was that the repetitive motifs before and after the spacers differed in MiSpa, MiSpb and MiSpd, a feature not described in literature before. A detailed analysis of all the reported MiSps is necessary to check if this trait is common to MiSps having spacers, and if there are MiSps without spacers in order to conclude if there are two different classes of spidroins, one with spacers and the other lacking spacers as discussed earlier.

Using bulk RNA-seq analysis, we identified 20 differentially expressed genes (DEGs) in the gland. All four MiSps were among the top ten DEGs. This supports the reasoning that the bulk of minor ampullate silk is composed of MiSps. Although a detailed characterization of the proteins encoded by the 20 DEGs is needed to determine their potential functional roles, preliminary analysis provided some insights. Hierarchical clustering of the 20 DEGs based on their expression profiles revealed three non-spidroin

DEGs (sp-1339.1_3, LASC010012, and LASC007099) that clustered with the MiSps. Amino acid composition analysis revealed that sp-1339.1_3 and LASC007099 have high Gly content, and identified two additional DEGs, LASC001730 and LASC000038, with high Ala content. These proteins lack terminal domains, indicating they are not spidroins, but all four have signal peptides, suggesting they could be components of the fiber. Only LASC001730 had homologs in other spiders, but the identified homologs were uncharacterized proteins with no known function.

Model for the minor ampullate glands and silk

Using scRNA-seq for the first time on the minor ampullate glands, we identified five cell types in this gland. Three methods were employed to identify the cell types: (1) comparing the scRNA-seq data with minor ampullate bulk RNA-seq data from glands roughly cut into the different zones (proximal and distal zone A, zone B and zone C) and duct; (2) comparing the marker genes with cell types identified in major ampullate gland (paper II); and (3) integration of scRNA-seq data with spatial transcriptomics data (generated from paper II) from both minor and major ampullate glands. With this combined strategy, two cell types were assigned to the duct (Duct1 and Duct2) and the other three to each epithelial zones (identified from paper I). The cell types were named based on their identified location and top marker genes as: ZoneA_MiSpabd, ZoneB_MiSpc, and ZoneC_SpiCE.

For step (3) mentioned above, it is important to note that the limited number of spatial transcriptomics spots covering the minor ampullate gland, coupled with low RNA counts in these spots, hindered a definitive assignment of cell types to specific zones based on minor ampullate gland spots alone. However, the striking morphological similarities between major and minor ampullate glands, along with identification of shared marker genes, provided a compelling rationale for extrapolating cell type localization within the minor ampullate gland.

As mentioned before, the results from paper I revealed a layered secretion in minor ampullate lumen. To further verify this, we performed image analysis on H&E-stained image of minor ampullate gland (paper I). While we couldn't distinguish the staining of vesicles from zone A and zone B, the vesicles in zone C had distinct staining, suggesting that at least two layers are formed from zone A/B and C secretions into the gland lumen. These results combined with previous reports of multilayered architecture in both

ampullate silks (Iachina et al., 2023; Sponner et al., 2007), and the striking similarities in morphology and transcriptional profiles between the major and minor ampullate glands, convince us that three layers are formed also in the minor ampullate silk, analogous to the major ampullate counterpart but with distinct protein compositions.

In our proposed model, fiber core is primarily composed of MiSpa, MiSpb, and MiSpd proteins secreted from zone A, while the middle layer likely contains all four MiSp proteins, but predominantly MiSpc secreted from zone B. The outermost layer is expected to be composed of non-spidroin proteins. Whether the non-spidroins can be classified as SpiCE-LMi (for minor ampullate silk) remains to be seen, as the SpiCE terminology is based on proteins identified in the fiber using proteomics. Likewise, proteomics studies are needed to confirm if the most highly expressed genes in each zone correspond to the most abundant proteins in the fiber. Several proteins identified in this study are promising candidates for further investigation into their role in fiber processing and properties.

Comparative analysis of two ampullate glands

Despite their striking morphological similarities, the presence of two distinct ampullate glands raises questions about their functional divergence. Both silk types share a common layered architecture, and the corresponding glands share many common genes, including the six SpiCE and the two AmSp-like proteins, that are among the 18 major ampullate spider silk genes. However, the silks have distinct mechanical properties and protein compositions. The minor ampullate fiber is more extensible but less strong, and it is more resistant to water compared to the major ampullate silk (Blackledge & Hayashi, 2006b; Guinea et al., 2012; Nakamura et al., 2023). These differences could originate from the different protein compositions, specifically from the corresponding layers containing spidroins since the bulk of the fiber is composed of MaSps (in major ampullate silk) and MiSps (in minor ampullate silk). However, the outer layer with non-spidroin proteins may also affect the fiber properties to some extent. Our analysis identified genes that are specifically expressed in minor ampullate zone C which should be studied further.

Another point worth discussing is limitations that comes with the current classification system used for the SpiCE proteins. Many of the SpiCE-LMa that in this paper was also found to be expressed in the minor ampullate gland could equally well be renamed as SpiCE-LMi (*Larinioides* minor ampullate

SpiCE) according to the current definition of SpiCE proteins. The numbering of these proteins in different species do not correspond to homologues which can be confusing and complicates inter-species comparisons. Furthermore, the SpiCEs are a very heterogeneous group of proteins, likely with diverse functions that are not reflected in their naming. Thus, nomenclature system leaves room for improvement, but understandably, it is not a trivial task to sort this terminology since we have only recently began unveiling these non-spidroin proteins in the fiber. However, increased understanding of the SpiCE proteins' functions and identification of SpiCE proteins in other spiders could lead to a more informative classification system, which should be a prioritized task for the community.

5. Conclusions and outlook

Following conclusions were drawn from the studies conducted in this thesis:

- All silk glands in *L. sclopetarius*, except the tubuliform gland, have at least two columnar epithelial cell types which are confined to specific zones in the glands (except in the aggregate glands in which the two cell types are dispersed). Both major and minor ampullate glands show similar epithelial zone organization and layered secretions. Active CA is present in all glands, except type I aciniform and tubuliform, suggesting potential pH gradients in most glands.
- Major ampullate silk is primarily composed of 18 silk proteins (belonging to MaSp1, MaSp2, MaSp3, MaSp4 and SpiCE-LMa classes) which are expressed and secreted from five cell types confined to the three epithelial zones (A, B, and C) in the gland.
- The secretions from the three major ampullate zones do not mix and form layers in the fiber. The inner layer is enriched in MaSp1, MaSp2, and MaSp4 proteins expressed and secreted from zone A cells. The middle layer is composed of MaSp3 proteins expressed in zone B cells. The outer layer consists of SpiCEs that are primarily expressed in zone C.
- The minor ampullate gland epithelium is composed of three cell types confined to the three glandular zones: ZoneA_MiSpabd, ZoneB_MiSpc, and ZoneC_SpiCE.
- The two ampullate glands share overall morphology and expression of several genes including six SpiCEs and AmSp-like genes, but there also significant differences.
- A model for minor ampullate silk is proposed with three layers, similar to the major ampullate silk but with distinct protein compositions: The innermost layer contains MiSpa, MiSpb, and MiSpc originating from

zone A cells, the middle layer contains primarily MiSpc but to some extent other MiSps, and the outer layer contains other non-spidroin proteins including LASC010012 and smoc1.

The findings from this thesis provide a deeper understanding of how silk is produced in *L. sclopetarius* at the cellular and molecular level and open avenues for future research in two key aspects as discussed below:

1) **Comprehension of spider silk biology**
   a) Characterization of important genes/proteins: We identified a number of proteins with unknown functions in the ampullate glands and silks, including SpiCE (paper II & III). These proteins are prime candidates for further investigation to obtain a more comprehensive understanding of the silk production process. Their functional characterization could involve a combination of methods, such as recombinant expression and protein purification followed by detailed biochemical and biophysical analyses, their secondary and tertiary structure determination, the effect of different pH on their structural conformations and stability, and if they can form oligomers or macromolecular complexes under specific conditions. The highly and differentially expressed genes in various cell types coding for non-secretory proteins are also worth investigating since they might play crucial regulatory roles. For instance, some proteins might function as chaperones, preventing spidroin aggregation. Investigating these proteins could provide valuable insights for the design of biomimetic systems and allow the replication of the complex environment of the native silk gland.
   b) Spidroin storage and layered structure: The layered secretion of proteins observed in multiple glands from our studies raise several intriguing questions that need detailed investigation – how are the spidroins stored in the lumen? What functions, if any, do the layers have? Could the presence of different MaSp subtypes confer distinct properties to the two inner layers of the major ampullate silk, for example? What is the importance of the outer layer that does not contain spidroins to any large extent? Furthermore, how would altering the composition of these layers affect the silk's properties? Recombinant production of these proteins followed by artificial

spinning methods offer a promising approach to explore some of these questions (see below). However, production of full-length spidroins in heterologous hosts still remains a challenge, hampering the investigation of biomimetic spinning of these large proteins. While knockdown or knockout of specific genes in spiders offer a potential approach for investigating the function of genes and may answer some of the questions, this complex task comes with challenges.

c) pH gradients in other glands: paper I revealed that active CA is present in most glands. A fundamental question for future research is whether pH gradients exist in other silk glands, and if so, to what extent they resemble those observed in the major ampullate glands. The smaller size of other glands, particularly the piriform and aciniform glands, presents a significant technical challenge for such a task. However, the use of ion-selective microelectrodes, as employed by Andersson et al. (Andersson et al., 2014) is a promising approach. Additionally, pH sensitive fluorescent probes or scanning ion conductive microscopy (SICM) could be other alternative methods for investigating pH in these glands.
d) Minor ampullate silk composition: proteomic studies of the minor ampullate silk fiber are needed to fully characterize its protein composition and confirm the presence of layers speculated in paper III.
e) Multi-omics approach for other silk glands: The successful application of multiomics approaches in this thesis for the major and minor ampullate glands paves the way for extending similar strategies to investigate other silk glands (flagelliform, aciniform, aggregate, piriform and tubuliform). Characterization of different silk glands and a comparative analysis between the glands will not only elucidate the distinctions between silk fibers and glues but also facilitate identification of novel cell types with specialized functions, identify genes shared across different silk glands, and potential regulating pathways essential for silk production in general. Understanding these common traits could inform the development of more generalized biomimetic strategies. These strategies could be applicable to a wider range of spider silks with diverse properties,

paving the way for the creation of biomimetic materials with a broader spectrum of potential applications.

2) Strategies for making biomimetic fibers

The knowledge from our findings could be used to develop new biomimetic fibers with improved and tunable mechanical properties.

a) Mimicking protein compositions: the presence of multiple spidroin types in the bulk of the silk fiber (MaSp 1/2/4 in major and MiSp a/b/d in minor) suggests that spiders may have adopted this strategy to produce fibers with superior mechanical properties. This concept can be mimicked by recombinantly expressing and purifying the MaSp variants and then combining them in concentrations and ratios reflecting those found in the gland to create a spinning dope which can subsequently be spun into fibers using aqueous buffers in order to fully mimic the spider's spinning system. Additionally, the SpiCE proteins can be included in the dope to study their effect on the fiber properties. This appraoch will also serve as a good model system for studying structure-function relationships

b) Mimicking layered architecture: The presence of layers in the ampullate silks suggests that layering is a potentially common trait among the fiber-forming silk types, including aciniform type II and piriform silks, where the cellular secretions form layers in the lumen (paper I). The layered fiber structure could be mimicked in the production of artificial fibers using for instance, microfluidic chips or co-axial devices. In such a setup, specific spidroins (MaSp 1/2/4 or and MiSp a/b/d), with subsequent layers containing other spidroins (MaSp3 or MiSpc) and potentially SpiCE and non-spidroins identified in this study could be mimicked.

References

Andersson, M., Chen, G., Otikovs, M., Landreh, M., Nordling, K., Kronqvist, N.,…Rising, A. (2014). Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biology*, *12*(8), 1-14. https://doi.org/10.1371/journal.pbio.1001921

Andersson, M., Holm, L., Ridderstrale, Y., Johansson, J., & Rising, A. (2013). Morphology and composition of the spider major ampullate gland and dragline silk. *Biomacromolecules*, *14*(8), 2945-2952. https://doi.org/10.1021/bm400898t

Andersson, M., Jia, Q., Abella, A., Lee, X. Y., Landreh, M., Purhonen, P.,…Rising, A. (2017). Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nature Chemical Biology*, *13*(3), 262-264. https://doi.org/10.1038/nchembio.2269

Arakawa, K., Kono, N., Malay, A. D., Tateishi, A., Ifuku, N., Masunaga, H.,…Numata, K. (2022). 1000 spider silkomes: Linking sequences to silk physical properties. *Sci Adv*, *8*(41). https://doi.org/10.1126/sciadv.abo6043

Arguelles, J., Baker, R. H., Perez-Rigueiro, J., Guinea, G. V., Elices, M., & Hayashi, C. Y. (2023). Relating spidroin motif prevalence and periodicity to the mechanical properties of major ampullate spider silks. *J Comp Physiol B*, *193*(1), 25-36. https://doi.org/10.1007/s00360-022-01464-3

Askarieh, G., Hedhammar, M., Nordling, K., Saenz, A., Casals, C., Rising, A.,…Knight, S. D. (2010). Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature*, *465*(7295), 236-238. https://doi.org/10.1038/nature08962

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*(1), 12-15. https://doi.org/10.1002/sca.4950220103

Ayoub, N. A., Garb, J. E., Kuelbs, A., & Hayashi, C. Y. (2013). Ancient properties of spider silks revealed by the complete gene sequence of the prey-wrapping silk protein (AcSp1). *Molecular Biology and Evolution*, *30*(3), 589-601. https://doi.org/10.1093/molbev/mss254

Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A., & Hayashi, C. Y. (2007). Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One*, *2*(6), e514. https://doi.org/10.1371/journal.pone.0000514

Babb, P. L., Gregoric, M., Lahens, N. F., Nicholson, D. N., Hayashi, C. Y., Higgins, L.,…Voight, B. F. (2022). Characterization of the genome and silk-gland transcriptomes of Darwin's bark spider (Caerostris darwini). *PLoS One*, *17*(6), e0268660. https://doi.org/10.1371/journal.pone.0268660

Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., Nicholson, D. N., Kim, E. J., Hogenesch, J. B.,…Voight, B. F. (2017). The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nature Genetics*, *49*(6), 895-903. https://doi.org/10.1038/ng.3852

Bakhshandeh, B., Nateghi, S. S., Gazani, M. M., Dehghani, Z., & Mohammadzadeh, F. (2021). A review on advances in the applications of spider silk in biomedical issues. *Int J Biol Macromol*, *192*, 258-271. https://doi.org/10.1016/j.ijbiomac.2021.09.201

Bergmann, F., Stadlmayr, S., Millesi, F., Zeitlinger, M., Naghilou, A., & Radtke, C. (2022). The properties of native dragline silk and its biomedical applications. *Biomaterials Advances*, *140*. https://doi.org/10.1016/j.bioadv.2022.213089

Blackledge, T. A., & Hayashi, C. Y. (2006a). Silken toolkits: biomechanics of silk fibers spun by the orb web spider
(Fabricius 1775). *Journal of Experimental Biology*, *209*(13), 2452-2461. https://doi.org/10.1242/jeb.02275

Blackledge, T. A., & Hayashi, C. Y. (2006b). Silken toolkits: biomechanics of silk fibers spun by the orb web spider (Fabricius 1775). *Journal of Experimental Biology*, *209*(13), 2452-2461. https://doi.org/10.1242/jeb.02275

Bohlin, J., & Pettersson, J. H. O. (2019). Evolution of Genomic Base Composition: From Single Cell Microbes to Multicellular Animals. *Computational and Structural Biotechnology Journal*, *17*, 362-370. https://doi.org/10.1016/j.csbj.2019.03.001

Bourzac, K. (2015). Spiders: Web of intrigue. *Nature*, *519*(7544), S4-S6. https://doi.org/10.1038/519S4a

Boutry, C., & Blackledge, T. A. (2010). Evolution of supercontraction in spider silk: structure-function relationship from tarantulas to orb-weavers. *Journal of Experimental Biology*, *213*(20), 3505-3514. https://doi.org/10.1242/jeb.046110

Challis, R. J., Goodacre, S. L., & Hewitt, G. M. (2006). Evolution of spider silks: Conservation and diversification of the C-terminus. *Insect Molecular Biology*, *15*(1), 45-56. https://doi.org/10.1111/j.1365-2583.2005.00606.x

Chaw, R. C., Clarke, T. H., 3rd, Arensburger, P., Ayoub, N. A., & Hayashi, C. Y. (2021). Gene expression profiling reveals candidate genes for defining spider silk gland types. *Insect Biochem Mol Biol*, *135*, 103594. https://doi.org/10.1016/j.ibmb.2021.103594

Chaw, R. C., Correa-Garhwal, S. M., Clarke, T. H., Ayoub, N. A., & Hayashi, C. Y. (2015). Proteomic Evidence for Components of Spider Silk Synthesis from Black Widow Silk Glands and Fibers. *Journal of proteome research*, *14*(10), 4223-4231. https://doi.org/10.1021/acs.jproteome.5b00353

Chaw, R. C., Saski, C. A., & Hayashi, C. Y. (2017). Complete gene sequence of spider attachment silk protein (PySp1) reveals novel linker regions and extreme repeat homogenization. *Insect Biochemistry and Molecular Biology*, *81*, 80-90. https://doi.org/10.1016/j.ibmb.2017.01.002

Chen, G., Liu, X., Zhang, Y., Lin, S., Yang, Z., Johansson, J.,…Meng, Q. (2012). Full-length minor ampullate spidroin gene sequence. *PLoS One*, *7*(12), e52293. https://doi.org/10.1371/journal.pone.0052293

Chen, X., Knight, D. P., & Vollrath, F. (2002). Rheological characterization of spidroin solution. *Biomacromolecules*, *3*(4), 644-648. https://doi.org/10.1021/bm0156126

Clarke, T. H., Garb, J. E., Haney, R. A., Chaw, R. C., Hayashi, C. Y., & Ayoub, N. A. (2017). Evolutionary shifts in gene expression decoupled from gene duplication across functionally distinct spider silk glands. *Sci Rep*, *7*(1), 8393. https://doi.org/10.1038/s41598-017-07388-1

Colgin, M. A., & Lewis, R. V. (1998). Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions". *Protein Science*, *7*(3), 667-672. https://doi.org/DOI 10.1002/pro.5560070315

Collin, M. A., Clarke, T. H., 3rd, Ayoub, N. A., & Hayashi, C. Y. (2018). Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. *International Journal of Biological Macromolecules*, *113*, 829-840. https://doi.org/10.1016/j.ijbiomac.2018.02.032

Cranford, S. W., Tarakanova, A., Pugno, N. M., & Buehler, M. J. (2012). Nonlinear material behaviour of spider silk yields robust webs. *Nature*, *482*(7383), 72-U91. https://doi.org/10.1038/nature10739

Davenport, H. W. (1939). Gastric carbonic anhydrase. *The Journal of physiology*, *97*(1), 32-43. https://doi.org/10.1113/jphysiol.1939.sp003790

De Oliveira, D. H., Gowda, V., Sparrman, T., Gustafsson, L., Pires, R. S., Riekel, C.,…Hedhammar, M. (2024). Structural conversion of the spidroin C-terminal domain during assembly of spider silk fibers. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-49111-5

Dicko, C., Knight, D., Kenney, J. M., & Vollrath, F. (2004). Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. *Biomacromolecules*, *5*(6), 2105-2115. https://doi.org/10.1021/bm034486y

Du, N., Xiang, Y. L., Narayanan, J., Li, L., Lim, M. L. M., & Li, D. (2006). Design of superior spider silk: From nanostructure to mechanical properties. *Biophysical Journal*, *91*(12), 4528-4535. https://doi.org/10.1529/biophysj.106.089144

Eisoldt, L., Thamm, C., & Scheibel, T. (2012). The role of terminal domains during storage and assembly of spider silk proteins. *Biopolymers*, *97*(6), 355-361.

Elghetany, M. T., & Saleem, A. (1988). Methods for staining amyloid in tissues: a review. *Stain technology*, *63*(4), 201-212. https://doi.org/10.3109/10520298809107185

Exler, J. H., Hümmerich, D., & Scheibel, T. (2007). The amphiphilic properties of spider silks are important for spinning. *Angewandte Chemie-International Edition*, *46*(19), 3559-3562. https://doi.org/10.1002/anie.200604718

Fan, Z., Yuan, T., Liu, P., Wang, L. Y., Jin, J. F., Zhang, F., & Zhang, Z. S. (2021). A chromosome-level genome of the spider Trichonephila antipodiana reveals the genetic basis of its polyphagy and evidence of an ancient whole-genome duplication event. *Gigascience*, *10*(3), 1-15. https://doi.org/10.1093/gigascience/giab016

Foelix, R. (1996). Biology of Spiders: Oxford University Press. *New York*, *330*.

Fraternali, F., Stehling, N., Amendola, A., Anrango, B. A. T., Holland, C., & Rodenburg, C. (2020). Tensegrity Modelling and the High Toughness of Spider Dragline Silk. *Nanomaterials*, *10*(8). https://doi.org/10.3390/nano10081510

Frische, S., Maunsbach, A. B., & Vollrath, F. (1998). Elongate cavities and skin-core structure in Nephila spider silk observed by electron microscopy. *Journal of Microscopy*, *189*(1), 64-70. https://doi.org/10.1046/j.1365-2818.1998.00285.x

Gaines, W. A., Sehorn, M. G., & Marcotte, W. R. (2010). Spidroin N-terminal Domain Promotes a pH-dependent Association of Silk Proteins during Self-assembly. *Journal of Biological Chemistry*, *285*(52), 40745-40753. https://doi.org/10.1074/jbc.M110.163121

Gao, Z., Lin, Z., Huang, W., Lai, C. C., Fan, J. S., & Yang, D. (2013). Structural Characterization of Minor Ampullate Spidroin Domains and Their Distinct Roles in Fibroin Solubility and Fiber Formation. *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056142

Garb, J. E., Ayoub, N. A., & Hayashi, C. Y. (2010). Untangling spider silk evolution with spidroin terminal domains. *BMC Evolutionary Biology*, *10*(1), 243. https://doi.org/10.1186/1471-2148-10-243

Garb, J. E., Haney, R. A., Schwager, E. E., Gregoric, M., Kuntner, M., Agnarsson, I., & Blackledge, T. A. (2019). The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness. *Communications Biology*, *2*(1), 275. https://doi.org/10.1038/s42003-019-0496-1

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*(Pt 23), 3295-3303.

Greco, G., Wolff, J. O., & Pugno, N. M. (2020). Strong and tough silk for resilient attachment discs: the mechanical properties of piriform silk in the spider Cupiennius salei (Keyserling, 1877). *Frontiers in Materials*, *7*, 138. https://doi.org/https://doi.org/10.3389/fmats.2020.00138

Grubb, D. T., & Ji, G. D. (1999). Molecular chain orientation in supercontracted and re-extended spider silk. *International Journal of Biological Macromolecules*, *24*(2-3), 203-210. https://doi.org/10.1016/S0141-8130(98)00086-5

Guan, J., Vollrath, F., & Porter, D. (2011). Two Mechanisms for Supercontraction in Spider Dragline Silk. *Biomacromolecules*, *12*(11), 4030-4035. https://doi.org/10.1021/bm201032v

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*(5258), 112-115. https://doi.org/10.1126/science.272.5258.112

Guinea, G. V., Elices, M., Plaza, G. R., Perea, G. B., Daza, R., Riekel, C.,…Pérez-Rigueiro, J. (2012). Minor Ampullate Silks from Nephila and Argiope Spiders: Tensile Properties and Microstructural Characterization. *Biomacromolecules*, *13*(7), 2087-2098. https://doi.org/10.1021/bm3004644

Hagn, F., Eisoldt, L., Hardy, J. G., Vendrely, C., Coles, M., Scheibel, T., & Kessler, H. (2010). A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature*, *465*(7295), 239-242. https://doi.org/10.1038/nature08936

Hayashi, C. Y., & Lewis, R. V. (1998a). Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Journal of Molecular Biology*, *275*(5), 773-784. https://doi.org/10.1006/jmbi.1997.1478

Hayashi, C. Y., & Lewis, R. V. (1998b). Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Journal of Molecular Biology*, *275*(5), 773-784. https://doi.org/10.1006/jmbi.1997.1478

Hayashi, C. Y., & Lewis, R. V. (2000). Molecular architecture and evolution of a modular spider silk protein gene. *Science*, *287*(5457), 1477-1479. https://doi.org/10.1126/science.287.5457.1477

Hayashi, C. Y., & Lewis, R. V. (2001). Spider flagelliform silk: Lessons in protein design, gene structure, and molecular evolution. *BioEssays*, *23*(8), 750-756. https://doi.org/10.1002/bies.1105

Hayashi, C. Y., Shipley, N. H., & Lewis, R. V. (1999). Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules*, *24*(1-3), 271-275. https://doi.org/10.1016/S0141-8130(98)00089-0

Hennecke, K., Redeker, J., Kuhbier, J. W., Strauss, S., Allmeling, C., Kasper, C.,…Vogt, P. M. (2013). Bundles of Spider Silk, Braided into Sutures, Resist Basic Cyclic Tests: Potential Use for Flexor Tendon Repair. *Plos One*, *8*(4). https://doi.org/10.1371/journal.pone.0061100

Herrera, E., Nguyen, L. T., Escobar, E., Ouriel, W., & Casem, M. L. (2015). Alteration of tubuliform silk gland cytoarchitecture with the reproductive cycle of the Western black widow spider, Latrodectus hesperus. *Invertebrate Biology*, *134*(4), 332-340. https://doi.org/10.1111/ivb.12106

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. *Biophysical journal*, *71*(6), 3442-3447. https://doi.org/10.1016/S0006-3495(96)79539-5

Holland, C., Numata, K., Rnjak-Kovacina, J., & Seib, F. P. (2019). The Biomedical Use of Silk: Past, Present, Future. *Advanced Healthcare Materials*, *8*(1). https://doi.org/10.1002/adhm.201800465

Holland, G. P., Creager, M. S., Jenkins, J. E., Lewis, R. V., & Yarger, J. L. (2008). Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. *Journal of the American Chemical Society*, *130*(30), 9871-9877. https://doi.org/10.1021/ja8021208

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*(3), 834-839. https://doi.org/10.1021/bm0343904

Hu, W., Jia, A., Ma, S., Zhang, G., Wei, Z., Lu, F.,…Wang, Y. (2023). A molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk. *Nat Commun*, *14*(1), 837. https://doi.org/10.1038/s41467-023-36545-6

Iachina, I., Fiutowski, J., Rubahn, H. G., Vollrath, F., & Brewer, J. R. (2023). Nanoscale imaging of major and minor ampullate silk from the orb-web spider Nephila Madagascariensis. *Sci Rep*, *13*(1), 6695. https://doi.org/10.1038/s41598-023-33839-z

Jenkins, J. E., Creager, M. S., Butler, E. B., Lewis, R. V., Yarger, J. L., & Holland, G. P. (2010). Solid-state NMR evidence for elastin-like $β$-turn structure in spider dragline silk. *Chemical communications*, *46*(36), 6714-6716.

Jenkins, J. E., Creager, M. S., Butler, E. B., Lewis, R. V., Yarger, J. L., & Holland, G. P. (2010). Solid-state NMR evidence for elastin-like β-turn structure in spider dragline silk. *Chemical Communications*, *46*(36), 6714-6716. https://doi.org/10.1039/c0cc00829j

Jin, H. J., & Kaplan, D. L. (2003). Mechanism of silk processing in insects and spiders. *Nature*, *424*(6952), 1057-1061. https://doi.org/10.1038/nature01809

Jorge, I., Ruiz, V., Lavado-Garcia, J., Vazquez, J., Hayashi, C., Rojo, F. J.,…Perez-Rigueiro, J. (2022). Expression of spidroin proteins in the silk glands of golden orb-weaver spiders. *J Exp Zool B Mol Dev Evol*, *338*(4), 241-253. https://doi.org/10.1002/jez.b.23117

Kelenyi, G. (1967). Thioflavin S fluorescent and Congo red anisotropic stainings in the histologic demonstration of amyloid. *Acta neuropathologica*, *7*(4), 336-348.

Kerkam, K., Viney, C., Kaplan, D., & Lombardi, S. (1991). Liquid Crystallinity of Natural Silk Secretions. *Nature*, *349*(6310), 596-598. https://doi.org/10.1038/349596a0

Keten, S., Xu, Z., Ihle, B., & Buehler, M. J. (2010). Nanoconfinement controls stiffness, strength and mechanical toughness of B-sheet crystals in silk. *Nature Materials*, *9*(4), 359-367. https://doi.org/10.1038/nmat2704

Knight, D. P., & Vollrath, F. (1999). Liquid crystals and flow elongation in a spider's silk production line. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1418), 519-523.

Knight, D. P., & Vollrath, F. (2001). Changes in element composition along the spinning duct in a Nephila spider. *Naturwissenschaften*, *88*(4), 179-182. https://doi.org/10.1007/s001140100220

Kono, N., Nakamura, H., Mori, M., Yoshida, Y., Ohtoshi, R., Malay, A. D.,…Arakawa, K. (2021). Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proc Natl Acad Sci U S A*, *118*(31), 1-10. https://doi.org/10.1073/pnas.2107065118

Kono, N., Nakamura, H., Ohtoshi, R., Moran, D. A. P., Shinohara, A., Yoshida, Y.,…Arakawa, K. (2019). Orb-weaving spider Araneus ventricosus genome elucidates the spidroin gene catalogue. *Scientific reports*, *9*(1), 8380. https://doi.org/10.1038/s41598-019-44775-2

Kono, N., Ohtoshi, R., Malay, A. D., Mori, M., Masunaga, H., Yoshida, Y.,…Arakawa, K. (2021). Darwin's bark spider shares a spidroin repertoire with Caerostris extrusa but achieves extraordinary silk toughness through gene expression. *Open Biol*, *11*(12), 210242. https://doi.org/10.1098/rsob.210242

Kornfeld, T., Nessler, J., Helmer, C., Hannemann, R., Waldmann, K. H., Peck, C. T.,…Radtke, C. (2021). Spider silk nerve graft promotes axonal regeneration on long distance nerve defect in a sheep model. *Biomaterials*, *271*, 120692-120692.

Kovoor, J. (1987). Comparative Structure and Histochemistry of Silk-Producing Organs in Arachnids. In *Ecophysiology of Spiders* (pp. 160-186). https://doi.org/10.1007/978-3-642-71552-5_12

Kovoor, J., & Zylberberg, L. (1980). Fine structural aspects of silk secretion in a spider (Araneus diadematus). I. Elaboration in the pyriform glands. *Tissue Cell*, *12*(3), 547-556. https://doi.org/10.1016/0040-8166(80)90044-0

Kronqvist, N., Otikovs, M., Chmyrov, V., Chen, G., Andersson, M., Nordling, K.,…Johansson, J. (2014). Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nat Commun*, *5*, 3254. https://doi.org/10.1038/ncomms4254

Kronqvist, N., Sarr, M., Lindqvist, A., Nordling, K., Otikovs, M., Venturi, L.,…Johansson, J. (2017). Efficient protein production inspired by how spiders make silk. *Nature Communications*, *8*(May). https://doi.org/10.1038/ncomms15504

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2024). Next-Generation Sequencing and Emerging Technologies. *Seminars in Thrombosis and Hemostasis*. https://doi.org/10.1055/s-0044-1786397

Kümmerlen, J., van Beek, 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*(8), 2920-2928. https://doi.org/DOI 10.1021/ma951098i

Landreh, M., Askarieh, G., Nordling, K., Hedhammar, M., Rising, A., Casals, C.,…Bergman, T. (2010). A pH-Dependent Dimer Lock in Spider Silk Protein. *Journal of Molecular Biology*, *404*(2), 328-336. https://doi.org/10.1016/j.jmb.2010.09.054

Lane, A. K., Hayashi, C. Y., Whitworth, G. B., & Ayoub, N. A. (2013). Complex gene expression in the dragline silk producing glands of the Western black widow (Latrodectus hesperus). *BMC genomics*, *14*(1), 1-12.

Larracas, C., Hekman, R., Dyrness, S., Arata, A., Williams, C., Crawford, T., & Vierra, C. A. (2016). Comprehensive Proteomic Analysis of Spider Dragline Silk from Black Widows: A Recipe to Build Synthetic Silk Fibers. *Int J Mol Sci*, *17*(9), 1-16. https://doi.org/10.3390/ijms17091537

Leal-Egana, A., & Scheibel, T. (2010). Silk-based materials for biomedical applications. *Biotechnol Appl Biochem*, *55*(3), 155-167. https://doi.org/10.1042/BA20090229

Lefèvre, T., Boudreault, S., Cloutier, C., & Pézolet, M. (2011). Diversity of molecular transformations involved in the formation of spider silks. *Journal of Molecular Biology*, *405*(1), 238-253. https://doi.org/10.1016/j.jmb.2010.10.052

Leppert, A., Chen, G. F., Lama, D., Sahin, C., Railaite, V., Shilkova, O.,…Landreh, M. (2023). Liquid-Liquid Phase Separation Primes Spider Silk Proteins for Fiber Formation via a Conditional Sticker Domain. *Nano Letters*, *23*(12), 5836-5841. https://doi.org/10.1021/acs.nanolett.3c00773

Li, S. F., McGhie, A. J., & Tang, S. L. (1994). New internal structure of spider dragline silk revealed by atomic force microscopy. *Biophysical Journal*, *66*(4), 1209-1212. https://doi.org/10.1016/S0006-3495(94)80903-8

Lin, T. Y., Masunaga, H., Sato, R., Malay, A. D., Toyooka, K., Hikima, T., & Numata, K. (2017). Liquid Crystalline Granules Align in a Hierarchical Structure To Produce Spider Dragline Microfibrils. *Biomacromolecules*, *18*(4), 1350-1355. https://doi.org/10.1021/acs.biomac.7b00086

Liu, Y., Shao, Z. Z., & Vollrath, F. (2005). Relationships between supercontraction and mechanical properties of spider silk. *Nature Materials*, *4*(12), 901-905. https://doi.org/10.1038/nmat1534

Liu, Y., Sponner, A., Porter, D., & Vollrath, F. (2008). Proline and processing of spider silks. *Biomacromolecules*, *9*(1), 116-121. https://doi.org/10.1021/bm700877g

MacIntosh, A. C., Kearns, V. R., Crawford, A., & Hatton, P. V. (2008). Skeletal tissue engineering using silk biomaterials. *J Tissue Eng Regen Med*, *2*(2-3), 71-80. https://doi.org/10.1002/term.68

Malay, A. D., Suzuki, T., Katashima, T., Kono, N., Arakawa, K., & Numata, K. (2020). Spider silk self-assembly via modular liquid-liquid phase separation and nanofibrillation. *Science Advances*, *6*(45). https://doi.org/10.1126/sciadv.abb6030

Miller, L. D., Putthanarat, S., Eby, R. K., & Adams, W. W. (1999). Investigation of the nanofibrillar morphology in silk fibers by small angle X-ray scattering and atomic force microscopy. *International Journal of Biological Macromolecules*, *24*(2-3), 159-165. https://doi.org/10.1016/S0141-8130(99)00024-0

Millesi, F., Weiss, T., Mann, A., Haertinger, M., Semmler, L., Supper, P.,…Radtke, C. (2021). Defining the regenerative effects of native spider silk fibers on primary Schwann cells, sensory neurons, and nerve-associated fibroblasts. *Faseb Journal*, *35*(2). https://doi.org/10.1096/fj.202001447R

Mohammadi, P., Sesilja Aranko, A., Landowski, C. P., Ikkala, O., Jaudzems, K., Wagermaier, W., & Linder, M. B. (2019). Biomimetic composites with enhanced toughening using silk-inspired triblock proteins and aligned nanocellulose reinforcements. *Science Advances*, *5*(9), 1-12. https://doi.org/10.1126/sciadv.aaw2541

Moon, M.-J. (1998). Changes of Tubuliform Silk Glands during the Cocoon Production in the Garden Spider, Argiope aurentia. *Applied Microscopy*, *28*(4), 539-549.

Moon, M.-J., & Kim, W.-K. (1989a). Ultrastructure of the Ampullate Gland in the Orb Web Spider, Nephila clavata L. Koch IV. Secretory Portion of the Small Ampullate Gland. *Applied Microscopy*, *19*(1), 59-69.

Moon, M.-J., & Kim, W.-K. (1989b). Ultrastructure of the Ampullate Glands in the Orb Web Spider, Nephila clav ata L. Koch III. Excretory Duct of the Small Ampullate Gland. *Applied Microscopy*, *19*(1), 49-58.

Moon, M. J. (2018). Fine structure of the aggregate silk nodules in the orb-web spider Nephila clavata. *Animal Cells and Systems*, *22*(6), 421-428. https://doi.org/10.1080/19768354.2018.1546227

Mullen, G. R. (1969). Morphology and Histology of the Silk Glands in Araneus sericatus Cl. *Transactions of the American Microscopical society*, *88*(2), 232-240. https://doi.org/10.2307/3224495

Nakamura, H., Kono, N., Mori, M., Masunaga, H., Numata, K., & Arakawa, K. (2023). Composition of Minor Ampullate Silk Makes Its Properties Different from Those of Major Ampullate Silk. *Biomacromolecules*, *24*(5), 2042-2051. https://doi.org/10.1021/acs.biomac.2c01474

Nocentini, A., Supuran, C. T., & Capasso, C. (2021). An overview on the recently discovered iota-carbonic anhydrases. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *36*(1), 1988-1995.

Nova, A., Keten, S., Pugno, N. M., Redaelli, A., & Buehler, M. J. (2010). Molecular and nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils. *Nano Letters*, *10*(7), 2626-2634. https://doi.org/10.1021/nl101341w

Otikovs, M., Chen, G. F., Nordling, K., Landreh, M., Meng, Q., Jörnvall, H.,…Jaudzems, K. (2015). Diversified Structural Basis of a Conserved Molecular Mechanism for pH-Dependent Dimerization in Spider Silk N-Terminal Domains. *Chembiochem*, *16*(12), 1720-1724. https://doi.org/10.1002/cbic.201500263

Papadopoulos, P., Ene, R., Weidner, I., & Kremer, F. (2009). Similarities in the Structural Organization of Major and Minor Ampullate Spider Silk. *Macromolecular Rapid Communications*, *30*(9-10), 851-857. https://doi.org/10.1002/marc.200900018

Parent, L. R., Onofrei, D., Xu, D., Stengel, D., Roehling, J. D., Addison, J. B.,…Holland, G. P. (2018). Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proceedings of the National Academy of*

Sciences of the United States of America, *115*(45), 11507-11512. https://doi.org/10.1073/pnas.1810203115

Parkhe, A. D., Seeley, S. K., Gardner, K., Thompson, L., & Lewis, R. V. (1997). Structural studies of spider silk proteins in the fiber. *Journal of Molecular Recognition*, *10*(1), 1-6. https://doi.org/10.1002/(Sici)1099-1352(199701/02)10:1<1::Aid-Jmr338>3.0.Co;2-7

Peakall, D. B. (1969). Synthesis of silk, mechanism and location. *American Zoologist*, *9*(1), 71-79. https://doi.org/10.1093/icb/9.1.71

Peters, H. M., & Kovoor, J. (1991). The silk-producing system of Linyphia triangularis (Araneae, Linyphiidae) and some comparisons with Araneidae - Structure, histochemistry and function. *Zoomorphology*, 1-17. https://doi.org/10.1007/BF01632706

Pham, T., Chuang, T., Lin, A., Joo, H., Tsai, J., Crawford, T.,…Vierra, C. (2014). Dragline silk: a fiber assembled with low-molecular-weight cysteine-rich proteins. *Biomacromolecules*, *15*(11), 4073-4081. https://doi.org/10.1021/bm5011239

Poza, P., Pérez-Rigueiro, J., Elices, M., & Llorca, J. (2002). Fractographic analysis of silkworm and spider silk. *Engineering Fracture Mechanics*, *69*(9), 1035-1048. https://doi.org/10.1016/S0013-7944(01)00120-5

Purkerson, J. M., & Schwartz, G. J. (2007). The role of carbonic anhydrases in renal physiology. *Kidney international*, *71*(2), 103-115. https://doi.org/10.1038/sj.ki.5002020

Qi, X. M., Wang, H., Wang, K. Z., Wang, Y., Leppert, A., Iashchishyn, I.,…Chen, G. F. (2024). Spiders Use Structural Conversion of Globular Amyloidogenic Domains to Make Strong Silk Fibers. *Advanced Functional Materials*. https://doi.org/10.1002/adfm.202315409

Ranade, V. (2024). Spider silk inspires a new route to organic magnets. *Mrs Bulletin*. https://doi.org/10.1557/s43577-024-00667-z

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics & Bioinformatics*, *13*(5), 278-289. https://doi.org/10.1016/j.gpb.2015.08.002

Riekel, C., Burghammer, M., Dane, T. G., Ferrero, C., & Rosenthal, M. (2017). Nanoscale Structural Features in Major Ampullate Spider Silk. *Biomacromolecules*, *18*(1), 231-241. https://doi.org/10.1021/acs.biomac.6b01537

Riekel, C., & Vollrath, F. (2001). Spider silk fibre extrusion: combined wide- and small-angle X-ray microdiffraction experiments. *Int J Biol Macromol*, *29*(3), 203-210. https://doi.org/10.1016/s0141-8130(01)00166-0

Rising, A. (2014). Controlled assembly: A prerequisite for the use of recombinant spider silk in regenerative medicine? *Acta Biomaterialia*, *10*(4), 1627-1631. https://doi.org/10.1016/j.actbio.2013.09.030

Rising, A., & Harrington, M. J. (2023). Biological Materials Processing: Time-Tested Tricks for Sustainable Fiber Fabrication. *Chem Rev*, *123*(5), 2155-2199. https://doi.org/10.1021/acs.chemrev.2c00465

Rising, A., Hjalm, G., Engstrom, W., & Johansson, J. (2006). N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. *Biomacromolecules*, *7*(11), 3120-3124. https://doi.org/10.1021/bm060693x

Rising, A., & Johansson, J. (2015). Toward spinning artificial spider silk. *Nat Chem Biol*, *11*(5), 309-315. https://doi.org/10.1038/nchembio.1789

Rodríguez, R., & Candelas, G. C. (1995). Flagelliform or coronata glands of Nephila clavipes. *Journal of Experimental Zoology*, *272*(4), 275-280. https://doi.org/10.1002/jez.1402720405

Salehi, S., Koeck, K., & Scheibel, T. (2020). Spider Silk for Tissue Engineering Applications. *Molecules*, *25*(3). https://doi.org/10.3390/molecules25030737

Sampath, S., Isdebski, T., Jenkins, J. E., Ayon, J. V., Henning, R. W., Orgel, J. P. R. O.,…Yarger, J. L. (2012). X-ray diffraction study of nanocrystalline and amorphous structure within major and minor ampullate dragline spider silks. *Soft Matter*, *8*(25), 6713-6722. https://doi.org/10.1039/c2sm25373a

Sampath, S., & Yarger, J. L. (2015). Structural hysteresis in dragline spider silks induced by supercontraction: an X-ray fiber micro-diffraction study. *Rsc Advances*, *5*(2), 1462-1473. https://doi.org/10.1039/c4ra13936d

Sanggaard, K. W., Bechsgaard, J. S., Fang, X., Duan, J., Dyrlund, T. F., Gupta, V.,…Wang, J. (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun*, *5*, 3765. https://doi.org/10.1038/ncomms4765

Sapede, D., Seydel, T., Forsyth, V. T., Koza, M. A., Schweins, R., Vollrath, F., & Riekel, C. (2005). Nanofibrillar structure and molecular mobility in spider dragline silk. *Macromolecules*, *38*(20), 8447-8453. https://doi.org/10.1021/ma0507995

Sarr, M., Kitoka, K., Walsh-White, K. A., Kaldmaee, M., Metlans, R., Tars, K.,…Kronqvist, N. (2022). The dimerization mechanism of the N-terminal domain of spider silk proteins is conserved despite extensive sequence divergence. *Journal of Biological Chemistry*, *298*(5). https://doi.org/10.1016/j.jbc.2022.101913

Schartau, W., & Leidescher, T. (1983). Composition of the Hemolymph of the Tarantula Eurypelma Californicum. *Journal of Comparative Physiology*, *152*(1), 73-77. https://doi.org/10.1007/Bf00689730

Schulz, S. (2001). Composition of the silk lipids of the spider Nephila clavipes. *Lipids*, *36*(6), 637-647. https://doi.org/10.1007/s11745-001-0768-7

Schwarze, S., Zwettler, F. U., Johnson, C. M., & Neuweiler, H. (2013). The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3815

Shao, Z., Hu, X. W., Frische, S., & Vollrath, F. (1999). Heterogeneous morphology of Nephila edulis spider silk and its significance for mechanical properties. *Polymer*, *40*(16), 4709-4711. https://doi.org/10.1016/S0032-3861(99)00072-5

Sheffer, M. M., Hoppe, A., Krehenwinkel, H., Uhl, G., Kuss, A. W., Jensen, L.,…Prost, S. (2021). Chromosome-level reference genome of the European wasp spider Argiope bruennichi: a resource for studies on range expansion and evolutionary adaptation. *Gigascience*, *10*(1), 148-148. https://doi.org/10.1093/gigascience/giaa148

Silva, L. P., & Rech, E. L. (2013). Unravelling the biodiversity of nanoscale signatures of spider silk fibres. *Nat Commun*, *4*, 3014. https://doi.org/10.1038/ncomms4014

Simmons, A., Michal, C. A., & Jelinski, L. W. (1996). Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science*, *271*(5245), 84-87. https://doi.org/10.1126/science.271.5245.84

Sonavane, S., Westermark, P., Rising, A., & Holm, L. (2023). Regionalization of cell types in silk glands of Larinioides sclopetarius suggest that spider silk fibers are complex layered structures. *Sci Rep*, *13*(1), 22273. https://doi.org/10.1038/s41598-023-49587-z

Sponner, A., Unger, E., Grosse, F., & Weisshart, K. (2005). Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat Mater*, *4*(10), 772-775. https://doi.org/10.1038/nmat1493

Sponner, A., Vater, W., Monajembashi, S., Unger, E., Grosse, F., & Weisshart, K. (2007). Composition and hierarchical organisation of a spider silk. *PLoS One*, *2*(10), e998. https://doi.org/10.1371/journal.pone.0000998

Stadlmayr, S., Peter, K., Millesi, F., Rad, A., Wolf, S., Mero, S.,…Radtke, C. (2024). Comparative Analysis of Various Spider Silks in Regard to Nerve Regeneration: Material Properties and Schwann Cell Response. *Advanced Healthcare Materials*, *13*(8). https://doi.org/10.1002/adhm.202302968

Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J.,…Frisen, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, *353*(6294), 78-82. https://doi.org/10.1126/science.aaf2403

Stehling, N., Abrams, K. J., Holland, C., & Rodenburg, C. (2019). Revealing Spider Silk's 3D Nanostructure Through Low Temperature Plasma Etching and Advanced Low-Voltage SEM. *Frontiers in Materials*, *5*. https://doi.org/10.3389/fmats.2018.00084

Stellwagen, S. D., & Burns, M. (2021). Repeat Variation Resolves a Complete Aggregate Silk Sequence of Bolas Spider *Mastophora phrynosoma*. *Integrative and Comparative Biology*, *61*(4), 1450-1458. https://doi.org/10.1093/icb/icab048

Stellwagen, S. D., & Renberg, R. L. (2019). Toward Spider Glue: Long Read Scaffolding for Extreme Length and Repetitious Silk Family Genes AgSp1 and AgSp2 with Insights into Functional Adaptation. *G3-Genes Genomes Genetics*, *9*(6), 1909-1919. https://doi.org/10.1534/g3.119.400065

Strauss, S., Diemer, M., Bucan, V., Kuhbier, J. W., Asendorf, T., Vogt, P. M., & Schlottmann, F. (2024). Spider silk enhanced tissue engineering of cartilage tissue: Approach of a novel bioreactor model using adipose derived stromal cells. *Journal of Applied Biomaterials & Functional Materials*, *22*. https://doi.org/10.1177/22808000241226656

Strickland, M., Tudorica, V., Rezac, M., Thomas, N. R., & Goodacre, S. L. (2018). Conservation of a pH-sensitive structure in the C-terminal region of spider silk extends across the entire silk gene family. *Heredity (Edinb)*, *120*(6), 574-580. https://doi.org/10.1038/s41437-018-0050-9

Supuran, C. T. (2018). Carbonic anhydrases and metabolism. In *Metabolites* (Vol. 8, pp. 25-25): Multidisciplinary Digital Publishing Institute.

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.

Proceedings of the National Academy of Sciences of the United States of America, *99*(16), 10266-10271. https://doi.org/10.1073/pnas.152162299

Vendrely, C., & Scheibel, T. (2007). Biotechnological production of spider-silk proteins enables new applications. *Macromolecular Bioscience*, *7*(4), 401-409. https://doi.org/10.1002/mabi.200600255

Vollrath, F., Barth, P., Basedow, A., Engstrom, W., & List, H. (2002). Local tolerance to spider silks and protein polymers in vivo. *In Vivo*, *16*(4), 229-234.

Vollrath, F., Holtet, T., Thøgersen, H. C., & Frische, S. (1996). Structural organisation of spider silk. *Proceedings of the Royal Society of London*, *263*(1367), 147-151. https://doi.org/10.1098/rspb.1996.0023

Vollrath, F., & Knight, D. P. (1999). Structure and function of the silk production pathway in the spider Nephila edulis. *Int J Biol Macromol*, *24*(2-3), 243-249. https://doi.org/10.1016/s0141-8130(98)00095-6

Vollrath, F., & Knight, D. P. (2001). Liquid crystalline spinning of spider silk. *Nature*, *410*(6828), 541-548. https://doi.org/10.1038/35069000

Wang, K., Wen, R., Wang, S., Tian, L., Xiao, J., & Meng, Q. (2020). The molecular structure of novel pyriform spidroin (PySp2) reveals extremely complex central repetitive region. *International Journal of Biological Macromolecules*, *145*, 437-444. https://doi.org/10.1016/j.ijbiomac.2019.12.027

Wang, Q. J., & Schniepp, H. C. (2018). Strength of Recluse Spider's Silk Originates from Nanofibrils. *Acs Macro Letters*, *7*(11), 1364-1370. https://doi.org/10.1021/acsmacrolett.8b00678

Wen, R., Liu, X., & Meng, Q. (2017). Characterization of full-length tubuliform spidroin gene from Araneus ventricosus. *International Journal of Biological Macromolecules*, *105*, 702-710. https://doi.org/10.1016/j.ijbiomac.2017.07.086

Wen, R., Wang, K. K., Liu, X. Q., Li, X., Mi, J. P., & Meng, Q. (2018a). Molecular cloning and analysis of the full-length aciniform spidroin gene from. *International Journal of Biological Macromolecules*, *117*, 1352-1360. https://doi.org/10.1016/j.ijbiomac.2017.12.090

Wen, R., Wang, K. K., Liu, X. Q., Li, X., Mi, J. P., & Meng, Q. (2018b). Molecular cloning and analysis of the full-length aciniform spidroin gene from *Araneus ventricosus*. *International Journal of Biological Macromolecules*, *117*, 1352-1360. https://doi.org/10.1016/j.ijbiomac.2017.12.090

Wen, R., Wang, S., Wang, K., Yang, D., Zan, X., & Meng, Q. (2022a). Complete gene sequence and mechanical property of the fourth type of major ampullate silk protein. *Acta Biomaterialia*, *155*, 282-291. https://doi.org/10.1016/j.actbio.2022.11.042

Wen, R., Wang, S., Wang, K., Yang, D., Zan, X., & Meng, Q. (2022b). Complete gene sequence and mechanical property of the fourth type of major ampullate silk protein Graphical abstract Abstract. *Acta Biomaterialia*. https://doi.org/10.1016/j.actbio.2022.11.042

Wen, R., Yang, D., Wang, K. K., & Zan, X. J. (2022). Characterization of two full-length major ampullate silk protein genes. *International Journal of Biological Macromolecules*, *213*, 297-304. https://doi.org/10.1016/j.ijbiomac.2022.05.178

Whaite, A. D., Wang, T., Macdonald, J., & Cummins, S. F. (2018). Major ampullate silk gland transcriptomes and fibre proteomes of the golden orb-weavers, Nephila plumipes and Nephila pilipes (Araneae: Nephilidae). *PLoS One*, *13*(10), e0204243. https://doi.org/10.1371/journal.pone.0204243

Wolff, J. O., Grawe, I., Wirth, M., Karstedt, A., & Gorb, S. N. (2015). Spider's super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. *Soft Matter*, *11*(12), 2394-2403. https://doi.org/10.1039/c4sm02130d

Work, R. W. (1977a). Dimensions, Birefringences, and Force-Elongation Behavior of Major and Minor Ampullate Silk Fibers from Orb-Web-Spinning Spiders - Effects of Wetting on These Properties. *Textile Research Journal*, *47*(10), 650-662. https://doi.org/10.1177/004051757704701003

Work, R. W. (1977b). Mechanisms of major ampullate silk fiber formation by orb-web-spinning spiders. *Transactions of the American Microscopical Society*, 170-189.

Xu, M., Jiang, Y. K., Pradhan, S., & Yadavalli, V. K. (2019). Use of Silk Proteins to Form Organic, Flexible, Degradable Biosensors for Metabolite Monitoring. *Frontiers in Materials*, *6*. https://doi.org/10.3389/fmats.2019.00331

Yang, Y., Chen, X., Shao, Z., Zhou, P., Porter, D., Knight, D. P., & Vollrath, F. (2005). Toughness of Spider Silk at High and Low Temperatures. *Advanced Materials*, *17*(1), 84-88. https://doi.org/10.1002/adma.200400344

Yang, Y. D., Gao, Z. W., & Yang, D. W. (2023). pH-dependent self-assembly mechanism of a single repetitive domain from a spider silk protein. *International Journal of Biological Macromolecules*, *242*. https://doi.org/10.1016/j.ijbiomac.2023.124775

Yang, Z., Liivak, O., Seidel, A., LaVerde, G., Zax, D. B., & Jelinski, L. W. (2000). Supercontraction and Backbone Dynamics in Spider Silk: 13C and 2H NMR Studies. *Journal of the American Chemical Society*, *122*(37), 9019–9025. https://doi.org/10.1021/ja0017099

Yazawa, K., Malay, A. D., Masunaga, H., & Numata, K. (2019). Role of Skin Layers on Mechanical Properties and Supercontraction of Spider Dragline Silk Fiber. *Macromolecular Bioscience*, *19*(3). https://doi.org/10.1002/mabi.201800220

Young, R. J., Holland, C., Shao, Z. Z., & Vollrath, F. (2021). Spinning conditions affect structure and properties of spider silk. *Mrs Bulletin*, *46*(10), 915-924. https://doi.org/10.1557/s43577-021-00194-1

Zhao, A.-C., Zhao, T.-F., Nakagaki, K., Zhang, Y.-S., SiMa, Y.-H., Miao, Y.-G.,…others. (2006). Novel molecular and mechanical properties of egg case silk from wasp spider, Argiope bruennichi. *Biochemistry*, *45*(10), 3348-3356.

Zylberberg, L., & Kovoor, J. (1979). Ultrastrueture du Canal des Glandes Agreges et Flagelliformes d'Araneus diadematus Clerek (Araneae, Araneidae) J. *Zoomorphologie*, *239*, 217-239.

Öksüz, K. E., Özkaya, N. K., Inan, Z. D. S., & Özer, A. (2021). Novel natural spider silk embedded electrospun nanofiber mats for wound healing. *Materials Today Communications*, *26*. https://doi.org/10.1016/j.mtcomm.2020.101942

Popular science summary

To date, over 52,000 unique spider species have been discovered and they are all capable of producing silk for various purposes. Some orb-weaving spiders can produce up to seven different silk types, each tailored for specific uses such as safety lines, prey capture, and building cocoons. The major ampullate silk, which is used for safety lines and frames of the web, is particularly impressive; it is tougher than many of the best synthetic materials, like Kevlar, and has therefore attracted a lot of attention over the years in the field of biomimicry. In the past decade, significant progress has been made in creating artificial silk by imitating how spiders spin it. However, these artificial fibers still don't match all the properties of natural spider silk. To produce artificial silk that truly replicates the natural silk, we probably need to closely mimic not just the spinning process but also the silk's composition and structure, which requires a comprehensive understanding of spider silk biology.

Silk is produced in glands located in the spider's abdomen and is pulled out through a structure called spinneret. The major ampullate gland that produces the major ampullate silk has a long tail that opens into a wide sac, which further connects to a narrow duct. The tail and sac consist of a single layer of epithelial cells organized into three distinct zones (A, B, and C) with sharp borders. Zone A is located in the tail and first part of the sac, zone B covers a small part in the middle of the sac, and zone C is closest to the duct. Spider silk proteins, known as spidroins, are produced and secreted by cells in zone A and B and they are stored in the sac at very high concentrations. The secretions from the three zones do not mix in the lumen (cavity) of the sac, but instead they form layers. When spiders spin silk, the proteins stored in the sac pass through the tapered duct. Here, factors like changes in pH, dependent on the enzyme carbonic anhydrase, ion composition, and

increased shear forces transform the soluble spidroins into a solid fiber. The composition of the secretions from the three zones are largely unknown and the relevance of the epithelial zones for the architecture of the fiber remains unclear. In addition, very little is known about the structure and function of other glands than the major ampullate. The overall aims of this thesis were to answer some of these outstanding questions.

In our approach to fill the knowledge gaps in the biology of spider silk production, we used Swedish bridge spiders (*Larinioides sclopetarius*) as a model in all the studies. The genetic information of an organism code for genes and proteins that regulate biological processes. Therefore, we sequenced the whole genome of *L. sclopetarius* and identified encoded spidroins representative of different silk classes. Next, we used a combination of sequencing techniques, which identified that the tail and sac of the major ampullate glands have six cell types that are spatially restricted to the three zones previously mentioned. We also revealed the specific combination of proteins produced by the different cell types.

Using proteomic techniques that detect essentially all proteins in a given sample, we defined the composition of the major ampullate silk fiber. Out of the more than 22000 proteins produced by *L. sclopetarius*, only 18 proteins form the bulk of the major ampullate fibers. These 18 proteins belong to six classes: major ampullate spidroin 1 (MaSp1), MaSp2, MaSp3, MaSp4, AmSp-like and spider silk constituting elements (SpiCE). The MaSp are specific to the major ampullate silk, AmSp-like are spidroins produced by two types of ampullate glands and SpiCE are a mixed group of proteins with unknown functions. These proteins were predominantly produced by specific cell types in the gland. To elaborate, MaSp2 and MaSp4 were produced by the cells found in the early part of the zone A (and hence labeled ZoneA_MaSp2 cells), MaSp1 were produced by cells found in the late part of the zone A (towards the sac, ZoneA_MaSp1), MaSp3 and AmSp-like were produced by cells restricted to zone B (ZoneB_MaSp3 cells) and finally, SpiCE proteins were produced by two cell types – one in zone A (ZoneA_SpiCE) and the other restricted to zone C (ZoneC_SpiCE).

Next, to study the architecture of the major ampullate silk fiber, we sequentially dissolved the silk. The results showed that the major ampullate silk fiber is made up of three layers – like a cylinder inside a cylinder inside a cylinder. The inner layer is rich in MaSp1, MaSp2 and MaSp4 proteins, the middle layer is rich in MaSp3 and AmSp-like proteins and the outermost

layer is rich in SpiCE proteins. Thus, the protein composition of each layer matched the combination of proteins produced by the specific cell types found in the glands. Next, by comparing the color stain intensity of the protein droplets in the cells of each zone to the corresponding layers found in the gland lumen, we gathered further evidence that the layers indeed are formed from secretions of cells in the different zones. Since we knew what proteins each cell type produces, we successfully identified the origin of the proteins found in each layer of the silk fiber.

In conclusion, the major ampullate glands are made of six cell types restricted to three zones that produce and secrete specific combinations of proteins. Their secretions do not mix but rather form three layers that persist in the final fiber. This work provides a detailed understanding of the biology of the major ampullate gland and the complex structure and composition of the major ampullate silk fiber. Our findings provide important knowledge for future efforts to spin biomimetic silk fibers.

In the next study, we extended similar methods to the minor ampullate gland. The minor ampullate glands are morphologically similar to the major ampullate glands with tail, sac and duct. Yet, these glands produce silk with comparable strength as the major ampullate silk, but they are more resistant to water than the major ampullate silk. If this silk type can be produced artificially, it could be used as a material for applications that require water resistance, for example textiles.

Using histological analysis, we revealed that the minor ampullate glands have a similar structure as the major ampullate glands with cell types restricted to three epithelial zones (A, B and C) and layered secretions in the gland. Using similar sequencing techniques used for the major ampullate glands, we found three cell types in the tail and sac of the glands which were restricted to the three zones. The cells in the zone A and B produced specific combinations of proteins, and specifically minor ampullate spidroins (MiSp) which are hallmarks of the minor ampullate glands. Zone C cells produced SpiCE proteins similarly to what was observed in major ampullate zone C cells. We performed a comparative analysis between the major and minor ampullate glands and revealed that many proteins are produced in corresponding zones in both glands, but also identified several proteins that are uniquely expressed in the minor ampullate gland and may have an impact on the properties of the silk fiber. Based on the presence of cell types restricted to specific zones, layered secretions in the gland lumen, and similar

protein profiles, we speculate that the minor ampullate silk also have a multi-layered structure, but with unique combination of proteins specific to the minor ampullate glands.

Apart from major and minor ampullate silks, three other silk fiber types are produced by spiders: flagelliform silk, which is used for the capturing spiral in the web; aciniform silk, which is used to wrap the captured prey; and tubuliform silk, which is used for building egg cocoons. The spiders also produce two other glue-like substances: aggregate silk which is used to coat the flagelliform so that the prey sticks to the web, and piriform silk which cements the junctions of the web and is used when spiders attach the safety line to a surface.

In a third study, we investigated the morphology of all silk glands of *L. sclopetarius* using histological methods. Staining the tissue sections with specific dyes revealed that most of the glands have multiple cell types that are restricted to specific zones. They also showed the presence of layers in the lumen as in the major and minor ampullate glands. In all glands, except one of the two types of aciniform and the tubuliform glands, we identified active carbonic anhydrase.

Populärvetenskaplig sammanfattning

Hittills har över 52 000 unika spindelarter upptäckts som alla producerar silke för olika ändamål. Vissa hjulnäts-spindlar kan producera upp till sju olika silkestyper, var och en skräddarsydd för specifika användningsområden, såsom säkerhetslinor, för infångande av bytesdjur, eller för äggsäckar. Flyktlinan (major ampullate silk), som används för säkerhetslinor och ramar till nätet, är särskilt imponerande; den är hållfastare än många syntetiska material, som Kevlar, och har rönt mycket uppmärksamhet inom biomimik. Under det senaste decenniet har betydande framsteg gjorts när det gäller att skapa konstgjord spindeltråd, genom att imitera hur spindlar spinner silke. Men dessa konstgjorda fibrer matchar fortfarande inte alla egenskaper hos naturligt spindelsilke. För att tillverka konstgjord tråd som verkligen liknar naturlig spindeltråd, måste vi nog efterlikna inte bara spinningsprocessen utan också trådens sammansättning och struktur, vilket kräver en omfattande förståelse av spindeltrådens biologi.

Spindeltråden produceras i körtlar i spindelns buk och dras ut genom en utförsgång. Den stora ampullkörteln som producerar flyktlinan har en lång svans som mynnar i en bred säck, som vidare ansluter till en smal kanal. Svansen och säcken består av ett enda lager av epitelceller organiserade i tre distinkta zoner (A, B och C) med skarpa övergångar. Zon A är belägen i svansen och första delen av säcken, zon B täcker en liten del i mitten av säcken, och zon C ligger närmast utförsgången. Spindeltrådsproteiner, spidroiner, produceras och utsöndras av celler i zon A och B och de lagras i säcken i mycket höga koncentrationer. Proteinerna som produceras av celler i de tre zonerna blandas inte när de utsöndras till säckens lumen (hålrum), utan istället bildar de lager. När spindlar spinner tråden passerar proteinerna som lagrats i säcken genom den avsmalnande kanalen. Här gör förändringar i pH, som kräver enzymet karbanhydras, och jonkoncentrationer, samt ökade

skjuvkrafter att de lösliga spidroinerna ändras och bildar en solid fiber. Sammansättningen av sekret från de tre zonerna är i stort sett okänd och sambandet mellan epitelzonerna och fiberns arkitektur är fortfarande oklar. Dessutom är väldigt lite känt om strukturen och funktionen hos andra körtlar än stora ampullkörteln. Det övergripande syftet med denna avhandling var att besvara några av dessa frågor.

För att fylla kunskapsluckorna i spindeltrådsproduktionens biologi använde vi svenska brospindlar (*Larinioides sclopetarius*) som modell i alla studierna. En organisms genetiska material kodar för gener och proteiner som reglerar biologiska processer. Därför sekvenserade vi först *L. sclopetarius* hela genom och identifierade gener som kodar för olika typer av spidroiner. Därefter använde vi en kombination av sekvenseringstekniker, vilket visade att svansen och säcken på de stora ampullkörtlarna har sex celltyper som är rumsligt begränsade till de tre tidigare nämnda zonerna. Vi bestämde också den specifika kombinationen av proteiner som produceras av de olika celltyperna.

Med hjälp av så kallad proteomik, som kan detektera i stort sett allla proteiner i ett givet prov, identifierade vi sammansättningen av flyktlinan. Av de mer än 22 000 proteiner som produceras av *L. sclopetarius* är det bara 18 som utgör huvuddelen av flyktlinan. Dessa 18 proteiner tillhör sex klasser: MaSp1, MaSp2, MaSp3, MaSp4, AmSp-like och SpiCE. MaSp är spidroiner som är specifika för flyktlinan, AmSp-liknande är spidroiner som produceras av två typer av ampullkörtlar och SpiCE är en blandad grupp av proteiner med okända funktioner. Dessa proteiner producerades huvudsakligen av specifika celltyper som finns i körteln. MaSp2 och MaSp4 producerades av celler som finns i den tidiga delen av zon A och döptes därför till ZoneA_MaSp2-celler, MaSp1 producerades av celler som hittades i den sena delen av zon A (mot säcken) vilka döptes till ZoneA_MaSp1, MaSp3 och AmSp-liknande producerades av celler begränsade till zon B (ZoneB_MaSp3-celler) och slutligen producerades SpiCE-proteiner av två celltyper - en i zon A (ZoneA_SpiCE) och den andra begränsad till zon C (ZoneC_SpiCE).

Därefter, för att studera flyktlinans arkitektur, löste vi upp tråden sekventiellt. Resultaten visade att flyktlinan består av tre lager – som en cylinder inuti en cylinder inuti en cylinder. Det inre lagret är rikt på MaSp1, MaSp2 och MaSp4, det mellersta lagret är rikt på MaSp3 och AmSp-liknande proteiner och det yttersta lagret är rikt på SpiCE-proteiner.

Proteinkombinationen i varje lager motsvarade den som produceras av de celltyper som återfinns i de tre zonerna (A, B och C) i körtlarna. Genom att jämföra färgintensiteten hos proteindroppar i cellerna med motsvarande skikt som finns i körtellumen, kunde vi stärka hypotesen att skikten verkligen bildas från utsöndring av celler i de olika zonerna. Eftersom vi vet vilka proteinkombinationer varje celltyp producerar, kunde vi identifiera ursprunget för proteinerna som finns i fiberns olika lager.

Sammanfattningsvis finns det sex celltyper i de stora ampullkörtlarna, vilka är begränsade till tre zoner, och som producerar och utsöndrar specifika kombinationer av proteiner. De utsöndrade substanserna blandas inte utan bildar snarare tre lager som kvarstår i den slutliga fibern. Detta arbete ger en detaljerad förståelse av den stora ampullkörtelns biologi och den komplexa strukturen och sammansättningen av flyktlinan. Våra resultat ger viktig kunskap för framtida ansträngningar att spinna biomimetisk spindeltråd.

I nästa studie utökade vi våra undersökningar till att omfatta en annan spindeltrådskörtel som kallas de mindre ampullkörtlarna (minor ampullate glands). Dessa liknar morfologiskt de stora ampullkörtlarna med svans, säck och kanal. Ändå producerar dessa körtlar en tråd med jämförbar styrka som flyktlinan men med högre motståndskraft mot vatten. Detta silke skulle kunna användas som ett material med tillämpningar som kräver vattenbeständighet, till exempel inom textilindustrin.

Med hjälp av histologisk analys fann vi att de mindre ampullkörtlarna har en liknande struktur som de stora ampullkörtlarna, består av celltyper begränsade till tre epitelzoner (A, B och C) och har skiktade utsöndringar i körteln. Med hjälp av liknande sekvenseringstekniker som används för de stora ampullkörtlarna, fann vi tre celltyper i tre zoner i svansen och säcken på körtlarna. Cellerna i zon A och B producerade specifika kombinationer av proteiner, specifikt mindre ampullspidroiner (MiSp) vilka är karakteristiska för de mindre ampullkörtlarna. Zon C-celler producerade SpiCE-proteiner på liknande sätt som observerades i stora ampullkörtelns zon C-celler. Vi gjorde en jämförande analys mellan de större och mindre ampullkörtlarna och identifierade många proteiner som producerades i motsvarande zoner, men också flera proteiner som var unika för de mindre ampullkörtlarna och som kan vara viktiga för trådens egenskaper. Baserat på närvaron av celltyper begränsade till specifika zoner, skiktade sekret i körtellumen och liknande proteinprofiler, spekulerar vi att tråden som de mindre

ampullkörtlarna producerar också har en flerskiktad struktur med unik kombination av proteiner.

Förutom silke från stora och mindre ampullkörtlar, produceras tre andra typer av fiber av spindlar: flagelliform silke, som används till spiralen i nätet; aciniform silke, som används för att linda in det fångade bytet; och tubuliform silke, som används för att konstruera äggsäckar. Spindlarna producerar även två andra limliknande ämnen: aggregatsilke som används för att belägga flagelliform silke så att bytet fastnar på nätet, och piriform silke som används när spindlar fäster säkerhetslinan på en yta.

I en tredje studie undersökte vi morfologin hos *L. sclopetarius* alla spindeltrådskörtlar med hjälp av histologiska metoder. Färgning av vävnadssnitt med specifika färgämnen möjliggjorde identifiering av flera celltyper i de flesta körtlar, vilka var begränsade till specifika zoner. De visade också närvaron av lager i lumen som i de stora och mindre ampullkörtlarna. I alla körtlar förutom aciniform och tubuliform körtlar identifierade vi aktivt karbanhydras.

Acknowledgements

The work described in this thesis was carried out at the Department of Animal Biosciences, SLU, Uppsala. It was generously funded by Formas, Morgongåva to Anna Rising, and the Faculty of Veterinary Medicine and Animal Science. I would like to acknowledge the bridge spiders that have been sacrificed for my research.

This thesis would not have been possible without support from many individuals. I am grateful to my supervisors for their invaluable guidance and support. **Anna**, you have been a great supervisor and I can't thank you enough for everything you've taught me! Your passion for research is truly inspirational and contagious. Thank you for always encouraging me to strive for excellence, for always being available to review my work and talk about science (even during weekends and holidays), for celebrating every small success, and for bringing over candies during my long working hours! Your unwavering support, especially during the demanding final phase of my PhD, has been invaluable.

Lena, I am grateful for your expertise, kindness and patience in guiding me (who started with absolutely no knowledge in histology!), and for providing fantastic photos of spiders (including the beautiful cover of this thesis). Your photographic skills are outstanding! Thanks for your wise words and for all the fun stories you shared during our fika sessions. **Janne,** thanks for remotely guiding me through my projects, and for always asking great questions. I admire your "think-outside-the-box" approach which has given me a unique perspective on things. **Marlene,** I wish we could have spent more time together. Your feedback and positivity have been truly motivating. **Johan**, there is no doubt that we make an awesome team! Thanks for always being patient and kind, including times when I asked silly

questions about coding or bioinformatics. It has been a blast working with you, and I will miss our never-ending “short” meetings ;)

I would like to thank all my collaborators and co-supervisors for their contributions and scientific discussions. My projects, especially paper II, would not have been possible without a great team effort from all authors.

I am grateful to have been part of an incredibly awesome group of people (past and present) in the Rising lab, especially those at SLU. **Viktoria, Tomás, Gabriele, Benjamin, Elin, Tobias, Emylin,** you all make the workplace so much more fun! Thanks for all the great conversations during lab meetings, lunch breaks, fika sessions. All the trips we had (both work-related and others) have been amazing and I am going to miss you guys a lot! Viktoria and Tomás, teaching the labs with you guys has been an exciting experience. **Urmi**, you have been a constant support during my entire PhD time. Those times with you in my early years trying to get the stubborn single cells from the glands and optimizing all experiments are some of my best memories! **Olga**, thanks to your support that makes everything in the group run smoothly. **Sameer**, thanks for mentoring me and for driving the bioinformatic project during the initial phase. I learned a lot from you. I extend my sincere gratitude to all members, past and present, of the AR and JJ group for their insightful contributions to our group meetings and for creating a wonderful team atmosphere during our group days, Christmas dinners, and potluck parties.

I want to express my gratitude to the department heads and the AFB study directors (past and present) who have guided me with official work and to AFB admin for resolving all my queries with a calm attitude, especially **Maria** – thanks a ton!

I would like to thank some people outside of AFB: **Aprajita** for all the fika sessions, for sharing my frustrations, and for encouraging me to keep a positive attitude. **Saritha** for always being there to listen to my problems and cheer me up when I was feeling down. Many past and present PhD students at VHC for sharing their experience and advice, and for the fika and game sessions.

My friends and family have been a great source of support and encouragement. **Sohan**, **Shibu,** and **Vlad** (& **Lili**)**,** thanks for all the entertaining Friday evenings, board game nights, parties, and long road trips, and, most of all, for always helping me with anything I needed. Sohan, thanks for helping me settle when I first moved to Sweden and for being an amazing

flatmate! **Mallika**, our long walks and your fun stories in the evenings helped me cope with my rough workdays. Thanks to **Gustav** and the **Jacobsson** family for hosting me during my first year and giving me so much love. Special thanks to **Juno** whose laughter fill my heart with joy. I am blessed to have many friends in Uppsala and Helsinki, especially **Vishal, Lamia, Nishant, Siddharth, Rajani, Gopi & family**, **Swagat**, **Diego, & Nelmary**, who have made this journey even more delightful.

I would like to express my deepest gratitude to my family. **Dad**, this achievement is as much yours as it is mine – thank you for everything! You are fondly missed. **Ma, Siddharth, Shweta & Sony**, thank you for your love and never-ending support. My three little sources of happiness and positivity, **Dhruva, Aditi & Chimmi**, thank you for being there; your smiles make me forget all my problems. My dear **Som**, I don't know what I would do without you! I am profoundly grateful for your endless love, support, and encouragement. You have kept me sane and strong through every step of this journey.

I

scientific reports

Check for updates

OPEN

Regionalization of cell types in silk glands of *Larinioides sclopetarius* suggest that spider silk fibers are complex layered structures

Sumalata Sonavane[1], Per Westermark[2], Anna Rising[1,3] & Lena Holm[1]✉

In order to produce artificial silk fibers with properties that match the native spider silk we likely need to closely mimic the spinning process as well as fiber architecture and composition. To increase our understanding of the structure and function of the different silk glands of the orb weaver *Larinioides sclopetarius*, we used resin sections for detailed morphology, paraffin embedded sections for a variety of different histological stainings, and a histochemical method for localization of carbonic anhydrase activity. Our results show that all silk glands, except the tubuliform glands, are composed of two or more columnar epithelial cell types, some of which have not been described previously. We observed distinct regionalization of the cell types indicating sequential addition of secretory products during silk formation. This means that the major ampullate, minor ampullate, aciniform type II, and piriform silk fibers most likely are layered and that each layer has a specific composition. Furthermore, a substance that stains positive for polysaccharides may be added to the silk in all glands except in the type I aciniform glands. Active carbonic anhydrase was found in all silk glands and/or ducts except in the type I aciniform and tubuliform glands, with the strongest staining in aggregate glands and their ductal nodules. Carbonic anhydrase plays an important role in the generation of a pH gradient in the major ampullate glands, and our results suggest that some other glands may also harbor pH gradients.

Spider silk fibers have unique mechanical and physical properties which make them interesting for applications in medicine and industry[1]. However, mass production of native silk using spiders is laborious and far from economically feasible since, unlike silkworms, individual spiders produce small amounts of silk. It is also challenging to breed spiders due to their territorial and cannibalistic nature. Alternative methods of producing silk by recombinant technologies offer promising solutions to these challenges and are being developed[2–4]. However, in order to truly mimic the pristine spider silk fiber, we need to develop better spinning methods, and we believe that the best way to do so is to characterize the spiders' spinning apparatus, which hopefully can give us clues for the design of novel biomimetic spinning devices as well as to increase our understanding of the composition and architecture of the silk fiber.

Swedish bridge spiders (*Larinioides sclopetarius*) were chosen for this study since they belong to the family Araneidae which means that the females can spin seven types of silk. The silks are produced in glands located in the opisthosoma and include the major ampullate, minor ampullate, flagelliform (or coronate), aggregate, piriform, tubuliform (or cylindrical), and aciniform glands. Each silk type has its characteristic structural, mechanical, and functional properties, and the silk glands that process them have distinct morphologies[5–12]. The major ampullate gland is the most well-studied silk gland and is composed of a tail where the main synthesis of the spidroins takes place, a wider sac where the proteins are stored, and an S-shaped duct where the soluble spidroins are converted into a fiber[13]. There are a number of studies on the histology and ultrastructural morphology of the major ampullate glands[5,11,14–19], and one of these have shown that the tail and the sac are made of three types of epithelial cells confined to three zones (A–C)[11]. Silk polymerization in the major ampullate glands occurs in response to changes in several factors such as decreased pH[20,21], changes in ion concentrations[22] as well as shear stress[23]. Several studies have suggested a skin–core architecture for the major ampullate silk[11,14,24–28], in which the core contains spider silk proteins (spidroins)[28,29] whereas the skin layer consists of carbohydrates, specifically

[1]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden. [2]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [3]Department of Biosciences and Nutrition, Karolinska Institutet, Neo, Huddinge, Sweden. ✉email: lena.holm@slu.se

glycogen/glycoproteins[24,28]. Others have further suggested a four-layer[30] and a five-layer[28] model for the major ampullate silk. The other silk glands have not been studied extensively, but a few investigations exist[5,15,31–38].

In the major ampullate gland, carbonic anhydrase (CA) is instrumental in upholding a pH gradient (from 7.6 to < 5.7 halfway through the duct)[21]. This pH gradient is one of the most important factors that regulate spidroin solubility and polymerization. CA is present in other silk-spinning species, e.g. the silkworm *Bombyx mori*, where it has been shown to have a similar function as in the spiders, i.e. generating protons that lower the pH along the silk gland[39]. CAs are a group of metalloenzymes that catalyze the reversible reaction –

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$$

These enzymes are ubiquitous in all kingdoms and eight genetically distinct CA families have been identified to date: α-, β-, γ-, δ-, ζ-, η-, θ-, and ι-CAs[40,41]. α-CAs are the most well-characterized and are present in several prokaryotic and eukaryotic organisms. CAs perform a variety of functions, i.e. in mammals CA is present in e.g., gastric mucosa where it is involved in acid secretion and kidney tubules where it regulates acid–base balance and water homeostasis[42–44].

This study aimed to identify all silk glands of *L. sclopetarius* and describe their morphological features using various methods. Resin-embedded tissue was used to provide good histological detail and a previously described histochemical method[45] was used to describe the localization of active CA in the glands. Thioflavin-S and Congo red were used to detect amyloid-like fibrils within silk glands, and periodic acid-Schiff (PAS) stain to detect the origin of carbohydrates. Since depletion of silk stored in the sac may result in morphological changes of the major ampullate gland[46], one group of spiders was stimulated to produce silk before sacrifice and the histological appearance of the major ampullate glands was compared.

Results

Morphology of silk glands

Seven types of silk glands were identified and could be isolated from the opisthosoma of *L. sclopetarius* (Fig. 1). Figure 2 shows a section of the opisthosoma stained with hematoxylin–eosin (HE) with the location of all the glands indicated.

Ampullate glands

Two primary major ampullate glands were located in the opisthosoma, close to the ventral wall. One pair of secondary major ampullate glands that are much smaller in size was attached to the duct of each major ampullate gland (Supplementary, SI Fig. 1). Two minor ampullate glands were located alongside the median plane of the opisthosoma and dorsal to the major glands. The ampullate glands were each composed of a narrow long tail, a wide bulged ampulla, and an S-shaped duct, however, the minor glands were smaller in size than the major glands (Fig. 1a and b). The sac and the duct of ampullate glands were bridged by a structure known as the funnel. The ducts of major glands were connected to the anterior lateral spinneret whereas those of the minor glands were connected to the posterior median spinnerets (Fig. 1h).

The tail and sac of ampullate glands were composed of three different zones as determined by the appearance of the epithelial cells by cell histology (Figs. 3a–c and 4a,b). The epithelium in the entire tail and the proximal third of the sac was identified as zone A, the central region of the sac as zone B, and the distal end of the sac until the funnel as zone C. The borders between the zones were sharp (Fig. 4b, Supplementary, SI Fig. 2). All zones were made of simple columnar tall epithelial cells with basally located nuclei. The cells of zone A contained big irregular granules that stained weakly with HE. In zone B, the cellular granules were somewhat smaller and stained lightly pink. In zone C, the granules were even smaller and stained intensely with HE. The secretory contents of the three zones stained similar to the granules of the corresponding cells. The secretions seemed to be sequentially added and formed separate layers in the lumen (Fig. 3a–c). The funnel was made of comb-like structures surrounding the lumen and was lined by a flat layer of cells on the basal side. The lumen of the duct was lined with a cuticle (shown by arrows in Fig. 3d) that originated from the funnel and continued till the spinnerets. The diameter of the duct decreased as it approached the spinnerets while the height of the surrounding epithelial cells increased, which was evident from a cross-section of the three limbs arranged in a *triad* fashion (Fig. 3e). Although the secondary major ampullate glands were seen macroscopically (Supplementary, SI Fig. 1), they could not be identified histologically. No histological differences were observed between the major ampullate glands of silked and unsilked spiders.

Aggregate glands

Two pairs of aggregate glands, anterior and posterior, were identified. The gland was branched (Fig. 1c,h) and spread across most of the anterior opisthosoma. The ducts were wide, partially convoluted and connected to the posterior lateral spinnerets. The convoluted part of the ducts was surrounded by numerous small nodules (Fig. 1c,h).

Aggregate glands were branched and characterized by a spacious lumen (Fig. 5a). The epithelium consisted of two types of cells (type A and B) (Fig. 5b), both simple columnar cells with basally located oval-shaped nuclei. Type A cells were tall and contained somewhat irregular granules that stained intensely with HE. Type B cells were slightly shorter, more abundant, and had heterogeneous granules. In addition to intensely staining granules similar to type A cells, they had two other types of granules that stained pink and light pink. The border between the two cell types was distinct, but they could not be confined to specific zones. Apically the cells were irregular and had bulging structures filled with numerous granules (Fig. 5b). The ducts of aggregate glands were composed of a cuticular intima that was surrounded by a single layer of cuboidal epithelial cells which was further

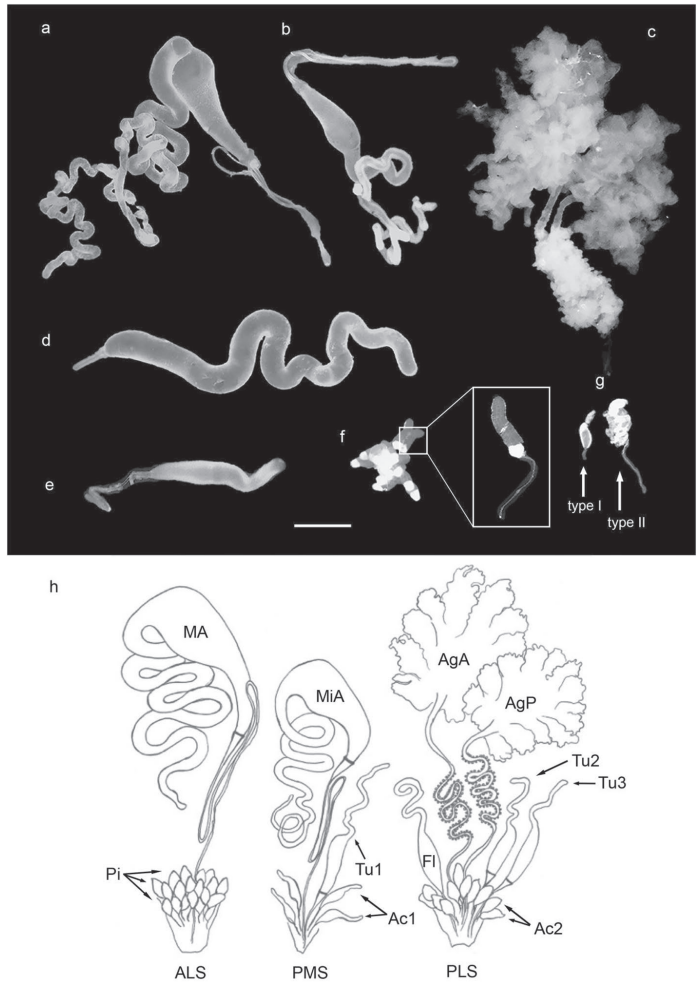

Figure 1. Silk glands of *L. sclopetarius*, only one set of the paired glands is shown for clarity. (**a**) Major ampullate gland. (**b**) Minor ampullate gland. (**c**) A pair of aggregate glands. (**d**) Tubuliform gland (from a spider with many mature eggs), only one of the three tubuliform glands from one side is shown. (**e**) Flagelliform gland. (**f**) A group of piriform glands, inset shows magnified image of single gland. (**g**) Aciniform glands (type I & type II). Scale bar in (**a**–**g**) = 1 mm. Parts of the minor ampullate & flagelliform gland tails were lost during dissection. (**h**) Schematic drawing of silk glands from one side of the opisthosoma in *L. sclopetarius* with an indication of which spinneret the gland is attached to (ALS- anterior lateral spinneret, PMS- posterior median spinneret, PLS- posterior lateral spinneret). MA- major ampullate, MiA- minor ampullate, Pi- group of piriform glands, Tu1–3- tubuliform glands, Fl- flagelliform gland, Ac1, Ac2- aciniform type I and type II glands, AgA- anterior aggregate gland, AgP- posterior aggregate gland).

surrounded by numerous nodules. The nodules were macroscopically visible (Fig. 1c) and consisted of a group of large cells with irregular nuclei and indefinite cell boundaries (Fig. 5c,d). In the opisthosoma, the ducts of two aggregate glands were located along with the duct of one flagelliform gland which can be visualized as a *triad* on the cross-sections (Supplementary, SI Fig. 3), close to the posterior lateral spinneret.

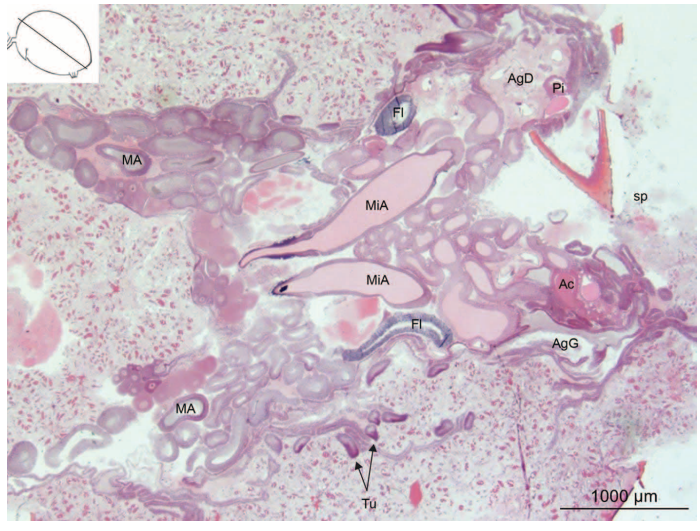

Figure 2. Frontal section of the opisthosoma of an adult female *L. sclopetarius* showing the localization of different glands. *MA* major ampullate, *MiA* minor ampullate, *Fl* flagelliform, *AgG* aggregate gland, *AgD* aggregate duct, *Ac* aciniform, *Tu* tubuliform, *Pi* piriform, *sp* spinnerets. The line in the inset indicates the approximate plane of the section in the opisthosoma. HE.

Tubuliform glands

Three tubuliform glands were located on each side of the median plane, one was connected to the posterior median spinneret and two were connected to the posterior lateral spinneret (Fig. 1h). These glands were long, tubular, and had thin short ducts (Fig. 1d). The tubular part of the gland narrowed towards the proximal end. The tubuliform glands were composed of simple columnar epithelium but the morphology of these glands varied between individuals. In some spiders, the gland had a single layer of tall columnar epithelial cells with very elongated, basally located nuclei and a constricted lumen (Fig. 6a). In spiders with large eggs, the epithelium was cuboidal or low columnar and nuclei were oval and basally located. In this case, the glandular lumen was very spacious and harbored secretory products that stained light pink with HE (Fig. 6b). The epithelial cells contained several concentric structures close to the nuclei which stained bright pink (Fig. 6b, blue arrowheads). The duct was lined with a cuticle and connected to the gland via a funnel similar to ampullate funnels (Fig. 6c).

Aciniform glands

Two types of aciniform glands were located close to the spinnerets. Type I aciniform glands consisted of a sac with a small tail at the proximal end. They occurred singly and their ducts connected to the posterior median spinneret. Type II aciniform glands were pear-shaped and presented in clusters. Their ducts connected to the posterior lateral spinnerets (Fig. 1g, h).

Type I aciniform glands were composed of two zones with histologically distinct features. The tail (zone A) had simple columnar epithelial cells with basally located nuclei and their granules stained light pink with HE. The distal part of the gland (zone B) had similar cells, but the granules stained bright pink (Fig. 7a).

The type II glands also consisted of two different regions. The proximal end was termed zone A and the distal end zone B. Both zones were composed of simple columnar epithelial cells with basally located nuclei. In zone A, the cellular granules were big and stained light pink. In zone B the cells contained two types of irregular granules that stained dark pink and bright purple. The secretory contents of the granules in zone B appeared to form an outer layer around the secretory content of zone A cells (Fig. 7b).

Piriform glands

Several piriform glands were located close to the spinnerets. These glands were macroscopically similar to aciniform type II glands, and their ducts exited through anterior lateral spinnerets (Fig. 1f,h). The gland contained two cellular zones, zone A (proximal end) and zone B (distal end), demarcated with a sharp boundary. In both zones, the cells were composed of a simple columnar epithelium with oval to round-shaped and basally located nuclei. The granules in the cells of zone A were big and stained light pink. In zone B, the cells had small granules which stained red. Zone B was connected to the duct which led to the spinnerets. The secretions of zone B cells formed a layer surrounding the weakly staining bulk content of the gland (Fig. 8).

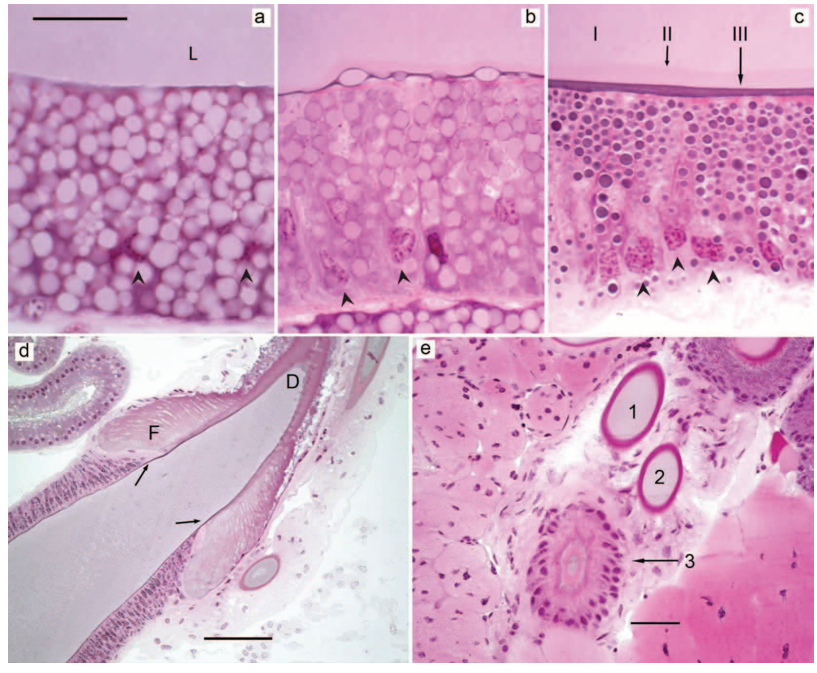

Figure 3. Major ampullate glands of *L. sclopetarius*. (**a**–**c**) Epithelial cells belonging to the three zones A, B and C, respectively, in the tail and sac of the major ampullate gland. The arrowheads indicate basally located nuclei. In (**c**), the three layers in the lumen (indicated as I, II and III), added by the three cell types A, B and C, respectively, can be distinguished. (**d**) Funnel region (F), the duct (D) that starts from the funnel is marked by arrows. (**e**) Cross section of the duct, the three limbs (numbered 1–3) are arranged in a triad fashion showing the decreasing diameter of the duct. Note that the cells surrounding the duct are taller in the third limb. L-lumen. HE, scale bars (**a**–**c**) = 25 µm, (**d**) = 100 µm and (**e**) = 50 µm.

Flagelliform glands

One pair of flagelliform glands were located dorsally of the minor ampullate glands. In appearance, the glands looked similar to tubuliform glands but had shorter tails and relatively larger sacs compared to their tails (Fig. 1h). We identified two distinct cellular regions, zone A and zone B (Fig. 9). The tail and about half of the sac was termed zone A, whereas the distal part that connected to the duct was termed zone B. Both zones were composed of simple columnar epithelium with basally located nuclei. Zone A cells had spherical nuclei and weakly-staining big, irregular granules. Apically these cells were irregular and appeared to secrete their content in the lumen that stained light pink (Fig. 9b). The cells harbored many granules with different staining content as indicated in Fig. 9b (red and blue arrowheads). Zone B had eosinophilic nuclei and cytoplasm full of evenly sized small secretory granules that stained dark pink (Fig. 9c).

Carbonic anhydrase activity

During the first step of CA staining, a chemical reaction occurs in locations with active CA whereby CO_2 leaves, the pH increases, and a cobalt-phosphate-carbonate complex is formed. In the subsequent steps, this complex converts into a black cobalt sulphide precipitate which can be visualized on sections. This method stains all active CA on the sections irrespective of the isoform. In the current study, active CA was found in all glands except the tubuliform glands and type I aciniform glands. Control sections incubated with the CA inhibitor acetazolamide were generally unstained. Low temperature euthanization of spiders produced the same staining results as CO_2 anesthesia.

In both major and minor ampullate glands, weak to moderate CA activity was observed in granules of the cells in zone C, funnel, and the cuticle of the duct (Fig. 10). No difference was observed in major ampullate glands of silked and unsilked individuals.

The basal membrane of the aggregate gland cells showed intense staining for CA (Fig. 11a). Cells in a small region between the distal end of the gland and the beginning of the duct had abundant small homogenous granules with strong black staining. The apical membrane of the cells in this region also showed intense staining for

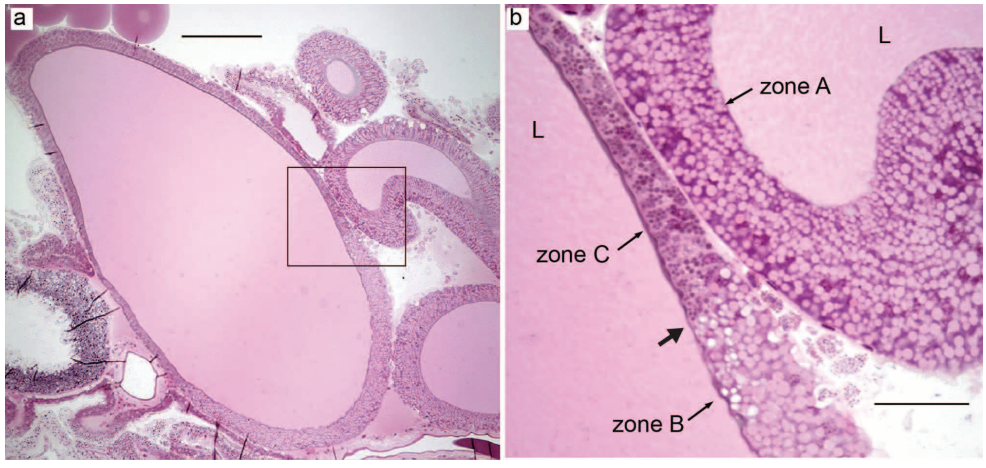

Figure 4. (**a**) Minor ampullate gland of *L. sclopetarius*. The inset is magnified in (**b**). Zones A–C are indicated by thin arrows, thick arrow shows the transition from zone B to C. L- lumen. HE, scale bars (**a**) = 200 µm, (**b**) = 50 µm.

CA (Fig. 11b). The ducts of the aggregate glands showed variable intensity of CA staining. In some parts of the ducts, the cuticular intima as well as the outer membrane of the surrounding cells showed strong black staining (Fig. 11c), whereas in other parts of the ducts the staining was very weak (Fig. 11e). Interestingly, some parts of the ducts displayed varying staining intensity within a cross-section (Fig. 11d). The cells of the nodules surrounding the epithelium in all parts of the aggregate ducts showed remarkably intense staining in the cytoplasm as well as their basal membrane. The nuclei of epithelial cells surrounding the duct and the cells in the nodules were unstained.

In the flagelliform gland, a few cells at the beginning of zone B had small, slightly irregular granules that stained intensely and some big granules that stained weakly for CA (Fig. 12a). The apical surface lining these cells showed abundant CA activity, while the basal membrane lining the whole flagelliform gland displayed weak CA staining.

In piriform glands, weak CA activity was seen in big granules of zone B cells (Fig. 12b) and in the secretion along the apical surface. No CA activity was detected in the cuticle of the duct. In zone B of type II aciniform glands, the CA activity was similar but slightly stronger as compared to piriform glands (Fig. 12c). These cells contained big granules that stained light grey as well as small granules that stained black. The secretion from zone B cells showed moderate CA activity.

Thioflavin-S

Thioflavin is a common dye used to detect amyloid fibrils since it gives a characteristic fluorescence upon binding to the fibrillar aggregates[47]. The luminal contents and the content of the granules in the tail of the major ampullate gland showed bright green fluorescence. Strikingly, the fluorescence ceased at the junction of zone B and C, indicating that the granules of both zone A and B cells contain thioflavin-S positive components (Fig. 13a). We observed differential staining in the luminal contents that appeared similar to the granules of zone A and B cells (indicated as I & II in Fig. 13a,b). Other silk glands did not show any detectable fluorescence. Major ampullate silk fibers were also thioflavin-S positive, as anticipated (Fig. 13c). As expected, we observed thioflavin-S fluorescence in control amyloid fibrils made from the Aβ42 peptide that is associated with Alzheimer's disease (Fig. 13d).

Congo red

Congo red staining and green birefringence are hallmarks of amyloid fibrils. Even without any counterstaining, the morphology of the opisthosoma was easy to examine due to a weak but evident greyish birefringence (Fig. 14). A weak Congo red affinity appeared in the lumen of ampullate glands close to the spinnerets (Fig. 14a). The walls and contents of the other glands showed no definite Congo red affinity but showed a grey birefringence between crossed polars. In one of the sections, a few round distinct structures appeared in a couple of tail cross-sections. They stained strongly with Congo red although no clear birefringence was seen (Fig. 14b–e). The exoskeleton showed a strong Congo red affinity and an equally strong green-yellow-orange birefringence.

PAS

PAS is a routinely used staining method to identify polysaccharides, mucosubstances, and mucins in tissue sections. In the ampullate glands, the zone C cells and the funnel stained weakly with PAS (Fig. 15a,b). In

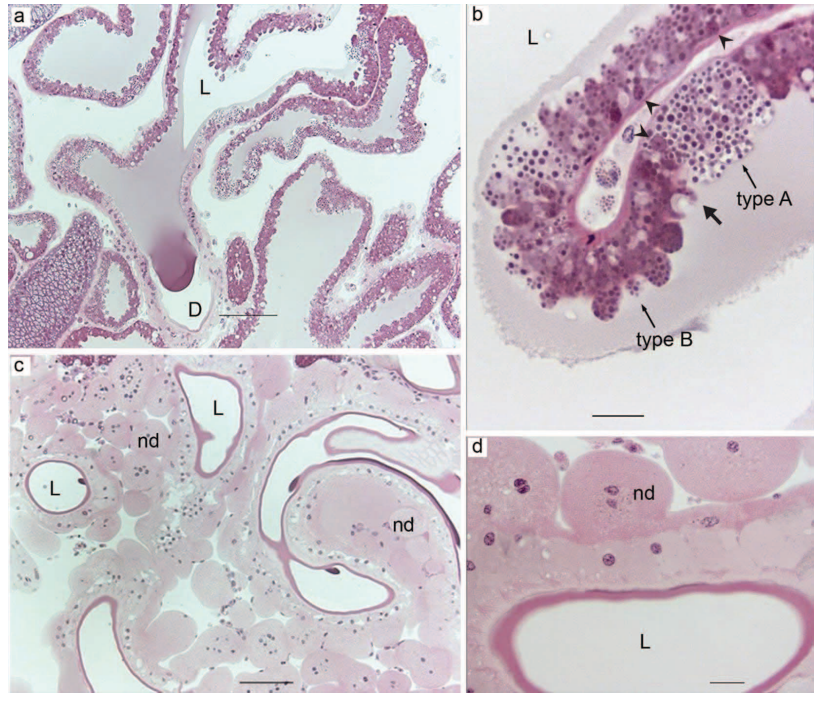

Figure 5. (**a**) Aggregate glands of *L. sclopetarius* showing a thin epithelium and spacious lumen (L), D-duct. (**b**) Two cell types (A & B) and the apical projections containing several granules. Thick arrow indicates the sharp transition between two cell types and the arrowheads indicate nuclei. (**c**) Ducts of aggregate glands. The cuticular intima is surrounded by a single layered epithelium. The ducts have numerous nodules (nd). (**d**) Magnified image of a duct showing the cells surrounding the cuticular intima and the cells of the nodules. HE, scale bar (**a**,**c**) = 100 µm and in (**b**,**d**) = 20 µm.

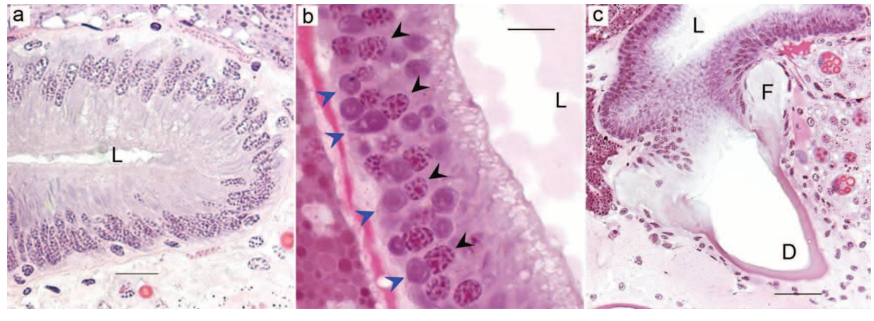

Figure 6. Tubuliform glands from two *L. sclopetarius*. (**a**) Tubuliform gland in female with no mature eggs. The lumen is narrow, cells are tall, and the nuclei are elongated. (**b**) Tubuliform gland in female with many mature eggs. The cells are low columnar with almost round nuclei and the lumen shows secretion. Black arrowheads indicate nuclei and blue arrowheads indicate structures resembling endoplasmic reticulae. (**c**) Funnel that connects the gland to the duct. L-lumen, D-duct. HE, scale bars: (**a**,**b**) = 10 µm, and (**c**) = 50 µm.

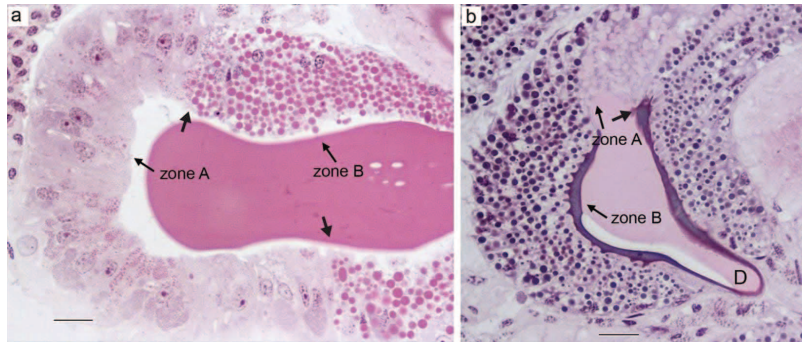

Figure 7. Aciniform glands of *L. sclopetarius*. (**a**) Type I aciniform gland showing two kinds of cells. (**b**) Type II aciniform gland showing two cell types that stain differently compared to type I aciniform cells. Zones are indicated in both types of glands. Thick arrows indicate the sharp transition between the zones, D-duct. HE, scale bars: (**a**) = 50 µm and (**b**) = 20 µm.

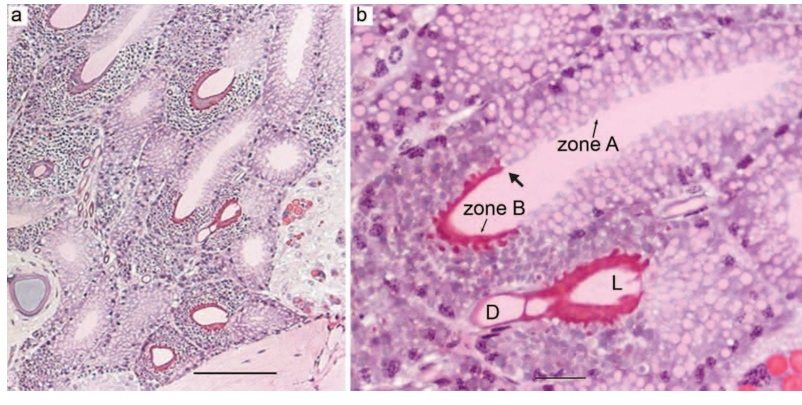

Figure 8. Piriform glands of *L. sclopetarius*. (**a**) A group of piriform glands in the opisthosoma. (**b**) Magnified image of the piriform glands showing two cell types, L-lumen, D-duct. Sharp border between two zones is indicated with an arrow. HE, scale bar: (**a**) = 100 µm, (**b**) = 20 µm.

flagelliform glands, the zone A & zone B cells stained very weakly but differently (Fig. 15c,d). The staining was very intense in aggregate glands, especially, in the granules of type B cells (Fig. 15e) and in the nodules surrounding the ducts (Fig. 15f). Very weak staining was observed in the apical surface and the lumen of the tubuliform glands (Fig. 15g). Zone B cells of piriform glands were characterized by intermediate PAS staining (Fig. 15h). An intense PAS staining was observed in the granules and the apical lining of zone B cells of type II aciniform glands (Fig. 15i).

Discussion

By using resin-embedded tissue, the morphology of all silk glands of *L. sclopetarius* could be visualized at a higher resolution compared to previously published work[5]. It allowed us to identify detailed morphological features including the presence of several cell types and distinct regionalization within all silk glands, some of which have not been described previously and that have implications for the architecture and composition of the silk fibers.

The major ampullate gland produces silk that is used by spiders as a lifeline as well as for the radial- and frame- lines of the web[15]. In agreement with previous work[11], the epithelium of the tail and sac of the major ampullate gland was herein shown to be composed of three zones, A–C. We observed that the cells of each zone of the gland harbored numerous secretory granules which stained distinctly with HE. According to Andersson et al.[11], zones A and B synthesize and secrete spidroins into the lumen, while the contents of zone C granules

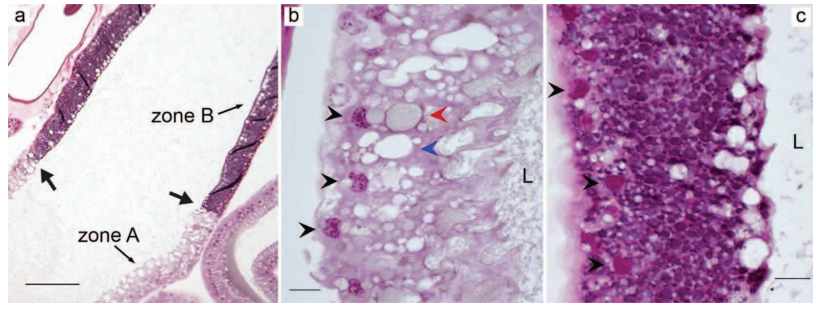

Figure 9. Flagelliform gland from *L. sclopetarius*. (**a**) Arrows indicate transition between zones. (**b**) Zone A cells with weakly stained granules. The apical surface is highly irregular in this region. Nuclei are indicated with black arrowheads. Blue and red arrowheads indicate empty and filled vesicles, respectively. (**c**) Zone B cells showing brightly stained granules and eosinophilic nuclei (black arrowheads), L-lumen. HE, scale bars: (**a**) = 100 µm, (**b**,**c**) = 10 µm.

Figure 10. Black staining show carbonic anhydrase activity in the ampullate glands of *L. sclopetarius*. (**a**) Zone C of major ampullate gland showing weak CA activity in granules and apical lining, L- lumen. Arrow indicates the border between zone B & C. (**b**) Funnel (F) and cuticular intima of the major ampullate duct (arrow) show weak CA activity. (**c**) Weak CA activity in zone C, and intermediate in the funnel (F) and cuticle (arrows) of minor ampullate glands. This section was counterstained with azure blue. Scale bars: (**a**) = 50 µm, (**b**) = 100 µm, (**c**) = 20 µm.

are unknown. In the present study, the contents secreted from each zone in the major ampullate gland did not mix, suggesting that the sequential addition of different components will lead to a complex and multi-layered structure of the fiber. In line with this, Sponner and coworkers demonstrated that the major ampullate silk has a multilayer arrangement, they identified four layers: the core, skin, glyco, and lipid layer[28]. Glycoproteins are most abundant in the glyco layer but can also be detected in the skin and the core layers. The function of the glyco and lipid layers is unknown, but they may aid in protecting against microbial degradation and may play a role in regulating the water content[24,28]. Here, we observed that the granules in zone C stained weakly for PAS, corroborating previous findings, and identifying the origin of the glycoproteins to be zone C epithelial secretions.

To detect morphological changes associated with increased protein production and silk secretion, we silked a group of spiders for 15 min before sacrifice and performed histological examination of the glands. However, we could see no difference in gland morphology or in CA activity compared to unsilked individuals. This is in contrast to a previous report[46], but we speculate that the difference may be due to that we used a quite short silking time (15 min compared to 60 min in Moon & Tillinghast[46]), which may be a too short time frame for visible changes to occur.

Spider major ampullate silk has been suggested to be amyloid-like since parts of the spidroins in the fiber form β-sheets that resemble those of amyloid fibrils[21,48]. Before converting to mainly β-sheet structures, the spidroins are in unordered or helical conformations during synthesis in zone A and B cells and during storage in the lumen[49–51]. In an attempt to shed light on silk polymerization, we therefore used Congo red and thioflavin-S staining to locate amyloid-like structures in the major ampullate glands. The thioflavin-S stained sections showed fluorescence in all locations of the major ampullate gland where spidroins are expected to be found, and thus suggest that the spidroins in the granules of zone A and B and in the lumen have converted to β-sheet rich structures. This could possibly be a consequence of the treatment associated with the embedding of the tissues, but it is still surprising that only secretions from zone A and B cells in the major ampullate gland stained with thioflavin-S. The lack of staining in the vesicles of zone C cells further strengthens the notion that these cells do

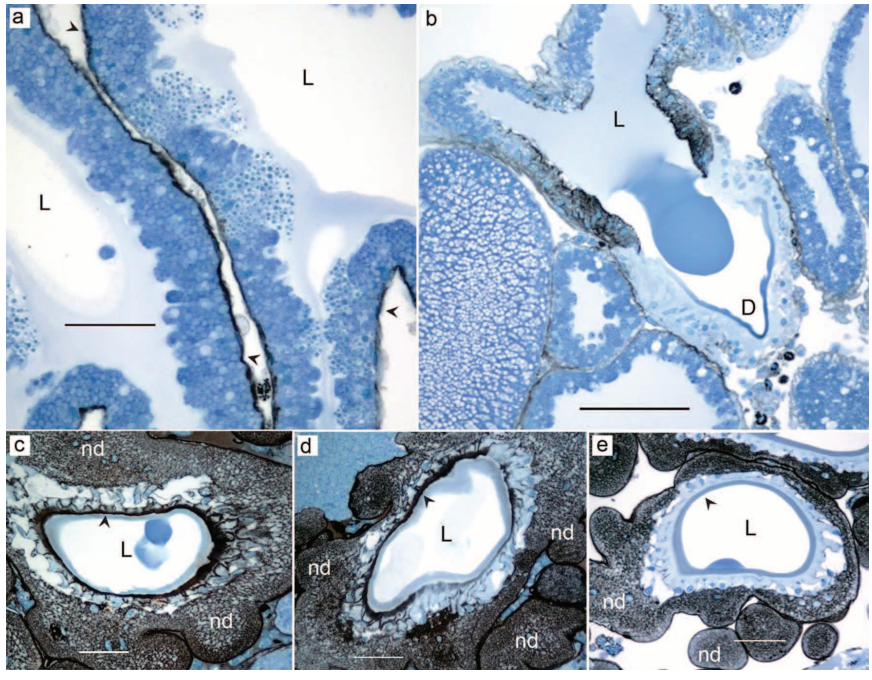

Figure 11. Black staining show carbonic anhydrase activity in aggregate glands of *L. sclopetarius*. (**a**) Active CA is seen throughout the basal membrane of the glands (arrowheads). (**b**) A group of cells close to the duct (D) show active CA. (**c**) Cross-section showing strong staining in cuticular intima of the duct and cell membrane of the surrounding epithelium. (**d**) Cross-section of a duct with variable CA staining. (**e**) Weak staining in the cuticle and no visible staining in the epithelium. In (**c**–**e**), cuticle is indicated with arrowheads. Note the strong staining for CA in the nodules (nd). L-lumen. Azure blue counterstain. Scale bars: (**a**,**c**–**e**) = 50 µm & (**b**) = 100 µm.

not secrete spidroins to a large extent. Congo red staining with concomitant birefringence under crossed polars is a more specific dye for amyloid than thioflavin-S, and a clear affinity of Congo red with typical birefringence is usually seen only after distinct amyloid fibrils have been formed. In this study, we could not detect distinct staining of any part of the gland in sections stained with Congo red. The strong affinity for the dye and concomitant birefringence of the exoskeleton likely depends on non-amyloid components.

The minor ampullate gland is similar in anatomy to the major ampullate and makes silk for the temporary scaffolding of the web[15]. To our knowledge, this is the first study which reports that the secretory portion (sac and tail) of the minor ampullate gland is composed of three cell types. The higher resolution of the thin (1 µm) resin sections made it possible to identify the transition between zone A and B compared to previous studies that have reported only two cell types in these glands[5,34]. Active CA was identified in zone C and funnel regions of the minor ampullate glands and zone C stained weakly for PAS. These results are very similar to major ampullate glands, and it is plausible to assume that minor and major ampullate silks are similar in structure.

Flagelliform glands produce silk for the capture spiral of the web[15] and in this study, two zones containing histologically distinct cells were identified in these glands. Numerous irregular unstained granules were observed in the secretory portion of the gland, which may indicate that the content was dissolved during processing. The morphology of the flagelliform glands of this study is similar to what Mullen regards as median ampullate glands[5]. We noted that CA activity was localized to a short distinct region in the beginning of zone B, which may indicate that these cells should be regarded as a third cell type, but these cells could not be distinguished from the other cells in zone B when examined with HE. An interesting observation was that the CA activity was found in the middle of the gland and not in the distal part as in other glands producing silk fibers.

We identified two types of aciniform glands in *L. sclopetarius*. Aciniform silk is used for swathing insects and for the inner layer of egg cocoons[15]. It is not known if these glands produce two different types of silk fibers and if so, whether they are produced by the two different types of aciniform glands. According to Moon et al. type I glands contain electron-dense protein granules, whereas type II glands contain electron-lucent lipid granules in *Trichonephila clavata*[38]. In the current study, the two aciniform gland types both consisted of two cell types each confined to different zones. We could neither detect CA activity nor any PAS-positive structures in the type I

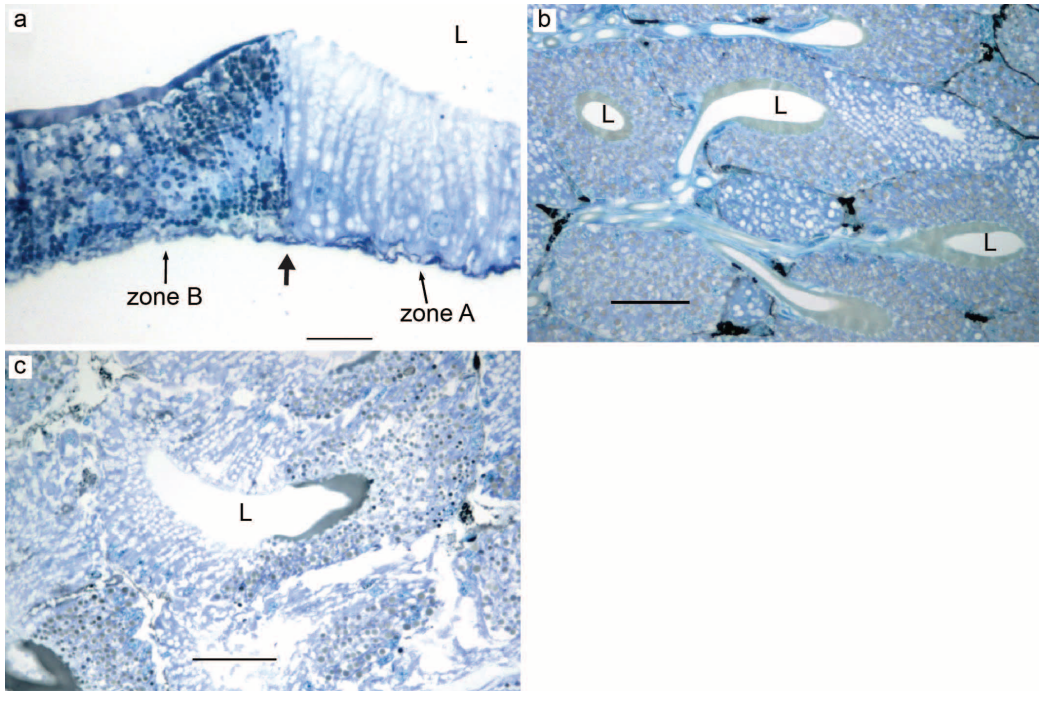

Figure 12. Carbonic anhydrase activity in flagelliform, piriform and aciniform glands of *L. sclopetarius*. (**a**) A small region in zone B of flagelliform glands show strong staining for CA in granules and cell membranes, arrow indicates border between zone A and B, L- lumen. (**b**) Weak CA staining in granules and apical lining of zone B cells of piriform glands. (**c**) Intermediate CA staining in some granules and the apical lining of Zone B cells of type II aciniform glands. Azure blue counterstain. Scale bars: (**a**–**c**) = 50 µm.

glands. However, active CA was found in zone B cells of type II aciniform glands, which also stained strongly with PAS indicating the presence of carbohydrates/glycoproteins. This indicates that at least type II aciniform glands, which are more numerous than type I glands, may produce silk by a similar pH-dependent mechanism as the major ampullate glands.

Aggregate glands produce a viscous solution that covers the core fiber produced by flagelliform glands[52]. We demonstrate, for the first time, that aggregate glands have two types of cells (A and B). The two cell types differed in height and their staining of granules. The different cell types were not confined to specific zones, but type B cells were more abundant than A cells. The aggregate ducts were convoluted and surrounded by several nodules. The ducts of two aggregate glands were arranged in a triad with one flagelliform duct in the center, as previously seen in two other species[12,37,53]. The triad arrangement of these ducts is thought to facilitate the extrusion of the silk so that the axial fibers formed by flagelliform glands are simultaneously covered with glue droplets produced by aggregate glands as the silk emerges out[54]. In the current study, the strongest staining for both CA and PAS was detected in the ducts of the aggregate glands, in the cytoplasm of the nodular cells, and in parts of the cuticular intima. In accordance, ultrastructural studies have shown that the nodular cells are rich in glycogen and mitochondria[55], and the internal cells of the ducts have characteristics of absorbing cells[56]. In two orb-weaving spiders, the nodules are shown to have glycogen phosphorylase, an enzyme vital for glycogenolysis[57]. Some studies suggest that these cells transport water, ions (especially phosphate), and possibly some organic compounds[58] using glycogen as energy[57], providing a possible explanation for how the viscous solution is produced. Additionally, it's worth noting that aggregate proteins are known to undergo post-translational glycosylation, which could also potentially account for the observed intense PAS staining in these glands[59–63].

In this study, all three pairs of tubuliform glands had identical morphology, consisted of only one cell type, and had a funnel that connected the gland and the duct as previously reported by[5]. The role of the funnel is not clearly understood, but it is seen only in the ampullate and tubuliform glands. The pink spiral structures observed in the epithelial cells of these glands (Fig. 6b) could be endoplasmic reticula which are reported to be abundant in these glands[64,65]. We did not detect CA activity in the tubuliform glands of *L. sclopetarius*, which is in contrast to the findings in other orb weavers[21]. Tubuliform silk is used for making the egg sack and is therefore only active

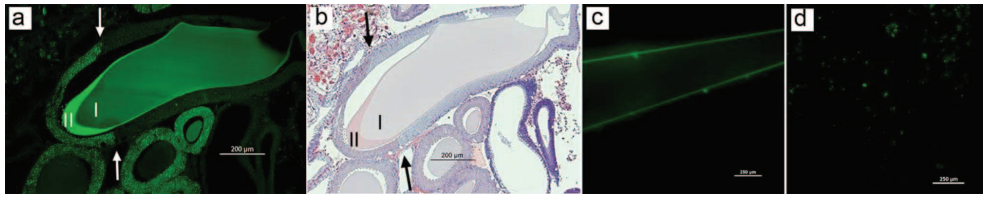

Figure 13. Thioflavin-S staining in major ampullate gland and silk fibers of *L. sclopetarius*. (**a**) Major ampullate gland showing green fluorescence in zone A & B cells. I & II indicate two layers distinguishable by Thioflavin-S. (**b**) Consecutive section stained with HE. I & II correspond to the layers detected in (**a**). In (**a**,**b**), arrows indicate the junction between zones B and C. (**c**) Native silk fibers stained with Thioflavin-S. (**d**) Aβ42 amyloid fibrils stained with Thioflavin-S served as a positive control. Scale bars: (**a**) = 200 µm, (**b**) = 100 µm, (**c**,**d**) = 250 µm.

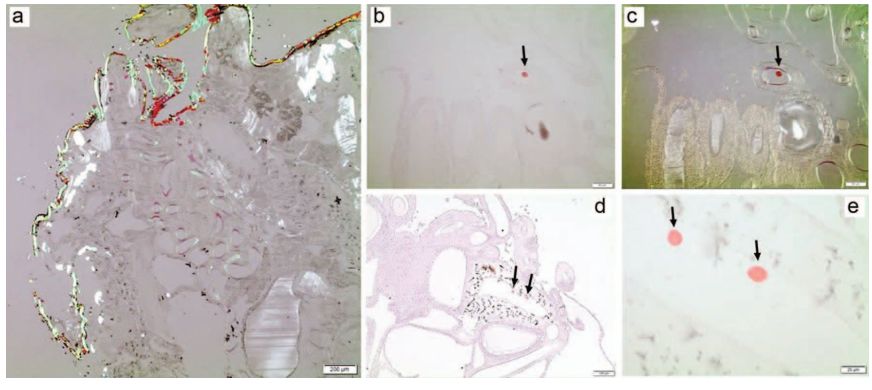

Figure 14. Congo red staining in silk glands of *L. sclopetarius*. (**a**) Part of a spider opisthosoma stained with Congo red and visualized in polarized light with crossed polars. A bright birefringence with predominantly green color is present in the exoskeleton. (**b**,**c**) A single strongly congophilic but not birefringent round structures found in one ampullate gland. (**b**) ordinary light, and (**c**) between crossed polars. (**d**,**e**) Two other congophilic structures, about 15 µm in diameter, visualized in ordinary light. At high magnification (**e**) there seems to be an outer demarcation of the structures. Scale bars: (**a**) = 200 µm, (**b**,**c**) = 50 µm, (**d**) = 100 µm, (**e**) = 20 µm.

during a phase in the spider´s life. Accordingly, the morphology of the gland varies with the development of oviducts and eggs[64]. Evidence for different developmental stages of the tubuliform glands was seen in our sections but we could not determine if this was the reason that no CA activity could be detected.

Piriform silk is used for cementing the junctions in the web as well as for making the attachment discs that secure the ampullate fibers to the substrate[15]. Results of the current study showed that these glands were composed of two zones with distinct cell types as described previously in *L. sclopetarius*[5] and *A. diadematus*[31]. Active CA and PAS-positive structures were detected in zone B cells. The PAS structures might be glycoproteins which are thought to serve an adhesive function to the silk equivalent to that of sericin[31]. The silk produced by piriform glands is composed of many small fibers which may contribute to the overall mechanical properties of the attachment discs[66]. The polymerization of these silk fibers may be aided by acidic environments similar to that in major ampullate glands which can be facilitated by active CA in these glands.

In the major ampullate gland, CA is responsible for generating and upholding a pH gradient, from 7.6 to < 5.7[21]. This pH gradient is important for major ampullate spidroin assembly and phase transition in the duct. It is the spidroin terminal domains that are primarily affected by the increasingly acidic conditions and mediate fiber formation[20,21,67–70]. Recent studies of the spider genome and transcriptome in the silk glands show that the terminal domains are present in all spidroin types[71–76]. This, combined with our findings that CA activity is found in all silk glands (except the tubuliform & type I aciniform glands) suggests that the pH gradient may be universal to silk glands and that polymerization of all different types of spider silk is dependent on a pH gradient generated by CA. However, the location of CA in the flagelliform glands is somewhat odd and how this relates to a possible pH gradient in these glands requires further studies. Regarding the lack of CA activity in the tubuliform gland, it should be noted that another study has identified CA activity also in this gland[21]. As mentioned earlier, the tubuliform gland is active only in association with egg laying and construction of the egg sack, which could be

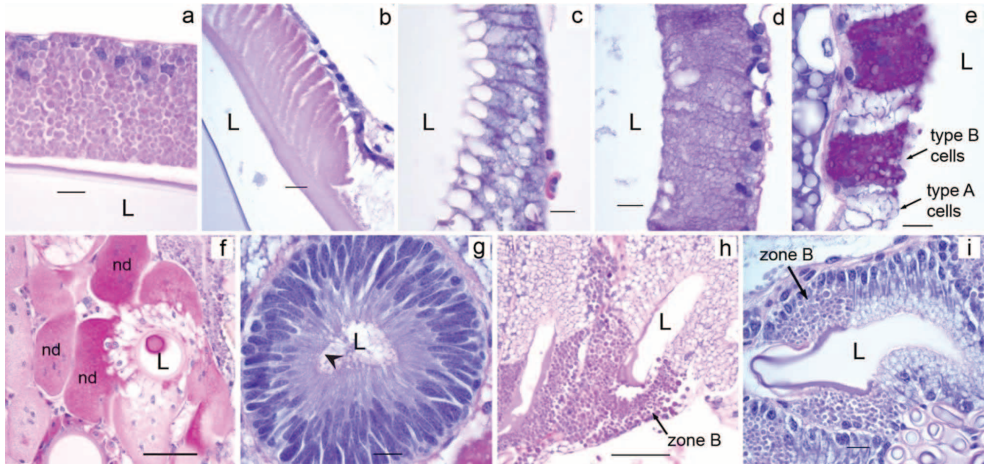

Figure 15. PAS staining in silk glands of *L. sclopetarius*. (**a**) PAS staining in granules of zone C cells of major ampullate gland. (**b**) Major ampullate funnel showing PAS staining. (**c**) Weak staining in zone A cells of flagelliform glands. (**d**) Intermediate staining in zone B cells of flagelliform glands. (**e**) Intense staining in type B cells of aggregate glands. (**f**) Nodular cells (nd) surrounding the aggregate ducts show intense PAS activity. (**g**) Weak staining in tubuliform cells apically (arrowhead). (**h**) Intermediate PAS staining in zone B cells of piriform glands. (**i**) Intermediate staining in zone B cells of type II aciniform glands, L – lumen. Scale bars: (**a**–**e**,**g**) = 10 µm, (**f**,**h**,**i**) = 50 µm.

true also for the aciniform type I glands. It is not established which of the aciniform glands that are responsible for making the inner lining of the egg sack but provided that aciniform type I glands produce this silk, a plausible explanation to the lack of detectable CA activity could be that both glands are dormant in our sections. In the aggregate glands, strong staining for CA was detected in the ducts, in the cytoplasm of the nodular cells, and in parts of the cuticular intima. We speculate that the abundant CA activity seen in these cells plays an active role in generating the bicarbonates or protons needed for glue formation.

In the α-CA class, 16 CA isozymes have been identified to date which are subdivided into the following groups: cytosolic, mitochondrial, secreted, GPI-anchored, transmembrane, and acatalytic[77]. Genes coding for two different orthologs (isozymes) of carbonic anhydrases (Ca13 and Ca14) have been identified as silk gland-specific transcripts in *Trichonephila clavipes*[78]. CA13 and CA14 belong to the cytosolic and transmembrane (extracellular active site) classes, respectively. Based on the staining in our sections, the CA activity is associated with cell membranes, granules, and/or cytoplasm, but it is beyond the scope of this study to assign the spider CAs to specific classes. We detected active CA in several other structures in the opisthosoma besides silk glands such as hemolymph, muscles, book lungs, and oviducts (results not shown). Except for the importance of CA to generate a pH gradient in the major ampullate gland, little is known about spider-specific CAs and their functions.

In conclusion, we have characterized all seven silk glands and ducts of *L. sclopetarius* using various morphological methods. The regionalization of multiple cell types in all glands, except tubuliform and aggregate glands, suggests that the secretions may be added in layers, indicating that spider silk fibers are complex layered structures. Based on the identification of different cell types, the addition of secretory products in layers, and the localization of active CA we speculate that the process of silk formation in minor ampullate, aciniform type II, and piriform glands may be similar to the mechanism previously described for major ampullate glands. In addition, PAS-positive structures in several glands support previous findings that carbohydrates are components of silk fibers.

Materials and methods

Adult females (n = 32) of *L. sclopetarius* were collected in Uppsala, Sweden between June and October of 2019. They were kept in individual containers, fed small crickets and water.

Tissue processing and light microscopy

Spiders were anesthetized using CO_2 gas and severed at the pedicel. One group of spiders (n = 7) was silked for about 15 min. To make the spiders spin major ampullate silk, they were gently made to fall several times from a wooden block on which the silk was collected, before they were sacrificed. The dissection was performed on a wax plate placed on ice using 9 mg/mL (154 mM) sodium chloride (Fresenius Kabi AG, Germany) with the help of a Leica M60 stereomicroscope equipped with a Leica IC80 HD camera. Individual glands and whole opisthosomas were fixed in 2.5% glutaraldehyde in 67 mM phosphate buffer, pH 7.2 for 24 h at 4 °C, and later rinsed in 67 mM phosphate buffer. The tissues were dehydrated in increasing concentrations of ethanol (50, 70, 90, and 100%, for

30 min each), infiltrated, and embedded in a water-soluble glycol methacrylate (Leica Historesin). Sections (1 and 2 µm) were cut with glass knives using a Leica RM 2165 microtome, stained with hematoxylin–eosin (HE), and mounted using Agar 100 resin. HE staining is one of the most common techniques used in histological analysis that allows for visualization of cells and tissues, with hematoxylin staining cell nuclei blue-purple and eosin staining cytoplasm and extracellular matrix pink.

Staining for carbonic anhydrase activity

CA activity was detected using a very selective histochemical method[45]. Serial sections were incubated floating on a solution containing 156.6 mM $NaHCO_3$, 3.5 mM $CoSO_4$, 52.6 mM H_2SO_4, and 11.7 mM KH_2PO_4. In locations with active CA, a cobalt-phosphate-carbonate complex was formed. Sections were moved to a solution containing $CoSO_4$ and $(NH_4)_2S$, and the complex was converted to a black cobalt sulphide precipitate. The sections were counterstained with azure blue. At least one section from each series was incubated with the CA inhibitor acetazolamide, to identify areas with unspecific staining. To check that the CO_2 anesthesia did not affect CA activity, a few spiders were euthanized using low temperature (− 20 °C, 5 min). Sections from these spiders were otherwise treated as above. All resin sections were evaluated using a Nikon Microphot-FXA (Tekno Optik AB) microscope equipped with a Nikon FX-35DX camera. Images were captured and edited using the software Eclipse Net v1.20.0.

Staining with thioflavin-S

Whole opisthosomas of six spiders were fixed in 4% formalin for 24 h at room temperature then rinsed in 67 mM phosphate buffer and embedded in paraffin. The samples were sectioned (4 µm) using a MICROM HM 355S microtome. For staining with thioflavin-S, deparaffinized sections were incubated in 0.1% thioflavin-S solution (prepared in 50% ethanol) for 1 h at room temperature (protected from light). The sections were washed in 80%, 95% ethanol, and pure water, respectively, dried, and mounted with coverslips. Native silk fibers collected from bridge spiders were secured on glass slides and incubated with 0.1% thioflavin-S solution for one hour at room temperature. Amyloid fibrils prepared from 24 µM Aβ42 monomers, which were recombinantly produced as described in Ref.[79], were used as positive control whereas silk fibers incubated in 50% ethanol were used as negative controls. Sections and fibers were evaluated using a Nikon Eclipse E600 fluorescence microscope equipped with a Nikon DXM1200 digital camera.

Staining with congo red and PAS

For Congo red staining, paraffin sections were stained with alkaline Congo red[80], modified as described in Westermark[81], without counterstain and examined, with and without crossed polars, in an Olympus BX51 polarization microscope equipped with a U-AN350P polarization device. For PAS, deparaffinized sections were incubated in 0.5% periodic acid (5 min), rinsed in distilled water, incubated in Schiff's reagent (20 min, protected from light), and rinsed in tap water. The sections were counter-stained with Mayer's hematoxylin for 1 min and mounted. Sections were evaluated using the Nikon Microphot-FXA (Tekno Optik AB) microscope equipped with Nikon FX-35DX camera. Images were captured and edited using the software Eclipse Net v1.20.0.

Data availability

The data from this study are available upon request from the corresponding author.

Received: 16 February 2023; Accepted: 9 December 2023
Published online: 14 December 2023

References

1. Bakhshandeh, B., Nateghi, S. S., Gazani, M. M., Dehghani, Z. & Mohammadzadeh, F. A review on advances in the applications of spider silk in biomedical issues. *Int. J. Biol. Macromol.* **192**, 258–271. https://doi.org/10.1016/j.ijbiomac.2021.09.201 (2021).
2. Andersson, M. *et al.* Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat. Chem. Biol.* **13**, 262–264. https://doi.org/10.1038/nchembio.2269 (2017).
3. Bittencourt, D. M. C. *et al.* Bioengineering of spider silks for the production of biomedical materials. *Front. Bioeng. Biotechnol.* **10**, 958486. https://doi.org/10.3389/fbioe.2022.958486 (2022).
4. Ramezaniaghdam, M., Nahdi, N. D. & Reski, R. Recombinant spider silk: Promises and bottlenecks. *Front. Bioeng. Biotechnol.* **10**, 835637. https://doi.org/10.3389/fbioe.2022.835637 (2022).
5. Mullen, G. R. Morphology and histology of the silk glands in Araneus sericatus Cl. *Trans. Am. Microscopic. Soc.* **88**, 232–232. https://doi.org/10.2307/3224495 (1969).
6. Palmer, J. M., Coyle, F. A. & Harrison, F. W. Structure and cytochemistry of the silk glands of the mygalomorph spider Antrodiaetus unicolor (araneae, antrodiaetidae). *J. Morphol.* **174**, 269–274. https://doi.org/10.1002/jmor.1051740303 (1982).
7. Palmer, J. M. The silk and silk production system of the funnel-web mygalomorph spider Euagrus (Araneae, Dipluridae). *J. Morphol.* **186**, 195–207. https://doi.org/10.1002/jmor.1051860205 (1985).
8. Peters, H. M. & Kovoor, J. The silk-producing system of Linyphia triangularis (Araneae, Linyphiidae) and some comparisons with Araneidae—Structure, histochemistry and function. *Zoomorphology* https://doi.org/10.1007/BF01632706 (1991).
9. Kovoor, J. & Munoz-Cuevas, A. Structure and function of the silk-gland system in Oxyopidae (Araneae). *Proceedings of the 17th European Colloquium of Arachnology, Edinburgh*, 133–141 (1998).
10. Casem, M. L., Tran, L. P. P. & Moore, A. M. F. Ultrastructure of the major ampullate gland of the black widow spider, Latrodectus Hesperus. *Tissue Cell* **34**, 427–436. https://doi.org/10.1016/S0040816602000836 (2002).
11. Andersson, M., Holm, L., Ridderstrale, Y., Johansson, J. & Rising, A. Morphology and composition of the spider major ampullate gland and dragline silk. *Biomacromolecules* **14**, 2945–2952. https://doi.org/10.1021/bm400898t (2013).
12. Chaw, R. C. & Hayashi, C. Y. Dissection of silk glands in the Western black widow Latrodectus hesperus. *J. Arachnol.* **46**, 159–161. https://doi.org/10.1636/JoA-16-S-063.1 (2018).
13. Work, R. W. Mechanisms of major ampullate silk fiber formation by orb-web-spinning spiders. *Trans. Am. Microscopic. Soc.* **96**, 170–189 (1977).

14. Vollrath, F., Knight, D. P. & Hu, X. W. Silk production in a spider involves acid bath treatment. *Proc. R. Soc. B Biol. Sci.* **265**, 817–820. https://doi.org/10.1098/rspb.1998.0365 (1998).
15. Kovoor, J. Comparative structure and histochemistry of silk-producing organs in arachnids. *Ecophysiol. Spiders* https://doi.org/10.1007/978-3-642-71552-5_12 (1987).
16. Bell, A. L. & Peakall, D. B. Changes in fine structure during silk protein production in the ampullate gland of the spider Araneus sericatus. *J. Cell Biol.* **42**, 284–295. https://doi.org/10.1083/jcb.42.1.284 (1969).
17. Vollrath, F. & Knight, D. P. Structure and function of the silk production pathway in the spider Nephila edulis. *Int. J. Biol. Macromol.* **24**, 243–249. https://doi.org/10.1016/s0141-8130(98)00095-6 (1999).
18. Moon, M.-J., Kim, C.-S. & Kim, W.-K. Ultrastructure of the ampullate gland in the orb web spider, Nephila clavata L. Koch I. Excretory duct of the large ampullate gland. *Appl. Microsc.* **18**, 77–90 (1988).
19. Moon, M.-J., Kim, C.-S. & Kim, W.-K. Ultrastructure of the ampullate gland in the orb web spider, Nephila clavata L. Koch II. Sac and tail portion of the large ampullate gland. *Appl. Microsc.* **18**, 91–101 (1988).
20. Kronqvist, N. *et al.* Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nat. Commun.* **5**, 3254. https://doi.org/10.1038/ncomms4254 (2014).
21. Andersson, M. *et al.* Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biol.* **12**, e1001921. https://doi.org/10.1371/journal.pbio.1001921 (2014).
22. Knight, D. P. & Vollrath, F. Changes in element composition along the spinning duct in a Nephila spider. *Naturwissenschaften* **88**, 179–182. https://doi.org/10.1007/s001140100220 (2001).
23. Knight, D. & Vollrath, F. Liquid crystals and flow elongation in a spider's silk production line. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **266**, 519–523 (1999).
24. Augsten, K., Muhlig, P. & Herrmann, C. Glycoproteins and skin-core structure in Nephila clavipes spider silk observed by light and electron microscopy. *Scanning* **22**, 12–15. https://doi.org/10.1002/sca.4950220103 (2000).
25. Frische, S., Maunsbach, A. B. & Vollrath, F. Elongate cavities and skin-core structure in Nephila spider silk observed by electron microscopy. *J. Microsc.* **189**, 64–70. https://doi.org/10.1046/j.1365-2818.1998.00285.x (1998).
26. Li, S. F., McGhie, A. J. & Tang, S. L. New internal structure of spider dragline silk revealed by atomic force microscopy. *Biophys. J.* **66**, 1209–1212. https://doi.org/10.1016/S0006-3495(94)80903-8 (1994).
27. Poza, P., Perez-Rigueiro, J., Elices, M. & Llorca, J. Fractographic analysis of silkworm and spider silk. *Eng. Fract. Mech.* **69**, 1035–1048. https://doi.org/10.1016/S0013-7944(01)00120-5 (2002).
28. Sponner, A. *et al.* Composition and hierarchical organisation of a spider silk. *PLoS One* **2**, e998. https://doi.org/10.1371/journal.pone.0000998 (2007).
29. Sponner, A., Unger, E., Grosse, F. & Weisshart, K. Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat. Mater.* **4**, 772–775. https://doi.org/10.1038/nmat1493 (2005).
30. Vollrath, F., Holtet, T., Thøgersen, H. C. & Frische, S. Structural organization of spider silk. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **263**, 147–151. https://doi.org/10.1098/rspb.1996.0023 (1996).
31. Kovoor, J. & Zylberberg, L. Fine structural aspects of silk secretion in a spider (Araneus diadematus). I. Elaboration in the pyriform glands. *Tissue Cell* **12**, 547–556. https://doi.org/10.1016/0040-8166(80)90044-0 (1980).
32. Kovoor, J. & Zylberberg, L. Fine structural aspects of silk secretion in a spider. II. Conduction in the pyriform glands. *Tissue Cell* **14**, 519–530 (1982).
33. Moon, M.-J. & Kim, W.-K. Ultrastructure of the ampullate glands in the orb web spider, Nephila clav ata L. Koch III. Excretory duct of the small ampullate gland. *Appl. Microsc.* **19**, 49–58 (1989).
34. Moon, M.-J. & Kim, W.-K. Ultrastructure of the ampullate gland in the orb web spider, Nephila clavata L. Koch IV. Secretory portion of the small ampullate gland. *Appl. Microsc.* **19**, 59–69 (1989).
35. Moon, M. & Kim, W. Fine structural study on the capture thread-producing organs in Nephila clavata L. Koch (Araneae: Araneidae), 1; Aggregate glands. *The Korean Journal of Zoology (Korea Republic)* (1989).
36. Moon, M. & Kim, W. Ultrastructural study on the tubuliform glands in Nephila clavata L. Koch (Araneae: Araneidae). *Korean Arachnol.* **5**, 43–55 (1989).
37. Moon, M. & Kim, W. Fine structural study on the capture threads-producing organs in Nephila clavata L. Koch (Araneae: Araneidae), 2; Flagelliform glands. *The Korean Journal of Zoology (Korea Republic)* (1990).
38. Moon, M. & Kim, W. Ultrastructure of the aciniform glands in Nephila clavata L. Koch (Araneae: Araneidae). *Korean Arachnol.* **5**, 195–206 (1990).
39. Domigan, L. J. *et al.* Carbonic anhydrase generates a pH gradient in Bombyx mori silk glands. *Insect Biochem. Mol. Biol.* **65**, 100–106. https://doi.org/10.1016/j.ibmb.2015.09.001 (2015).
40. Supuran, C. T. *Metabolites* Vol. 8, 25–25 (Multidisciplinary Digital Publishing Institute, 2018).
41. Nocentini, A., Supuran, C. T. & Capasso, C. An overview on the recently discovered iota-carbonic anhydrases. *J. Enzyme Inhibit. Med. Chem.* **36**, 1988–1995 (2021).
42. Davenport, H. W. Gastric carbonic anhydrase. *J. Physiol.* **97**, 32–43. https://doi.org/10.1113/jphysiol.1939.sp003790 (1939).
43. Davenport, H. W. The inhibition of carbonic anhydrase and of gastric acid secretion by thiocyanate. *Am. J. Physiol.-Legacy Content* **129**, 505–514 (1940).
44. Purkerson, J. M. & Schwartz, G. J. The role of carbonic anhydrases in renal physiology. *Kidney Int.* **71**, 103–115. https://doi.org/10.1038/sj.ki.5002020 (2007).
45. Ridderstråle, Y. *The Carbonic Anhydrases* 133–144 (Springer, 1991).
46. Moon, M.-J. & Tillinghast, E. K. Silk production after mechanical pulling stimulation in the ampullate silk glands of the barn spider, Araneus cavaticus. *Entomol. Res.* **34**, 123–130. https://doi.org/10.1111/j.1748-5967.2004.tb00101.x (2004).
47. Elghetany, M. T. & Saleem, A. Methods for staining amyloid in tissues: A review. *Stain Technol.* **63**, 201–212. https://doi.org/10.3109/10520298809107185 (1988).
48. Kenney, J. M., Knight, D., Wise, M. J. & Vollrath, F. Amyloidogenic nature of spider silk. *Eur. J. Biochem.* **269**, 4159–4163. https://doi.org/10.1046/j.1432-1033.2002.03112.x (2002).
49. Hijirida, D. H. *et al.* 13C NMR of Nephila clavipes major ampullate silk gland. *Biophys. J.* **71**, 3442–3447. https://doi.org/10.1016/S0006-3495(96)79539-5 (1996).
50. Hronska, M., van Beek, J. D., Williamson, P. T., Vollrath, F. & Meier, B. H. NMR characterization of native liquid spider dragline silk from Nephila edulis. *Biomacromolecules* **5**, 834–839. https://doi.org/10.1021/bm0343904 (2004).
51. Jenkins, J. E., Holland, G. P. & Yarger, J. L. High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft Matter* **8**, 1947–1954 (2012).
52. Tillinghast, E. K. & Townley, M. A. Silk glands of araneid spiders selected morphological and physiological aspects. *Silk Polym.* https://doi.org/10.1021/bk-1994-0544.ch003 (1994).
53. Park, J. G. & Moon, M. J. Fine structural analysis on triad spinning spigots of an orb-web spider's capture threads. *Entomol. Res.* **44**, 121–129 (2014).
54. Peters, H. M. *Ecophysiology of Spiders* 187–202 (Springer, 1987).
55. Moon, M.-J. & Kim, T.-H. Microstructural analysis of the capture thread spinning apparatus in orb web spiders. *Entomol. Res.* **35**, 133–140. https://doi.org/10.1111/j.1748-5967.2005.tb00149.x (2005).

56. Moon, M. J. Fine structure of the aggregate silk nodules in the orb-web spider Nephila clavata. *Anim. Cells Syst.* **22**, 421–428. https://doi.org/10.1080/19768354.2018.1546227 (2018).
57. Tillinghast, E. K. & Chase, S. F. Phosphorylase activity and glycogen utilization in the aggregate duct of orb weaving spiders. *Compar. Biochem. Physiol. B Biochem.* https://doi.org/10.1016/0305-0491(85)90113-0 (1985).
58. Zylberberg, L. & Kovoor, J. Ultrastrueture du Canal des Glandes Agreges et Flagelliformes d'Araneus diadematus Clerek (Araneae, Araneidae). *J. Zoomorphologie* **239**, 217–239 (1979).
59. Ayoub, N. A. *et al.* Protein composition and associated material properties of cobweb spiders' gumfoot glue droplets. *Integr. Compar. Biol.* **61**, 1459–1480. https://doi.org/10.1093/icb/ocan086 (2021).
60. Ayoub, N. A. *et al.* Orb weaver aggregate glue protein composition as a mechanism for rapid evolution of material properties. *Front. Ecol. Evol.* https://doi.org/10.3389/fevo.2023.1099481 (2023).
61. Jain, D. *et al.* Composition and function of spider glues maintained during the evolution of cobwebs. *Biomacromolecules* **16**, 3373–3380. https://doi.org/10.1021/acs.biomac.5b01040 (2015).
62. Vollrath, F. & Tillinghast, E. K. Glycoprotein glue beneath a spider web's aqueous coat. *Naturwissenschaften* **78**, 557–559. https://doi.org/10.1007/BF01134447 (1991).
63. Tillinghast, E. K., Townley, M. A., Wight, T. N., Uhlenbruck, G. & Janssen, E. The adhesive glycoprotein of the orb web of Argiope aurantia (Araneae, Araneidae). *MRS Online Proc. Libr.* **292**, 9–23. https://doi.org/10.1557/PROC-292-9 (1992).
64. Herrera, E., Nguyen, L. T., Escobar, E., Ouriel, W. & Casem, M. L. Alteration of tubuliform silk gland cytoarchitecture with the reproductive cycle of the Western black widow spider, Latrodectus Hesperus. *Invertebr. Biol.* **134**, 332–340. https://doi.org/10.1111/ivb.12106 (2015).
65. Moon, M.-J. Changes of tubuliform silk glands during the cocoon production in the garden spider, Argiope aurentia. *Appl. Microsc.* **28**, 539–549 (1998).
66. Greco, G. *et al.* Properties of biomimetic artificial spider silk fibers tuned by PostSpin bath incubation. *Molecules* **25**, 3248–3248. https://doi.org/10.3390/molecules25143248 (2020).
67. Hagn, F. *et al.* A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* **465**, 239–242. https://doi.org/10.1038/nature08936 (2010).
68. Askarieh, G. *et al.* Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **465**, 236–238. https://doi.org/10.1038/nature08962 (2010).
69. Schwarze, S., Zwettler, F. U., Johnson, C. M. & Neuweiler, H. The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. *Nat. Commun.* **4**, 2815. https://doi.org/10.1038/ncomms3815 (2013).
70. Landreh, M. *et al.* A pH-dependent dimer lock in spider silk protein. *J. Mol. Biol.* **404**, 328–336. https://doi.org/10.1016/j.jmb.2010.09.054 (2010).
71. Kono, N. *et al.* Orb-weaving spider Araneus ventricosus genome elucidates the spidroin gene catalogue. *Sci. Rep.* **9**, 8380. https://doi.org/10.1038/s41598-019-44775-2 (2019).
72. Babb, P. L. *et al.* Characterization of the genome and silk-gland transcriptomes of Darwin's bark spider (Caerostris darwini). *PloS One* **17**, e0268660–e0268660. https://doi.org/10.1371/journal.pone.0268660 (2022).
73. Babb, P. L. *et al.* The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nat. Genet.* **49**, 895–903. https://doi.org/10.1038/ng.3852 (2017).
74. Kono, N. *et al.* Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proc. Natl. Acad. Sci. U. S. A.* https://doi.org/10.1073/pnas.2107065118 (2021).
75. Kono, N. *et al.* Darwin's bark spider shares a spidroin repertoire with Caerostris extrusa but achieves extraordinary silk toughness through gene expression. *Open Biol.* **11**, 210242. https://doi.org/10.1098/rsob.210242 (2021).
76. Hu, W. *et al.* A molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk. *Nat. Commun.* **14**, 837. https://doi.org/10.1038/s41467-023-36545-6 (2023).
77. Di Fiore, A., D'Ambrosio, K., Ayoub, J., Alterio, V. & De Simone, G. *Carbonic Anhydrases* 19–54 (Elsevier, 2019).
78. Babb, P. L. *et al.* The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nat. Genet.* https://doi.org/10.1038/ng.3852 (2017).
79. Abelein, A. *et al.* High-yield production of amyloid-beta peptide enabled by a customized spider silk domain. *Sci. Rep.* **10**, 235. https://doi.org/10.1038/s41598-019-57143-x (2020).
80. Puchtler, H., Sweat, F. & Levine, M. On the binding of Congo red by amyloid. *J. Histochem. Cytochem.* **10**, 355–364 (1962).
81. Westermark, P. Subcutaneous adipose tissue biopsy for amyloid protein studies. *Methods Mol. Biol.* **849**, 363–371. https://doi.org/10.1007/978-1-61779-551-0_24 (2012).

Acknowledgements

The authors would like to thank Henrik Biverstål (Karolinska Institutet) for providing the amyloid fibrils which were used as controls for thioflavin-S staining and Shaffi Manchanda (Karolinska Institutet) for suggestions regarding thioflavin-S staining. Spider species was identified by Gabriele Greco (SLU) and the Museum of Natural History, Stockholm. This work was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 815357), Formas (2019-00427), and Olle Engkvist Stiftelse (207-0375).

Author contributions

A.R. and L.H. developed the study concept. L.H., A.R., and S.S. designed the experiments. L.H. and S.S. performed the experiments and collected data. The data interpretation was done by S.S., L.H. and A.R. The Congo red staining, analysis and interpretation was done by P.W. S.S. wrote the original draft of the manuscript which was reviewed, revised, and further improved by L.H., A.R. and P.W.

Funding

Open access funding provided by Swedish University of Agricultural Sciences. This work was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 815357), Formas (2019–00427), and Olle Engkvist Stiftelse (207–0375).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-49587-z.

Correspondence and requests for materials should be addressed to L.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

II

Origin, Structure and Composition of the Spider Major Ampullate Silk Fiber revealed by Genomics, Proteomics, and Single-Cell and Spatial Transcriptomics

Authors and affiliations

Sumalata Sonavane[1], Sameer Hassan[2], Urmimala Chatterjee[2], Lucile Soler[3,4], Lena Holm[1], Annelie Mollbrink[5], Gabriele Greco[1], Noah Fereydouni[2], Olga Vinnere Pettersson[6], Ignas Bunikis[6], Allison Churcher[7], Henrik Lantz[3,4], Jan Johansson[2], Johan Reimegård[3*], Anna Rising[1,2*]

[1]*Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden.*
[2]*Department of Biosciences and Nutrition, Karolinska Institutet, Neo, Huddinge, Sweden.*
[3]*National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory (SciLifeLab), Uppsala University, Uppsala, Sweden.*
[4]*Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden.*
[5]*Department of Gene Technology, KTH Royal Institute of Technology, SciLifeLab, Sweden.*
[6]*Department of Immunology, Genetics and Pathology, National Genomics Infrastructure SciLifeLab, Uppsala, Sweden.*
[7]*Department of Molecular Biology, NBIS, SciLifeLab, Umeå University, Umeå, Sweden.*
Sweden.
**Corresponding authors*

Abstract

Spiders produce nature's toughest fiber using renewable components at ambient temperatures and with water as solvent, making it highly interesting to replicate for the materials industry. Despite this, much remains to be understood about the bioprocessing and composition of spider silk fibers. Here, we identify 18 proteins that make up the spiders' strongest silk type, the major ampullate fiber. Single-cell RNA sequencing and spatial transcriptomics revealed that the secretory epithelium of the gland harbors six cell types. These cell types are confined to three distinct glandular zones that produce specific combinations of silk proteins. Interestingly, image analysis of histological sections showed that the secretions from the three zones do not mix, and proteomic analysis revealed that these secretions form layers in the final fiber. Using a multi-omics approach, we provide substantial advancements in the understanding of the structure and function of the major ampullate silk gland as well as of the architecture and composition of the fiber it produces.

Teaser

Unveiling silk gland secrets: Multi-omics expose spider silk's intricate three-layered structure and unique protein makeup.

Introduction

Spider major ampullate silk, distinguished by its exceptional mechanical properties, is a product of renewable resources and is synthesized within fractions of a second under ambient conditions (*1*). Its unique combination of rapid production, sustainability, and superior performance characteristics positions spider silk as a compelling environmentally friendly material for high-performance fiber applications (*2*). To unlock its potential, the development of artificial replicas of the pristine silk fiber is imperative, which necessitates a comprehensive understanding of the bioprocessing of the silk fiber.

During the past 400 million years, spiders have refined their silk spinning abilities, and today, individual spiders of some species can spin seven different types of silks from specialized glands located in the abdomen (opisthosoma) (3) (Fig. 1A). These include major ampullate silk for making the radial threads and frame of the web, minor ampullate silk for making the inner spiral, flagelliform silk for making the extensible capture spiral that is coated by sticky droplets of aggregate silk, pyriform silk for making attachment discs, tubuliform silk for making the outer layer of the egg sac, and aciniform silk for making the inner part of the egg sac and/or swathing prey (4). The different silk types are produced in separate sets of glands, which have secretory ducts connected to spigots located caudoventrally in the abdomen (5). The spider silks are primarily composed of spider silk proteins (spidroins), characterized by a distinctive architectural arrangement that features non-repetitive N- and C-terminal domains flanking an extensive repetitive region (6-13). The terminal domains, herein referred to as NT and CT, respectively, are evolutionarily conserved (14), while the nature of the repetitive regions differs between the different spidroin types (7-11).

The major ampullate silk is the strongest of the different silks, and the main components of this fiber are called major ampullate spidroins (MaSps) (*7-13*). The repetitive regions of the MaSps are rich in glycine and alanine that organize into typical motifs, which, together with clustering of the terminal domains in phylogenetic analyses, have led to the classification of MaSps into five types (MaSp1-5). MaSp1 (*15*) typically have poly-Ala blocks interspersed with Gly-rich repeats. MaSp2 (*16*) repeats are rich in proline and glutamine with abundant Gly-Pro-Gly and Gln-Gln motifs that MaSp1 lacks. MaSp3 (*17*) is similar to MaSp1 but is void of the Gly-Gly motifs commonly found in MaSp1. In contrast, MaSp4 and MaSp5 (*18*) repeats lack poly-Ala blocks, but MaSp4 contains Gly-Pro-Gly motifs common to MaSp2 whereas MaSp5 have Gly-Gly motifs like MaSp1. Evidently, the major ampullate fiber is a composite of different MaSps, which has been suggested to be related to the supreme mechanical properties of the fiber (*19*). For example, MaSp1 likely contributes primarily to fiber strength, while MaSp2 is more related to the fiber strain (*12, 20*), and expression of MaSp3 and MaSp4 have been suggested to be correlated to fiber toughness (*12, 18*). Although recent work has extended our understanding of the spider genomes (*7-11, 13, 21*), the spidroin genes are difficult to assemble because they contain extended stretches of repeats and are often found in tandem arrays in the genome. Therefore, it appears that previous descriptions of spidroin gene repertoires are incomplete.

In addition to MaSps, the major ampullate silk contains several proteins that do not conform to the overall architecture of a classical spidroin. These proteins form a heterogeneous group termed spider silk-constituting elements (SpiCE) that differ in molecular weight and amino acid composition (*8-10*). A subgroup within this group of proteins, known as cysteine-rich proteins (CRP), stands out for their high cysteine content (*22*), and are suggested to have a structural role in the fiber, possibly by forming large disulphide-linked complexes (*23*).

The major ampullate silk is produced in paired glands, each consisting of a long tubular tail, a wide sac, and a narrow s-shaped duct (*24*) (Fig. 1B). The lumen of the duct is lined by a cuticular intima, a hard and chitin-rich structure believed to aid in the removal of water during the spinning process and to protect against tissue laceration during spinning (*24, 25*). Histological studies of the gland of distantly related spider species have shown that the tail and sac are composed of three types of columnar epithelial

cells confined to three different zones (A, B and C) (*26*). Zone A cells are found in the tail and proximal parts of the sac (toward the tail), the middle part of the sac harbors zone B cells, while the most distal part of the sac next to the duct is composed of zone C cells (Fig. 1B). According to immuno-transmission electron microscopy (TEM) studies performed using an antibody that identified the conserved spidroin NT domain, zone A and B cells produce spidroins, while the nature of the secretion produced by zone C cells remains uncharacterized (*26*). Investigations of the major ampullate silk have shown that the fiber has a core-skin structure (*26-33*), wherein the central core contains the spidroins (*32, 34*) while the surrounding skin layer is composed of unknown proteins and other components like carbohydrates, glycoproteins and lipids (*30, 32, 33*). Consequently, the compositions of the secretions from zone A, B and C are largely unknown, and the link between the epithelial zones and the architecture of the fiber remains to be revealed.

In the gland, the spidroins are stored as a soluble and viscous dope at high concentrations (~50% w/v) (*35*) and at around neutral pH (*36*). During storage, the MaSps are dimeric, interconnected in the CT domain (*37*). As the dope travels from the sac and through the duct, the pH is lowered from ca. 6.3 to < 5.7, which induces antiparallel dimerization of the NT domain (*36, 38, 39*). Since the MaSps are already interconnected via their CT domains (*36, 37*), the dimerization of NT leads to the interlocking of the spidroins into large networks (*40*). At the same time, shear forces acting on the dope in the narrowing duct combined with the increasingly acidic environment destabilize the CT domain, which transitions into β-sheet conformation (*36, 37*). When the fiber is pulled out from the spider (*41*), the pulling forces will be propagated along the protein chains (since they are linked in macromolecular networks), which aids in the alignment and conversion of the poly-Ala blocks in the repetitive region into β-sheet crystals (*1, 40*).

Here, we present a complete spidroin repertoire from an orb-weaving spider, *Larinioides sclopetarius*. Spatial transcriptomics and single-cell RNA (scRNA) sequencing showed that the tail and the sac of the major ampullate gland are composed of six silk-producing cell types found in the three distinct zones (A, B and C). This, combined with image analyses of histological sections and proteomics analyses of sequentially dissolved fibers, indicate that the silk fiber is a three-layered structure in which the core is composed of proteins belonging to the MaSp1, 2 and 4 families, the middle layer is dominated by MaSp3, and the outermost layer contains several uncharacterized non-spidroin proteins. Thus, we present a detailed description of the structure and composition of the major ampullate silk gland and fiber, in which we successfully link the layered epithelial secretions in the gland to those in the silk fiber. Our results bridge the long-standing knowledge gaps necessary for the production of biomimetic artificial silk fibers.

Results

The *L. sclopetarius* major ampullate fiber is composed of 18 silk proteins

In this study we focus on the Swedish bridge spider (*L. sclopetarius*) that produces major ampullate silk fibers with impressive mechanical properties (fig. S1). First, the genome of this spider species was sequenced, assembled, and annotated (Table 1, Supplementary Notes, tables S1 to S11, and fig. S2). Manual curation resulted in a catalog of 35 complete spidroin genes (Fig. 1 and figs. S3 to S10).

Table 1: Assembly and annotation statistics of *Larinioides sclopetarius* genome.

Genome assembly	
Assembly size (bp)	2274070471
GC %	30.5
Number of Contigs	1602
Longest Contig (bp)	29089857
N50 (bp)	5487611
N90 (bp)	1366626
BUSCO complete (%)	97.3

Repeat statistics	
Number of elements	2761064
Length (bp) [% Genome]	47.50%

Genome annotations	
Protein-coding genes	22860
Transcript Isoforms	46911
BUSCO complete (%)	96.8

Next, to reveal the protein composition of the major ampullate gland and silk fiber, respectively, proteomic analysis was performed using liquid chromatography-tandem mass spectrometry (LC-MS/MS) (table S1). In the gland samples, 3985 proteins were identified in at least one of the three replicates (table S12). Dissolving the silk fiber can be challenging and different chemical treatments can extract different proteins (*22*). Therefore, three solvents, urea, hexafluoroisopropanol (HFIP) and lithium bromide (LiBr), were used to extract proteins from the fibers. A total of 1835 proteins were identified (table S12), of which 325 were identified in all three biological replicates from at least one treatment. However, since the fibers are easily contaminated by other silk types during spinning and by unrelated proteins during handling, only the proteins that were also found in the major ampullate gland proteomic data, were predicted to have a signal peptide, and with spectral counts exceeding 0.1% of the total spectra in LC-MS/MS analysis were considered (Fig. 2A). This process rendered 18 silk proteins that were named “the 18 silk proteins” (Table 2 and table S13) and corresponding genes were named “the 18 silk genes” (Table 2). The 18 silk proteins included ten MaSps, two ampullate spidroin-like proteins (AmSp-like1 and AmSp-like2) and six proteins with unknown functions (table S13). The six proteins of unknown function were annotated as spider silk-constituting elements (*8-10*) of *Larinioides* major ampullate silk (SpiCE-LMa1 to SpiCE-LMa6) in order of abundance. LC-MS/MS analysis of fibers that were completely dissolved showed that peptides mapping to MaSp1 (a–c), MaSp2 (b, c, e, f), MaSp3 (a, b) and MaSp4 represented 46%, 20%, 30% and 1%, respectively, when considering percent of total spectra (Fig. 2A). Peptides mapping to SpiCE-LMa proteins accounted for 2% while the AmSp-like proteins only constituted 1% (Fig. 2A). It should be noted that the large difference in molecular weight between the 18 proteins, and the extreme repetitiveness and sequence similarity of the MaSps are problematic when quantifying the protein composition of these samples. We therefore analyzed the data with three additional methods: (1) intensity-Based Absolute Quantification (iBAQ), (2) Normalized Spectral Abundance Factor (NSAF), and (3) Label-Free Quantification (LFQ) intensity. The estimates of the composition of the fiber obtained from these four methods slightly differ (fig. S16). However, the order of abundance of MaSps and SpiCE-LMa proteins was similar across the methods, i.e., MaSp1, MaSp2, and MaSp3 were always the three most abundant classes of proteins and the

SpiCE-LMa proteins only contributed a small fraction. The methods based on intensity (iBAQ and LFQ) suggested that MaSp2 proteins were the most abundant whereas the MaSp1 proteins were the most abundant when using the methods based on spectral counts (total spectra and NSAF). Consequently, the intra-sample protein composition estimations reported herein, that are derived from label-free LC-MS/MS analyses, should be interpreted with caution. Comparisons across samples (see below), on the other hand, are more reliable since technical biases are the same.

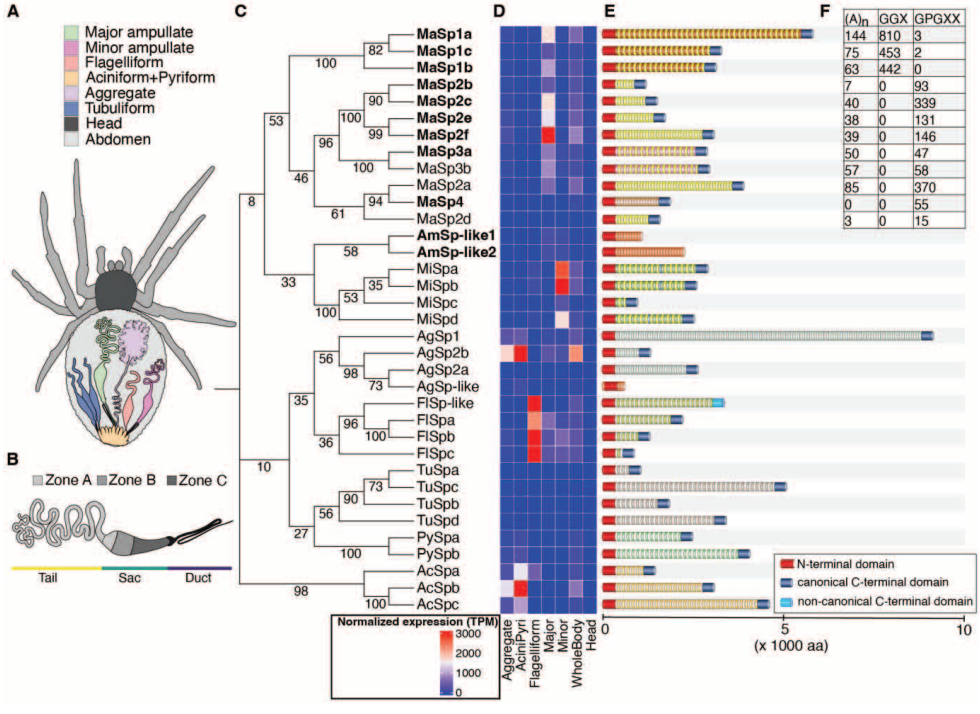

Figure 1: Spidroin catalogue from *Larinioides sclopetarius*. (A) Schematic view of an orb-weaving spider with one set of each type of silk gland indicated. **(B)** Schematic figure of the major ampullate gland. The gland has three anatomical parts: the tail (yellow), the sac (green), and the duct (blue). The tail and sac are composed of a single layered epithelium in which three morphologically distinct cell types are found, each localized to one of three zones (A–C), indicated as shades of grey. **(C)** Phylogenetic tree of the NT domain from the 35 spidroins identified in *L. sclopetarius*. Numbers on the branches indicate bootstrap values. Spidroins in bold were identified by proteomics analysis of the major ampullate gland and silk. **(D)** Heat map with the expression of all spidroins in different tissues as determined by bulk RNA sequencing. The colors correspond to transcripts per million shown in the scale (bottom left inset). **(E)** Schematic illustration of the spidroin genes. All spidroin genes encoded proteins with a signal peptide (not shown) and an NT domain (red block). Most spidroin genes encoded a canonical CT domain (dark blue block), except FlSp-like which was found to have a non-canonical CT domain (light blue block), and AmSp-like1 and AmSp-like2 and AgSp-like spidroins which completely lacked the C-terminal non-repetitive region. The repetitive motifs in each spidroin are represented as colored blocks. **(F)** Table showing the number of typical MaSp repeat motifs found in each of the MaSps. $(A)_n$ refers to poly-alanine motifs, X in GGX and GPGXX represents any amino acid residue.

The six SpiCE-LMa proteins generally had lower molecular weight compared to the MaSps (table S13). In terms of amino acid composition, the SpiCE-LMa1 and SpiCE-LMa2 resembled the MaSp proteins and SpiCE-LMa2 had spidroin-like repeat motifs. SpiCE-LMa3 to SpiCE-LMa6 were rich in Cys and resembled the amino acid composition of the AmSp-like1 and AmSp-like2 (Fig. 2E and figs. S17 to S19). The predicted secondary structure content and AlphaFold2 structural predictions (*42, 43*) of the SpiCE-LMa proteins suggested large heterogeneity within this group of proteins (figs. S20 to S26).

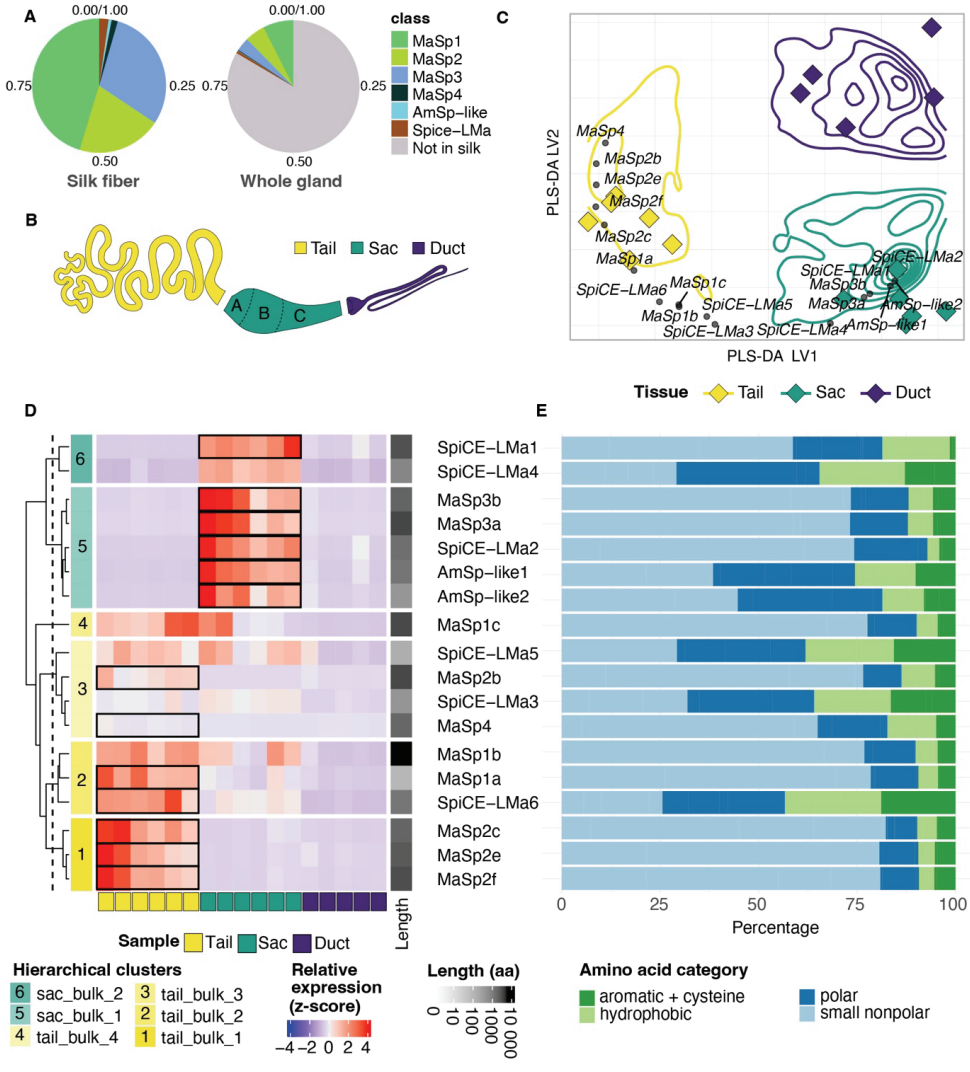

Figure 2: Expression of the 18 silk genes in the major ampullate tail, sac, and duct. (**A**) Relative quantification of the peptides identified in the major ampullate gland and in silk fibers, dissolved in 2–8M urea, using LC-MS/MS proteomics, colored according to protein classes (MaSp1, MaSp2, MaSp3, MaSp4, AmSp-like, and SpiCE-LMa). (**B**) Schematic figure of the gland showing the three parts (tail, sac, and duct), that were separated for RNA sequencing experiments. (**C**) PLS-DA analysis of the bulk RNA data separates the three different parts (tail, sac, and duct). The *x* and *y* axes represent the values or the loading of genes on the first and second latent variables, LV1 & LV2, respectively. Colors indicate different sample types (tail/sac/duct) as shown in (B). Diamonds represent samples, small circles represent the overlay of the 18 silk genes. (**D**) Heatmap showing the relative expression levels (z-score) for the 18 silk genes in the tail, sac, and duct samples (indicated as yellow, green, and blue, respectively, in *x* axis). Gene names are shown in the right. Hierarchical clustering separates the genes into six subclusters based on their expression profiles shown as a dendrogram on the *y* axis. Black boxes indicate upregulated differentially expressed genes ($P < 0.05$). The six subclusters identified are colored as shades of yellow or green based on whether the highest gene expressions were in the tail or sac parts, respectively. The bar on the right indicates protein length. The scale ranges from 0 (white) to 10000 aa residues (black). (**E**) Percentage of different categories of amino acid residues in each of the 18 silk proteins (%). The colors indicate different categories (small nonpolar: A, G, P, S, T; hydrophobic: I, L, M, V; polar: D, E, H, K, N, Q, R; aromatic and cysteine: C, F, W, Y).

Table 2: Ranking of the 18 silk proteins/genes as markers in different methods used.

* DEG analysis on tail, sac, and duct parts of the major ampullate glands. Rank is based on average log_2-fold change. The higher the rank the more differential expressed the gene is. ** Marker gene identification on spots manually annotated as different zones within the spatial transcriptomic sections. Rank is based on average log_2-fold change, the higher the better. *** Marker gene identification in cell types identified with single-cell data. Rank is based on average log_2-fold change, the higher the better.

	Differential gene expression					
	Bulk*		Spatial**		Single-cell***	
Gene	**part**	**rank**	**zone**	**rank**	**celltype**	**rank**
MaSp2b	Tail	19	Zone A	31	ZoneA_MaSp2	3
MaSp2c	Tail	9	Zone A	3	ZoneA_MaSp2	2
MaSp2e	Tail	6	Zone A	6	ZoneA_MaSp2	4
MaSp2f	Tail	1	Zone A	5	ZoneA_MaSp2	1
MaSp4	Tail	34			ZoneA_MaSp2	12
MaSp1a	Tail	38	Zone A	1	ZoneA_MaSp1	1
					ZoneA_MaSp2	13
MaSp1b			Zone A	4	ZoneA_MaSp1	2
MaSp1c			Zone A	32	ZoneA_MaSp1	3
SpiCE-LMa6			Zone A	12	ZoneA_SpiCE-LMa	6
SpiCE-LMa3					ZoneA_SpiCE-LMa	28
SpiCE-LMa5					ZoneA_SpiCE-LMa	11
MaSp3a	Sac	19	Zone B	2	ZoneB_MaSp3	5
MaSp3b	Sac	18	Zone B	3	ZoneB_MaSp3	1
AmSp-like2	Sac	42	Zone B	4	ZoneB_MaSp3	8
			Zone C	61		
AmSp-like1	Sac	49	Zone B	9	ZoneB_MaSp3	12
SpiCE-LMa4	Sac	39	Zone B	72	ZoneB_MaSp3	68
					ZoneC_SpiCE-LMa	85
SpiCE-LMa2	Sac	63	Zone B	23	ZoneB_MaSp3	6
			Zone C	16	ZoneC_SpiCE-LMa	11
SpiCE-LMa1	Sac	68	Zone C	13	ZoneC_SpiCE-LMa	1

The 18 silk genes are expressed in the tail and the sac

Bulk RNA sequencing of whole major ampullate glands was used to verify that the RNA levels correlated with the protein abundance in the gland (fig. S27A). Notably, the 18 silk genes ranked among the most highly expressed genes (fig. S28A). However, this analysis did not permit localization of the expression of specific genes to different parts of the gland. To enhance the resolution of the gene expression profiles, the major ampullate gland was sequenced after being cut into three distinct anatomical parts: tail, sac, and duct (Fig. 2B). This implies that the tail samples contained transcripts from zone A while the sac samples contained transcripts from all three zones (Figs. 1B and 2B, and table S1). The major ampullate gene set (table S14) was identified by principal component analysis (PCA) of the enriched genes in the gland (fig. S28). This gene set was then used to perform Partial Least Square Discriminant Analysis (PLS-DA) regression on the tail, sac, and duct samples. This analysis separated the three types of samples into distinct clusters using only two latent variables with high predictive relevance ($Q^2 = 0.857$), verifying that the transcriptome in the tail, sac and duct are diverse (Fig. 2C). To identify which part of the major ampullate gland the 18 silk proteins were originating from, the 18 silk genes were superimposed on the PLS-DA plot which revealed that all the MaSp genes were expressed in the major ampullate tail, except *MaSp3a* and *MaSp3b* which were expressed in the sac along with *AmSp-like1* and *AmSp-like*2 (Fig. 2C). Three genes encoding proteins of unknown function, *SpiCE-LMa1, SpiCE-LMa2*, and *SpiCE-LMa4* were also among genes expressed in the sac. The expression profiles of the silk genes in the three parts (tail/sac/duct) were further visualized by a heatmap (Fig. 2D), which clearly indicated that the genes encoding the 18 silk proteins are expressed in the tail and the sac but not in the duct.

Hierarchical clustering grouped the 18 silk genes into 6 clusters based on the similarity in their expression profiles (Fig. 2D and fig. S27B). Of these, cluster 1 genes showed differential expression in the tail samples and were assigned to the tail, while all the genes in clusters 5 and 6 showed differential expression in sac samples and were assigned to the sac. Clusters 2, 3, and 4 had some genes that were differentially expressed in the tail and clustered in the same node as the cluster 1, and therefore were assigned to the tail. The tail clusters are colored as shades of yellow while sac clusters are shown as shades of green. Based on these results, cluster 1 that contained *MaSp2*, *c*, *e*, and *f* genes, which were solely expressed in the tail, could be assigned to zone A (Fig. 1B and 2D) but the rest of the genes, which were also expressed in the sac, could not be assigned to one of the zones since the sac samples contained tissue from all the three zones.

Expression of the silk genes is spatially resolved in the three zones

To improve the resolution of the silk gene expression, we next used spatial transcriptomics (10X Visium). This is an unbiased and elegant technique that allows mapping of the gene expression in tissue sections with a resolution of 50 mm (*44*) (Supplementary Notes). Six sections of the whole abdomen from four female individuals were used and the spots on the hematoxylin and eosin (H&E) stained sections were manually annotated to different silk glands based on histology (one section is shown in Fig. 3, A and B; figs. S12 and S13 show additional sections). The sequencing data from all spots from all the sections were then isolated and visualized using Uniform Manifold Approximation and Projection (UMAP) (*45*). In the UMAP, the spots assigned as silk glands clustered together and separated from the spots belonging to other tissue types (fig. S14). Within the silk gland cluster, the spots from different silk glands clustered together, in line with the manual annotation (Fig. 3C). On the spatial sections, the spidroin expression was specific to corresponding glands (fig. S15) and confirmed the results from the bulk-RNA expression profiles (Fig. 1D), attesting to the quality of the spatial data.

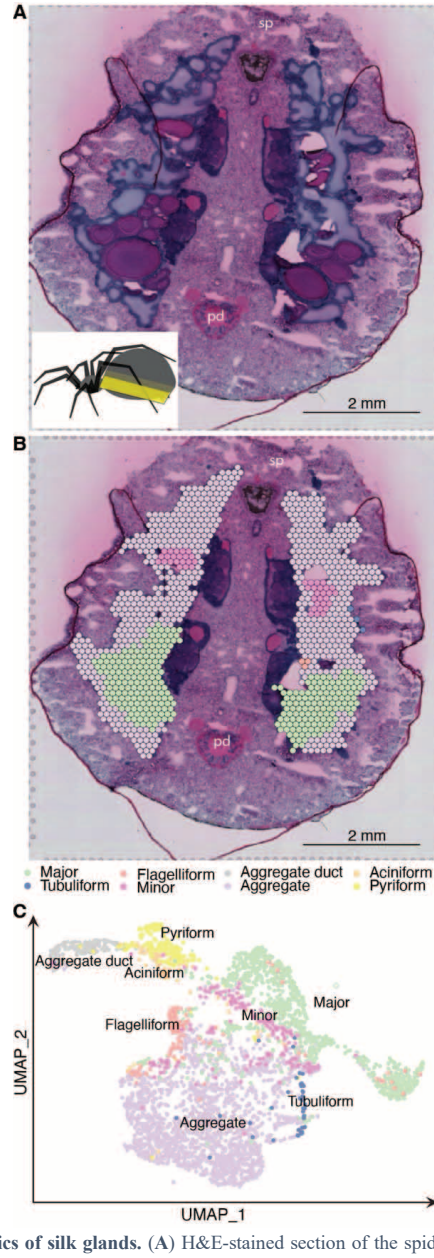

Figure 3: Spatial transcriptomics of silk glands. (**A**) H&E-stained section of the spider abdomen. The inset shows a top view of a lateral section of the abdomen with the approximate plane where the section was made. (**B**) The spots were annotated as different silk glands based on morphology of the tissue and the spatial location. In (A) and (B), sp and pd indicate the location of the spinnerets and the pedicel, respectively, scale bar, 2 mm. (**C**) UMAP analysis of all spots, from eight sections, that were manually annotated as silk glands. The axes (UMAP-1 and UMAP-2) represent the first and second UMAP dimensions, respectively. Each dot represents a spot in the spatial sections.

The spots annotated as covering major ampullate gland tissue could be further assigned as zone A, B or C based on the morphology of the epithelial cells in six of the sections (Fig. 4, A and B and high-resolution H&E-stained image of zones in fig. S29A). This resulted in 847 spots from six sections that clustered according to zone in the UMAP (fig. S29B), indicating that the expression profiles in these zones are indeed different. Furthermore, marker genes for zone A, B and C identified from the spatial data overlapped with the bulk-RNA PLS-DA plot of tail and sac but not the duct (fig. S30A). The expression of the 18 silk genes on all the spots annotated as zone A, B or C was then visualized as a heatmap (Fig. 4C). Statistical analyses of the expression levels revealed that 15 of the 18 silk genes were predominantly expressed in at least one of the zones (Table 2). *MaSp1 (a–c), MaSp2 (b, c, e, f)* and *SpiCE-LMa6* genes had significantly higher expression in zone A cells, *MaSp3 (a, b)*, *AmSp-like1*, and *SpiCE-LMa4* in zone B cells, and *SpiCE-LMa1* in zone C cells. The expression of *SpiCE-LMa2* and *AmSp-like2* were significantly higher in both zone B and zone C. The expression of the remaining three genes, *MaSp4*, *SpiCE-LMa3*, and *SpiCE-LMa5* did not differ significantly between the zones (Fig. 4C). The expression profiles of selected genes in one of the major ampullate glands in section 1 are shown as examples in Fig. 4D. Expression profiles of the 18 silk genes in all other spatial sections can be found in figs. S31 to S35.

The 18 silk genes in the three zones are specifically expressed in six cell types

Interestingly, the *MaSp1* and *MaSp2* genes were both expressed in the zone A cells, but the spatial transcriptomics analyses revealed that their expression profiles differ in different regions of zone A (Fig. 4 and figs. S31 to S35). This suggests that the epithelial zones could harbor several different cell types. To elucidate this, single-cell RNA sequencing was performed on whole major ampullate glands isolated from 7 individuals (table S1). After quality control (QC), filtering, and analysis, 9700 cells were obtained which clustered into eight groups. Based on the gene overlap with the first two components of PLS-DA analysis of the bulk RNA gene sets, these eight clusters were found to originate from either tail, sac, or duct of the gland (fig. S30B). Three of the cell clusters were specific for the tail, three for the sac and two for the duct. The clusters were then compared with the spatial transcriptomic data which allowed them to be classified as zone A/B/C cells based on their gene overlap with the marker genes of the three different zones (fig. S30C). Three clusters were identified in zone A, one in zone B, one in zone C and one cluster was found in all three zones. The clusters were assigned as cell types and named according to the zone and top marker spider silk protein gene (Fig. 5A and table S15). One of the zone A cell types had *MaSp1a–c* as the top three marker genes and was named as ZoneA_MaSp1. The second zone A cell type had *MaSp2b*, *c*, and *f* as the top three marker genes and was therefore annotated as ZoneA_MaSp2. This cell type included *MaSp4* as one of the top marker genes. In the last zone A cell type, *SpiCE-LMa3*, *SpiCE-LMa5*, and *SpiCE-LMa6* were the top marker genes among silk genes and the cell type was named ZoneA_SpiCE-LMa. In zone B cells, the top marker genes were *MaSp3a* and *MaSp3b* and the cell type was named as ZoneB_MaSp3 cells. These cells also had *AmSp-like1* and *AmSp-like2*, two of the spidroins lacking the CT domain (Fig. 1E), and *SpiCE-LMa2* and *SpiCE-LMa4* as top marker genes. The zone C cells had *SpiCE-LMa1*, *SpiCE-LMa2*, and *SpiCE-LMa4* as the top marker genes among the silk genes and hence named as ZoneC_SpiCE-LMa cells. The final cell type was mainly found in the sac but could not be assigned to a specific zone and was therefore named ZoneABC (fig. S30). The remaining two cell types were named according to their similarity in expression profile to the bulk-RNA data from the duct as Duct_1 and Duct_2 (fig. S30B), and these two cell types did not have any of the silk proteins as marker genes (Fig. 5B).

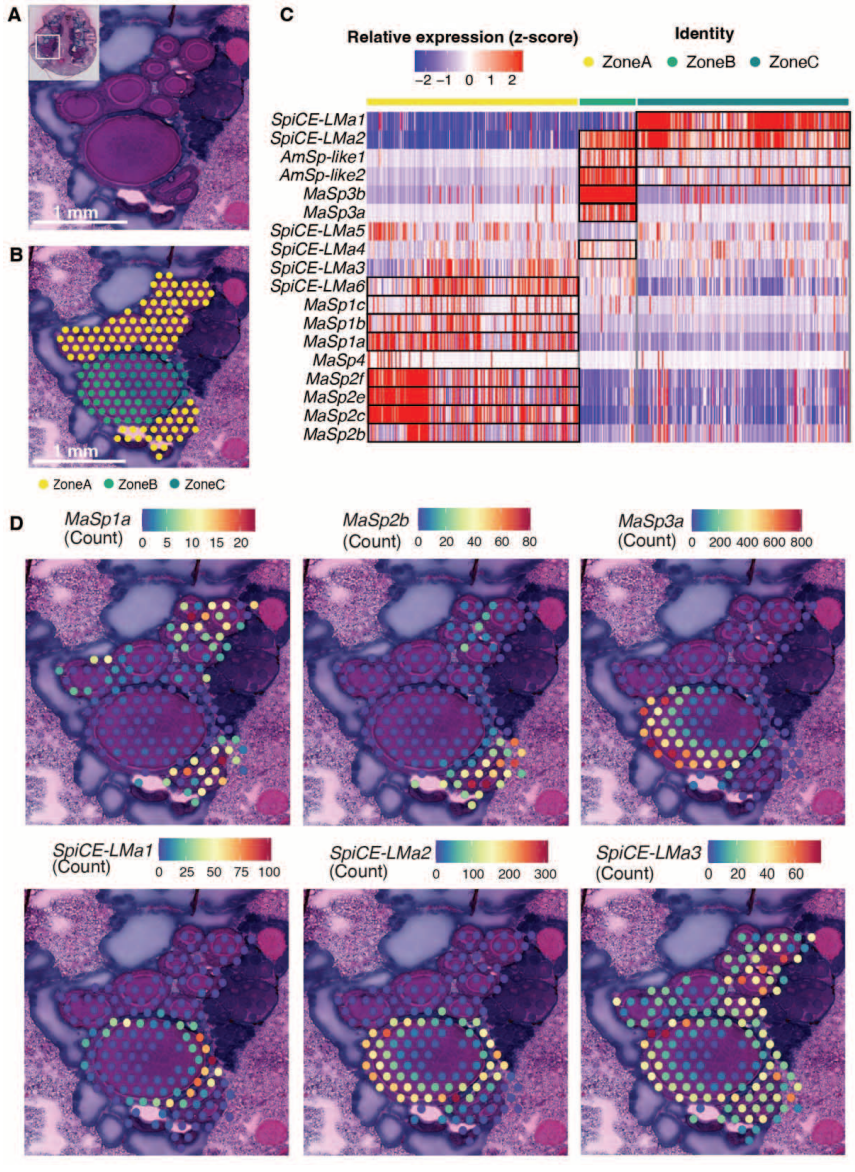

Figure 4: Spatial resolution of silk protein expression in the zone A, B and C. (**A**) One of the major ampullate glands in section 1 contains several cross sections of the tail and a cross sectioned sac (H&E staining). Inset shows the original image from which the region was magnified (white square). (**B**) The spots corresponding to the major ampullate gland were annotated as zone A, B or C based on the morphology and staining of the epithelium overlaying each spot. In (A) and (B), scale bar, 1 mm. (**C**) Heatmap showing the expression of the 18 silk genes in the 847 spots annotated as zone A, B and C, respectively. Each bar on the heatmap represents a spot on the spatial section and black boxes indicate marker genes ($P < 0.05$) in different zones. (**D**) Expression profiles of *MaSp1a, MaSp2b, MaSp3a, SpiCE-LMa1, SpiCE-LMa2* and *SpiCE-LMa3* (in order) in the three zones of major ampullate gland shown in (A) and (B).

A compelling observation is that all the 18 silk genes were among the top 100 marker genes in at least one of the six cell types when considering all the 22860 protein-coding genes. In most cases, they were among the top ten (Table 2 and Fig. 5B). All the cell types express a distinct set of silk genes as evidenced by that 15 of 18 silk genes were significantly expressed in only one of the six cell types, while the remaining three genes were expressed in two cell types (Table 2). In summary, the presence of the 18 silk proteins in the fiber can be directly related to the expression of the corresponding genes in the six cell types.

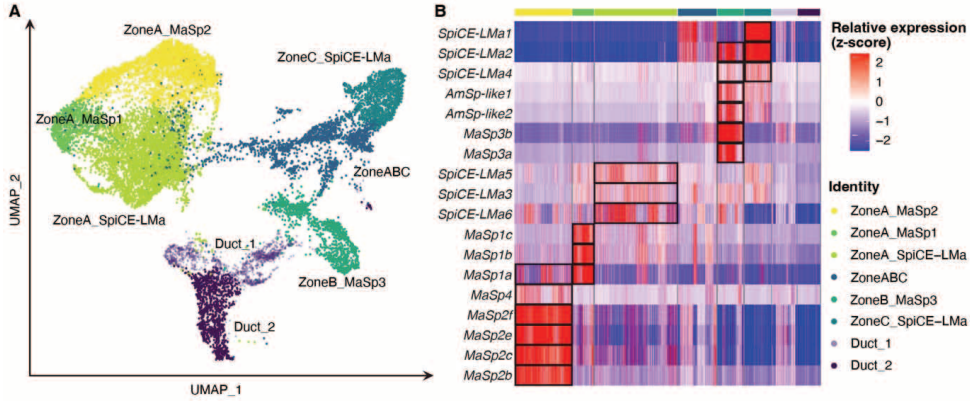

Figure 5: Single-cell RNA sequencing analysis of major ampullate gland reveals eight cell types. (**A**) Eight cell types in the major ampullate gland visualized in UMAP. The axes (UMAP_1 and UMAP_2) represent the first and second UMAP dimensions, respectively. Each dot represents a cell. The cell types were annotated by correlating the gene expression with the bulk RNA and spatial transcriptomic data. This revealed that six cell types make up the secretory epithelium of the tail and sac; three cell types were confined to zone A, one to zone B, one to zone C and one could be found in all three zones. Two cell types were found in the duct. (**B**) Heatmap of relative gene expression (z-score) of the 18 silk proteins in different single-cell types. Black boxes indicate the marker genes (P <0.05) in different cell types.

The six silk producing cell types can be spatially resolved along the gland

To identify the spatial location of the cell types in the major ampullate gland, the scRNAseq data was combined with spatial transcriptomics data and deconvolution of the two datasets was performed. This allowed us to visualize the spatial distribution of cell types in the spots annotated as major ampullate glands on the spatial sections (Fig. 6, A, B, and C and figs. S36 and S37). The pattern that emerged suggested that the cell types confined to zone A may not be evenly distributed along this zone. To address this, we took advantage of the tapering nature of the gland's tail to order the cross sections from proximal (small perimeter) to distal (large perimeter) and employed image analysis using QuPath (*46*). The sectioned major ampullate glands were annotated according to zones, as shown in Fig. 6B, resulting in 101 cross-sections for zone A, 2 for zone B, 9 for zone C and 5 for the duct across all sections. This annotation was confirmed by digital color deconvolution in QuPath which generates values for the degree of hematoxylin and eosin staining, respectively, for each area of interest (fig. S38). Furthermore, the perimeters of all the cross-sectioned parts of the gland were determined using QuPath and associated with the spots on the spatial transcriptomic sections. The spatial spots corresponding to each cross-section in zone A were extracted and the mean expression of marker genes in these spots as a function of the perimeter of the cross-section was visualized (Fig. 6D and fig. S39). This revealed higher *MaSp2* expression in zone A cross-sections with smaller perimeters (proximal tail parts), while *MaSp1* and *SpiCE-LMa3* genes exhibited higher expression in zone A cross-sections with larger perimeters (distal tail, toward the sac). Notably, a significant negative correlation was observed between hematoxylin

values and the cross-section perimeter in zone A, further supporting the presence of different cell types along this zone (Fig. 6E). Next, based on the perimeter values, the zone A cross-sections were split into three parts: proximal, middle, and distal. The distribution of cell types in the spots corresponding to different parts of zone A was obtained by combining the scRNAseq, the spatial and the QuPath data and visualized in Fig. 6F. ZoneA_MaSp2 cells showed higher abundance in proximal zone A regions, which gradually decreased in middle and distal parts. Conversely, ZoneA_MaSp1 cells were more prevalent in distal parts compared to the most proximal region. ZoneA_SpiCE-LMa cells were found along the whole length of zone A, but most frequently in the middle and distal parts. ZoneB_MaSp3 cells were almost exclusively found in spots annotated as zone B, affirming their identification as zone B cells. Zone C spots were dominated by ZoneC_SpiCE-LMa cells, with a minor fraction of other cell types. ZoneABC cells were found in all the zones with low abundance. The distribution of the different cell types along the major ampullate gland is illustrated in Fig. 6G.

The location of the silk producing cell types determine the composition of layers in the silk fiber

In line with previous reports for other spider species (*26*), the secretions originating from zones A–C within *L. sclopetarius* major ampullate glands stained distinctly with H&E and were separated in the gland lumen (Fig. 7A). To confirm the zone-specific origin of these secretions, the vesicles in the cells of each zone and the secreted substances forming layers in the lumen were annotated and evaluated for staining intensity using QuPath image analysis. The obtained H&E intensity values for the vesicle content and the secreted substances within the lumen were plotted and color-coded according to their respective zones. The data obtained for vesicle content and secretions from the respective zones formed clusters, clearly indicating that the zone A secretion formed the bulk of the silk feedstock in the sac lumen, the secretion from zone B contributed a surrounding middle layer, while the secretion from zone C formed the outermost thin layer (Fig. 7B and fig. S40).

To verify that the three layers in the liquid silk feedstock persist in the major ampullate silk fiber, a solubilization protocol involving the use of different concentrations of urea (2M, 4M and 8M) was employed (*28*). Low concentrations of urea (2M) will not solubilize the entire silk and, therefore, enrich for proteins present at the surface, while the highest concentrations (8M) will dissolve the whole fiber. LC-MS/MS analysis of the solubilized fractions provided the relative fraction for each protein in the different urea concentrations. The SpiCE-LMa1 to SpiCE-LMa6, and AmSp-like 1 & 2 proteins were significantly enriched in 2M urea supernatants compared to the samples dissolved in 8M urea ($P < 0.05$), while the MaSp1a–c, MaSp2b-c, and MaSp2e-f were significantly enriched in the 8M compared to 2M urea supernatants ($P < 0.05$) (Fig. 7E and fig. S41). MaSp3a and b were most enriched in 4M urea supernatants. Next, we compared the protein composition in our proteomics data to the corresponding spatial gene expression in the gland. All the eight proteins enriched in the 8M urea samples overlapped with the marker genes of ZoneA_MaSp1 and ZoneA_MaSp2 cells. Interestingly, *MaSp3a* and *MaSp3b*, the most significant marker genes in ZoneB_MaSp3 cells, showed the highest relative fraction in the 4M urea samples (Fig. 7F). Among the eight proteins enriched in the 2M urea samples, five overlapped with marker genes of either ZoneB_MaSp3 or ZoneC_SpiCE-LMa cells (Fig. 7, E and F).

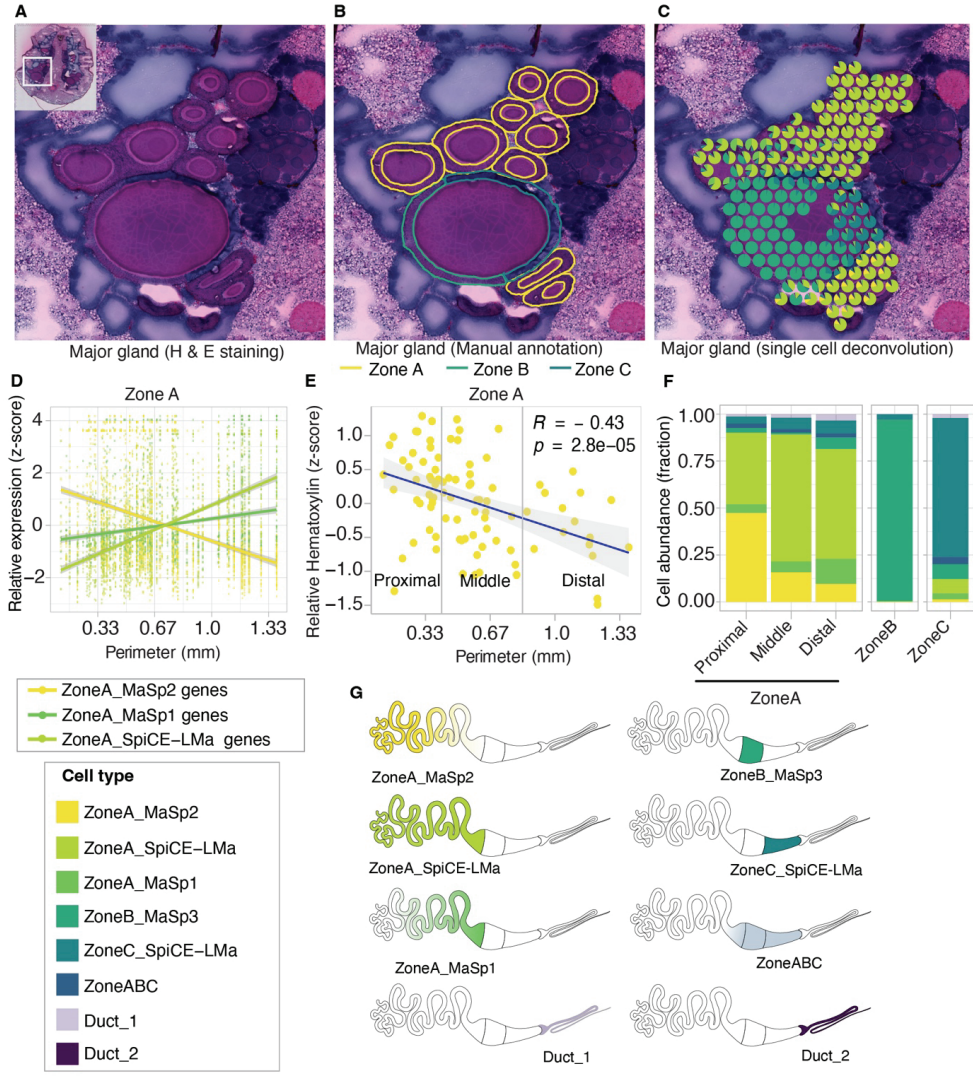

Figure 6: Spatial distribution of the eight major ampullate cell types. (**A**) H&E-stained section of the spider major ampullate gland, inset shows the area magnified. (**B**) The same section as in (A) with the zones (A in yellow, B in green and C in dark green) of the major ampullate gland indicated. The zones were identified based on the cell morphology and 11 cross-sections were obtained for this gland (9 for zone A, 1 for zone B and 1 for zone C). (**C**) Deconvolution of the spatial spots using the scRNAseq data generates a pie chart for each spot, where the colors represent cell types identified from the scRNAseq data and shows the fraction of cells that belongs to each cell type. (**D**) Relative gene expression of the silk protein genes that are marker genes for the cell types in zone A as a function of the perimeter of the gland cross-sections. The colors correspond to the three zone A cell types and the lines are the linear models of the gene expression profiles for the respective cell types. (**E**) Pearson correlation (R = -0.43, P = 2.8*10^{-5}, N = 80) of Hematoxylin mean values to the perimeter of the major ampullate gland cross sections in zone A. (**F**) Average fraction of cell types identified from scRNAseq data in zone A (proximal, middle, and distal), zone B and zone C spots as evaluated from the deconvolution plots. (**G**) Schematic figures showing the spatial distribution of the eight cell types in the major ampullate glands.

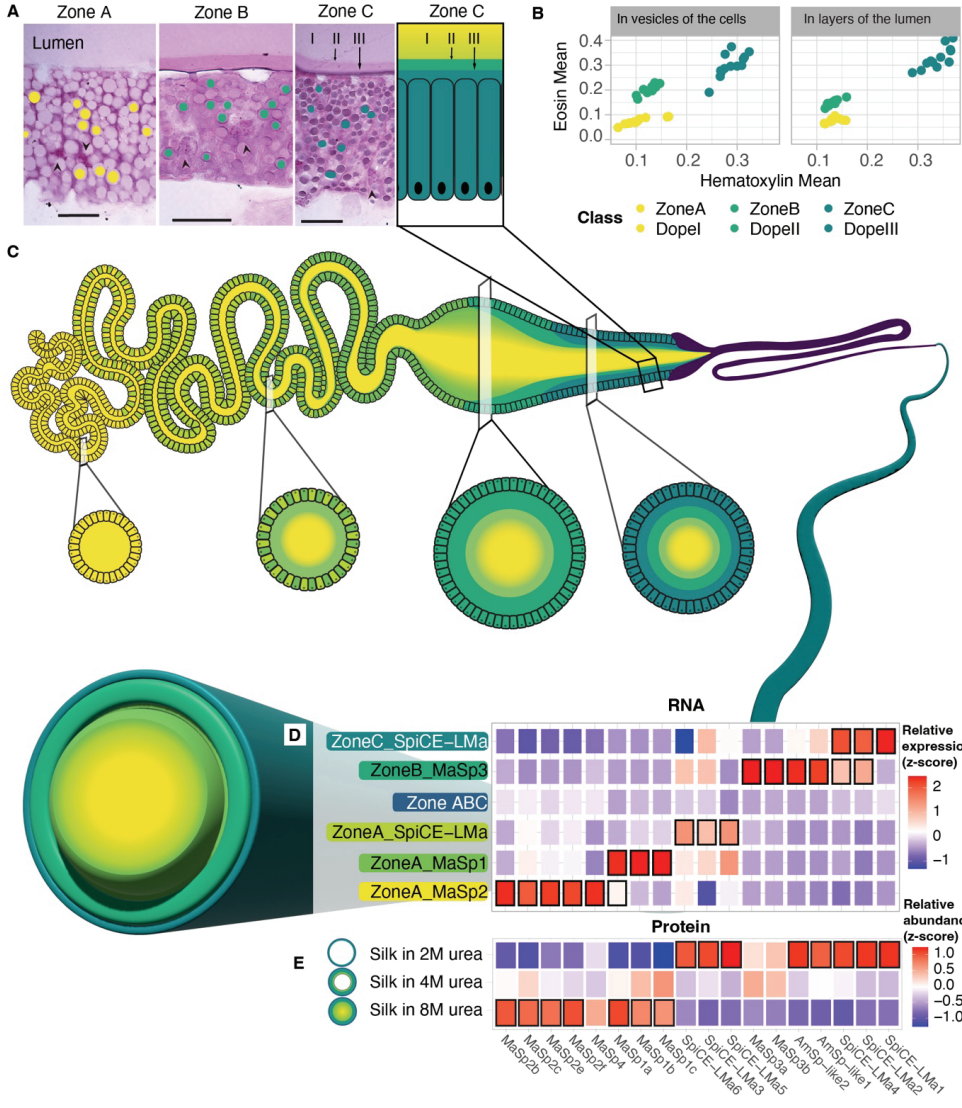

Figure 7: Origin and model of the multi-layered architecture of the major ampullate silk. (**A**) Histological sections of the *L. sclopetarius* major ampullate gland zone A, B and C, respectively (H&E staining). Secretions from zone A, B and C form three layers in the lumen (indicated as I, II and III in the zone C panels). Filled circles around the intracellular vesicles in the first three panels indicate the annotated regions for image analysis in QuPath. Scale bar, 20 µm. (**B**) H&E intensity plots of the annotation objects shown in (A)**.** (**C**) Schematic image of the major ampullate gland and model of the multi-layered architecture of the major ampullate silk. The localization of the six cell types and the layered secretions are illustrated in the gland schematic. Below the gland, schematic drawings of cross-sections from four locations along the gland are shown. In the model of the major ampullate silk, the colors correspond to the cell types the layers originate from, and the thickness of each layer is based on the proteomics analysis of the fiber (% total spectra). It should be noted that the histological sections suggest that the middle and outer layers could be thinner than indicated in the schematic figure of the fiber. **(D)** Heatmap showing gene expression values of the 18 silk proteins in the six cell types that are found in zone A, B and C. (**E**) Heatmap showing the peptides mapping to each of the 18 silk proteins in soluble extracts from major ampullate fibers incubated in 2, 4 and 8M urea, respectively. In (D) & (E), colors indicate expression values (z-scores), and the black frames indicate genes/proteins that were identified as significantly enriched ($P < 0.05$ when comparing samples between 2M & 8M urea). The 18 silk genes/proteins are listed below the panels.

To ensure that the results obtained from the sequential solubilization of the fibers reflected the proteins' localization in the fiber and not only their relative solubility in urea, major ampullate silk fibers were completely dissolved in formic acid and subsequently dried. Next, the dried samples were incubated in 2M, 4M and 8M urea, respectively, and the presence of different proteins in the supernatants were determined by LC-MS/MS. If the distribution of the proteins in the supernatants from the intact fibers and dissolved fibers would be similar, the results would support the absence of layers in the fiber. However, proteins that in our model reside in the outer layer of the fiber, i.e., AmSp-like1, AmSp-like2, SpiCE-LMa1, SpiCE-LMa2 and SpiCE-LMa4, show a higher relative abundance in the 2M urea supernatants from intact fibers compared to the dissolved fiber samples (fig. S42). In the 2M urea supernatants from the dissolved fiber samples, SpiCE-LMa5 and SpiCE-LMa6 show a significantly higher relative abundance compared to those obtained from the intact major ampullate fibers. The reason for the high abundance of these proteins in the 2M fraction, despite that they are primarily produced by zone A cells, could be attributed to their higher solubility compared to the MaSps (fig. S42). Another contributing factor could be that both these SpiCEs are expressed also in zone B and C, albeit to a lower degree (Figs. 4 and 5).

We further compared the relative abundance of the 18 proteins in the silk extracted at different urea concentrations for both the intact and the dissolved silk samples using pairwise Pearson correlation for all combinations of the samples (fig. S43). The results indicate that the 2M and 4M intact samples differed the most compared to the other samples (see dendrogram in fig. S43A). There was a higher variation between the 2M and 8M intact silk samples than the variation between the 2M and 8M dissolved samples (fig. S43A). Additionally, there was a low correlation between the 2M dissolved and the 2M intact silk samples, similar to that found between the 2M and 8M intact silk samples (fig. S43, B and D). This suggests that the obtained protein profiles in the intact samples are not a result of the differences in protein solubility but reflect the differences in protein composition of the different fiber layers. When comparing any of the intact 8M samples to any of the dissolved samples, regardless of urea concentration, or when comparing any of the dissolved samples to each other, they all showed a significant positive correlation ($P < 0.001$) (fig. S43, A, C, E, and F) which further contradicts the idea that the solubility of different proteins is the main explanation for the variation observed between the intact silk samples at different urea concentrations. Additionally, certain proteins expressed in zone B and C, such as SpiCE-LMa1, SpiCE-LMa2, SpiCE-LMa4, AmSp-like 1, and AmSp-like 2, are enriched in the 2M urea extracts from intact silk (fig. S43B). In summary, these data support the inhomogeneous distribution of proteins in the major ampullate silk fiber.

Discussion

Advanced molecular methodologies, such as single-cell RNA sequencing and spatial transcriptomics, serve as pivotal tools in advancing our comprehension of diverse tissue biology. However, their effectiveness depends on well-annotated genomes, posing a substantial challenge when investigating non-model organisms. This challenge has been particularly pronounced in investigations of spider silk production due to the nature of spidroin genes, characterized by their substantial size and repetitiveness. Addressing this limitation, we present a high-quality genome assembly of *L. sclopetarius*, featuring nearly complete annotations for all coding genes. Notably, our annotation reveals the presence of 35 full-length spidroins which is higher than previously reported (*7-11, 13*). Consistent with prior findings (*7-11, 13*), the spidroins exhibit variable lengths spanning 575 to 9146 amino acid residues (Fig. 1 and table S11). All the spidroin genes encoded proteins with signal peptides, the NT domain, a repetitive region, and had a stop codon in-frame. Manual inspection of the 3' downstream sequences revealed no sign of shifted reading frames, which means that the spidroin gene catalogue described herein encompasses exclusively complete genes.

Because the focus of this work was to provide a detailed understanding of the structure and function of the major ampullate silk gland, as well as the fiber it produces, we next sought to identify the proteins that make up the major ampullate fiber. This is not a trivial task, since the major ampullate fiber is easily contaminated by other silk types during spinning and collection of the fibers. To avoid these problems, we used LC-MS/MS, both on major ampullate silk fibers continuously collected from a spider mounted under a microscope and on isolated major ampullate glands. By only considering the secretory proteins that were found in both data sets, we could ensure that any contaminating proteins from other glands were removed. Assuming that the most abundant proteins are contributing most to the properties of the fiber, we filtered proteins with abundance of more than 0.1 % based on the percent of total spectra, which allowed us to identify 18 proteins as the predominant constituents in the major ampullate silk. Ten of the 18 came from the four distinct classes of the major ampullate spidroins, namely MaSp1 to MaSp4. Together, these spidroin proteins constitute 96% of the total protein content of the fiber (Fig. 7G), which is in line with the reported MaSp content in the major ampullate silks from other spider species (*9, 10, 13, 22, 47*), but the abundance of the different MaSp types differ between fibers from some species (*8, 9*).

In addition to the MaSps, a previously unreported type of spidroin, which we named AmSp-like, was also found in the fiber. The AmSps were characterized by having an NT domain that grouped with the minor and major ampullate spidroins and a repetitive region, but intriguingly, these spidroins lacked the CT domain and displayed an amino acid composition more similar to the SpiCE proteins than the MaSps (Fig. 2E and fig. S17). Six proteins, designated as SpiCE-LMa1 to SpiCE-LMa6, were identified as constituents of the fiber. Despite their common naming 'SpiCE-LMa', their properties are diverse in terms of molecular weight (table S13), tertiary structure predicted by AlphaFold2 (fig. S26) and amino acid composition (Fig. 2E and fig. S17). The amino acid composition of SpiCE-LMa1 and SpiCE-LMa2 that are expressed in zone C was similar to that of the MaSps, yet they lack both NT and CT domains. While the other SpiCE-LMa proteins either lacked homologs or exhibited homology to hypothetical proteins from other spiders, SpiCE-LMa3 and SpiCE-LMa5 demonstrate similarity to previously reported SpiCE proteins (tables S16 to 20). Specifically, SpiCE-LMa3 shares similarity with SpiCE-CMa2 (produced by spiders from the Caerostris family), while SpiCE-LMa5 shares similarity with SpiCE-NMa4 (Nephilinae family), the latter of which has been designated as a CRP (*9*). Notably, SpiCE-LMa3 to SpiCE-LMa6 proteins, that were found to be expressed and secreted from zones A and B have a high cysteine content. To date, only one study has experimentally investigated the impact of SpiCE on the mechanical properties of materials made from recombinant mini-spidroins. A composite film composed of a mini-spidroin and SpiCE-NMa1 had higher tensile strength compared to films comprising the mini-spidroin alone. However, corresponding composite silk fibers display reduced tensile strength (*9*), highlighting the need for further investigation into the functional role of SpiCE proteins in the silk fiber.

To determine which cell types that express the 18 silk proteins and their spatial distribution in the gland, we combined three unbiased transcriptomics techniques. First, we used bulk RNA sequencing to reveal that all 18 silk genes are expressed in the tail and the sac and not in the duct (Fig. 2D). Thirteen of these were among the top 100 significantly differentially expressed genes in either the tail or the sac of the gland indicating that the expression profile indeed differs along the gland. Second, single-cell RNA sequencing analysis of whole major ampullate glands identified eight cell types. Notably, the marker genes of five of these eight cell types overlapped with all 18 silk genes (Table 2 and Fig. 5B). By cross-referencing the marker genes of the scRNA cell types with the differentially expressed genes found in the bulk RNA data, we were able to assign three cell types to the tail, three to the sac, and two to the duct. The transcriptional profile of cells expressing spider silk proteins could be matched to the bulk-RNA sequencing data from the tail and sac samples, but not to the samples derived from the duct. This means that the 18 silk proteins are produced by the five cell types located in the tail and sac. In

order to spatially resolve the distribution of the cell types, we used a third transcriptomic technique, 10X Visium. In the sections used for spatial transcriptomics, we first manually annotated the transcriptomic spots within the major ampullate gland zone A, B, and C using the distinct H&E staining pattern and morphology of the epithelium in the respective zones (Fig. 4B). Comparing the expression of genes between the zones revealed that 15 out of the 18 silk genes are differentially expressed in at least one of these three zones (Fig. 4C). Next, by integrating the spatial transcriptomics with the single-cell data, the precise localization of the cell types was revealed. In line with the bulk RNA data, the three cell types assigned to the sac were predominately present in the zone B and zone C. Notably, by using the spatial transcriptomics data, we could see a clear distinction between ZoneB_MaSp3 cells that were confined to zone B and ZoneC_SpiCE-LMa cells that were dominant in the zone C epithelium. The three cell types that were assigned to the tail were indeed found in zone A. The tail is long and winding, and by taking advantage of the observation that the cross-section of the tail increases along the gland, we could generate information about the spatial location of cell types even within zone A. By determining the perimeter of each cross sectioned part of the tail, we could order them from proximal (small) to distal (large) (Fig. 6). We found that the ZoneA_MaSp2 cells were primarily localized to the proximal part of zone A, while ZoneA_MaSp1 cells were most abundant in the distal portion of zone A (closer to zone B cells). ZoneA_SpiCE-LMa cells were present in all parts of the zone A (Fig. 6). Finally, again in line with the bulk RNA data, the two cell types assigned to the duct were not detected in the tail or sac. Taken together, our data support that the proteins that make up the major ampullate silk are produced by five cell types that have specific regional anatomical localizations in zone A, B and C, but not by the cell types confined to the duct.

Next, we sought to connect the expression of the genes in the cell types along the gland to the multiple layers observed in the silk feedstock (Fig. 7A). By using digital image analysis, we showed that the H&E staining of the layers in the dope (the liquid feedstock stored in the sac) as determined by QuPath corresponds to the staining of the intracellular vesicles in the corresponding epithelium. The inner layer of the dope matched the staining of the vesicles in the zone A, the middle layer stained as the vesicles in the zone B and the outer layer matched the staining of the vesicles in the zone C epithelial cells (Fig. 7, A and B). This indicates that the three zones indeed produce layered secretions with different protein compositions. Moreover, since we knew the presence of different cell types in zone A, B and C, respectively, and their expression profiles (Fig. 7D), we could predict the protein composition of the different layers in the fiber. Given this model (Fig. 7C), the inner layer of the silk fiber contains the MaSp1, MaSp2, MaSp4, SpiCE-LMa3, SpiCE-LMa5, and SpiCE-LMa6 proteins, the middle layer primarily contains the MaSp3, AmSp-like1 and AmSp-like2, SpiCE-LMa2, and SpiCE-LMa4 proteins, while the outer layer is dominated by SpiCE-LMa1, SpiCE-LMa2, and SpiCE-LMa4 but contains no classical spidroins. In order to test the hypothesis that the layers identified in the gland lumen persist to form layers in the fiber, we ran proteomic analysis on silk fiber extracts after exposure of major ampullate silk fibers to different concentrations of urea (2M, 4M, and 8M, respectively). This approach was established by Vollrath *et al.* to sequentially dissolve major ampullate silk fibers (*28*). Supernatants from fibers incubated in 2M urea contained proteins primarily expressed in zone B and zone C, whereas in 8M urea, that completely dissolve the silk fibers, proteins that are expressed in zone A were enriched (Fig. 7E). When using the 4M urea, MaSp3a and MaSp3b, which are marker genes for zone B, had the highest relative abundance. These results were further verified to be related to the proteins' localization in the fiber (figs. S42, S43). The data presented herein allows us to present a detailed model of the composition of the three-layered major ampullate silk fiber, and to conclude that each layer has a distinct protein composition that is derived from specific cell types confined to zone A, B and C, respectively (Fig. 7). Since the ZoneA_MaSp2 cells are the dominating cell type in the most proximal part of the tail, it is logical that MaSp2 proteins form the core of the fiber and that the more peripheral regions of the core are dominated by MaSp1 proteins, secreted by the ZoneA_MaSp1 cells which are located more distally in Zone A. This finding contrasts to a report from Hu *et al*. (*13*), which shows that the central

core of the *Trichonephila* major ampullate fiber is dominated by MaSp1 proteins and MaSp2 are found more peripherally, but is in line with work by Sponner *et al*. (*32*), who by biochemical and immunohistochemical investigations of the *Trichonephila* major ampullate silk revealed the presence of both MaSp1 and MaSp2 in the inner core but exclusively MaSp1 in the peripheral parts of the core. The latter study also concludes that the layer surrounding the core of the fiber (referred to as skin layer) is more tough and resistant to chemical treatment than the outermost layer (coat) and the central core. If these notions are combined with the data presented herein, a plausible conclusion is that the skin layer is dominated by MaSp3 and corresponds to the middle layer in our model.

Thus, what is the purpose of the layered fiber structure? Despite decades of research, the function of each layer is not completely understood. The outermost layer and the skin layer have been suggested to protect against environmental impact (*32*), the skin layer is resistant to protease degradation, and the core of the fiber accounts for most of the mechanical properties of the fiber (*48*). However, Arakawa *et al*. (*12*) have recently published over 1000 transcriptomes from different spider species along with corresponding mechanical data for the major ampullate fibers and showed that expression of *MaSp3* correlates with higher fiber toughness. Provided that the *MaSp3* is expressed mainly in zone B also in other species, this would suggest that the skin layer could be important for obtaining a high fiber toughness. On the other hand, the presence of layered structures and three epithelial zones in the major ampullate glands from species that do not express *MaSp3* have been reported (*12, 26*), which suggests that not only the specific protein composition of the fiber layers but also the layered structure in itself may be important for the fiber's properties. For example, spider species that do not express MaSp3 (*12*) (*e.g.* Pisauridae and Agelinidae) also have three epithelial zones in the major ampullate gland (*26*), and the Pisauridae spider *Euprosthenops australis* spins a fiber with one of the highest tensile strengths reported (*49*).

Recent publications have pointed toward the idea that formation of recombinant spidroin heterodimers, interconnected in the CT domain, may result in improved mechanical properties of the silk fiber (*19*). Our data suggest that heterodimer formation of spidroins via their CT domains is limited, since the dimerization would occur in the endoplasmic reticulum (ER) of individual cells (*36, 37*). As can be seen in Fig. 5, there is some overlap of the expression of the MaSps in some cells, but *MaSp1*, *MaSp2* and *MaSp3* are generally expressed in separate cell types (Fig. 5B). This implies that the spidroins in the soluble state would primarily form dimers composed of one type of spidroin. In the lumen of the sac, just before the duct begins, there is clear evidence that the secretions from the three zones are separated (Fig. 7A) (*26, 50*). This means that the MaSp1 and MaSp2 spidroins would be found in the zone A "core" secretion, which is surrounded by the zone B secretion that contains MaSp3 (Fig. 4C and 7E). While some diffusion would occur between the layers, our data indicates that the protein composition of the layers mostly stay intact during fiber processing. When the dope travels down the duct, the pH gradient will cause the NT domains to dimerize (*36, 38, 39, 51*), which leads to that the spidroins are linked together in large complexes (*40*). Since the protonatable residues of the *L. sclopetarius* MaSp NT domain are evolutionary conserved (fig. S4) (*38*), inter-class NT dimerization between spidroins, even originating from different layers, could occur. Experimental support for this is currently lacking, but the fiber layers should, in any case, probably not be considered as completely isolated.

Conclusively, we present a model of the major ampullate spider silk which contains three layers with specific protein compositions. Diffusion may take place across the interface of the layers, possibly making the transition between these gradual, but even if so, the distribution of proteins in the fiber is clearly inhomogeneous. The model is based on several observations: First, the results from the transcriptomics analyses (Figs. 3 and 4) clearly show that the MaSps, AmSp and SpiCE-LMa proteins are primarily produced in three specific zones along the gland. Second, the H&E staining of the intracellular granules in the epithelium of said zones stain differentially, and the respective staining

matches the staining of the three layers observed in the gland lumen, i.e., in the dope. The layered secretions from zone A, B and C have been observed before and in several distantly related spider species (*26, 50*), which suggest that this is a common feature of the major ampullate gland and its silk. Third, the proteomics data presented herein supports the inhomogeneous distribution of proteins in the fiber and reflects the compositions that would be expected based on a layered structure originating from secretions from the three epithelial zones. Fourth, numerous studies using fiber diffraction, electron microscopy and light microscopy support the conclusion that the major ampullate silk fiber is layered (*27-33*).

This work provides a detailed understanding of the major ampullate gland's biology and the intricate structure and composition of the major ampullate silk fiber. Specifically, eight distinct glandular cell types were characterized, five of which contribute proteins to the silk feedstock. A high-resolution spatial mapping of these cell types within the major ampullate gland of *L. sclopetarius* was presented along with the identification of several previously uncharacterized genes exhibiting significant differential expression across the cell types. Additionally, 18 spider silk proteins were found to make up the bulk of the fiber. Finally, the protein compositions of the three enigmatic layered secretions in the gland were elucidated and correlated to the protein composition of sequentially dissolved fibers. These insights are important for improved bioprocessing of artificial spider silk fibers and should be incorporated in molecular dynamic simulations (*52*) and generative machine learning models (*53*) which hold promise to reveal molecular features that govern the mechanical performance of the fiber.

Methods

Spider Samples

L. sclopetarius adult female spiders were collected in the wild in a small habitat in Uppsala, Sweden. Taxonomic identity of the spider was verified by the Museum of Natural History, Stockholm, Sweden. The spiders were kept in big containers that allowed them to spin webs. They were fed with meal worms or *Drosophila* flies weekly and watered daily.

Extraction and sequencing of genomic DNA

High molecular weight (HMW) genomic DNA was extracted from the whole body of one individual *L. sclopetarius* spider using MagAttract HMW DNA Kit (Qiagen). One adult female spider was anesthetized using dry ice and dissected on ice. The exoskeleton was removed, and all soft tissue was collected to extract HMW DNA. The DNA extraction was done according to the manufacturer's protocol, except for tissue incubation in RNAse and proteinase-K at 50°C for 30 min and elution of DNA that was done twice by adding an extra 100uL of buffer AE to the beads. The purified DNA was run on a 0.5% agarose gel to assess DNA integrity. Absorbance ratios were evaluated on a Nanodrop spectrophotometer and determined to be as follows: 260/280: 1.82 and 260/230: 1.91, resulting in a total of 36.2 mg DNA. The extracted genomic DNA was also subjected to a quality check using a BioAnalyzer (Agilent), which revealed a single peak at around 11 kb. 10.3 µg DNA was used to make a 20kb library. The National Genomics Infrastructure (NGI) platform at SciLifeLab, Uppsala University, performed the library preparation and sequencing using PacBio long-reads and 10X Genomics linked-reads.

PacBio genomic library preparation and sequencing

The QC-passed DNA samples were sent to the NGI platform at SciLifeLab, Uppsala University, for library preparation and sequencing. The SMRTbell libraries obtained by the TPK1 kit, according to the manufacturer's instructions, were size-selected at 20 kb using a BluePippin instrument (SAGE) and

sequenced on 60 SMRT cells of the RSII instrument using P5-C3 chemistry. For each SMRT cell, 10 h movies were captured. A total of 798 Gbp of data with an insert size of 11 kb was produced.

Library preparation and sequencing using 10X Chromium linked reads
The HMW DNA was used to generate the 10X linked read libraries on 10X genomics Chromium platform (Genome Library Kit & Gel Bead Kit v2 PN-1000017, genome Chip Kit v2 PN120257) following the manufacturer's guidelines. The 10X libraries were sequenced on Illumina NovaSeq6000 instrument (NovaSeq Control Software 1.6.0/RTA version 3.4.4) with 151bp paired-end setup using NovaSeqXp workflow in S4 flow cell. The Bcl to FastQ conversion was performed using bcl2fastq_v2.19.1.403 from the CASAVA software suite. Sanger / phred33 / Illumina 1.8+ was used as the quality scale.

Bulk RNA extraction and sequencing of silk glands, head, and abdomen
The spiders were anesthetized using dry ice before they were dissected on ice. After making an incision at the pedicel, the abdomen was gently pinned to a wax plate placed under a Zeiss Stemi 305 stereo microscope, and the exoskeleton was carefully removed with micro scissors to visualize the silk glands. Phosphate-buffered saline (PBS, pH 7.4) was used to wash off the excess non-silk tissue. With the help of micro tweezers, silk glands (major ampullate glands, minor ampullate glands, flagelliform glands, aggregate glands, tubuliform glands) were isolated separately by holding their ducts. The aciniform and piriform glands from these five spiders were extracted as a single sample due to difficulties in separating them owing to their small sizes. The major ampullate glands from six additional individuals were cut into three parts: tail, sac and duct and used for RNA extraction. In another preparation, after removing the exoskeleton the soft tissue from the whole abdomen was scraped out and used for extracting RNA. The RNA from the head was extracted similarly after removing the legs and the thick exoskeleton. Five replicates were collected for each sample type and RNA was extracted from each replicate separately.

A total of 57 RNA samples were extracted using RNeasy Plus Mini kit (QiaGen) by following manufacturer's protocol. The integrity of the samples was estimated using Tapestation (Agilent Technologies). The transcriptome libraries from the different tissues were generated using Illumina TruSeq Stranded mRNA kit following manufacturer's protocol and 151 bp paired end reads were sequenced using Novaseq6000 instrument.

PacBio long-read Iso-Seq library construction and sequencing for major ampullate glands
RNA was extracted from the major ampullate gland from one individual and homogenized in TriZol. The extracted RNA was sequenced at NGI, Uppsala University, Sweden. RNA QC was performed on the Agilent Bioanalyzer instrument, using the Eukaryote Total RNA Nano kit. The sequencing library was prepared according to PacBio's Procedure & Checklist – Iso-Seq™ Express Template Preparation for Sequel® and Sequel II Systems, PN 101-763-800 version 02 (October 2019) using the NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module, the Iso-Seq Express Oligo Kit, ProNex beads and the SMRTbell Express Template Prep Kit 2.0. The sample (300 ng) was first amplified to 12 cycles, followed by 3 additional cycles, according to the protocol. In the purification of amplified cDNA, the Long Transcripts workflow was applied to obtain material enriched for longer transcripts (>3kb). The quality control of the SMRTbell libraries was performed with the Qubit dsDNA HS kit and the Agilent Bioanalyzer High Sensitivity kit. Primer annealing and polymerase binding was performed using the Sequel II binding kit 2.0. The samples were sequenced on the Sequel II instrument, using the Sequel II sequencing plate 2.0 and the Sequel® II SMRT® Cell 8M, with 24 h of movie time and 2 h of pre-extension time.

Single-cell preparation from major ampullate glands of *L. sclopetarius*
In total 20 spiders were used for single-cell sequencing. The first ten spiders were anesthetized in dry ice and dissected. All buffers were bubbled with carbogen during and prior to use. The major ampullate

glands were taken out in ringer solution, pH 7.4. The duct was removed. The glands were washed with PBS, pH 7.4 and then incubated into pre-warmed trypsin-EDTA (Gibco, 0.5%) at 37°C for 1 min in a low-binding micro-centrifuge tube. The glands were triturated with a pipette briefly and centrifuged at 300 *g* for 3 min at 4°C. The pellets were resuspended in 300 µL Dulbecco's Modified Eagle's Medium (DMEM; Gibco) containing 1% BSA (Sigma). The DMEM and the tubes were briefly bubbled with carbogen prior to resuspension. The suspension was strained through a 40 µm cell strainer and transferred to low-binding micro-centrifuge tubes. Cells were counted using Trypan blue dye.

Because of problems with contaminating droplets of dope that made the separation of single cells challenging, a slightly modified protocol was used for the following ten spiders. These were processed as described above, but with the exception that the duct was not removed, and the sac was cut and kept in PBS, pH 7.4 for 10 min to allow the dope to flow out. The pieces of glands were picked up and incubated in pre warmed trypsin-EDTA (Gibco, 0.5%) at 37°C for 2 min in a low-binding micro-centrifuge tube. The suspension was triturated with fire-polished glass Pasteur pipette for 1 min, incubated in trypsin-EDTA at 37°C for 2 min, and triturated again with fire polished glass Pasteur pipette for 3 min. The suspension was centrifuged at 300 *g* for 3 min at 4°C and the pellet was resuspended in 200 µL DMEM containing 3% BSA. The suspension was strained through 40 µM pre-washed cell strainers into low-binding micro-centrifuge tubes. The strainer was further washed with 100 µL DMEM containing 3% BSA to reduce the loss of cells. The library preparation was done using 10X 3' GE kit on the 10X Chromium Single Cell 3' Platform following manufacturer's protocol (Dual Index 10X_3'_V3.1) and sequenced using Novaseq (100 cycles).

Sample preparation for spatial transcriptomics

Whole opisthosomas of spiders were flash-frozen in optimal cutting temperature (OCT) medium on an isopentane-dry ice bath and samples were stored at –80°C until use. The samples were sectioned in a cryotome (10 µm) with knife temperature set at –23°C and sample holder temperature at –10°C. Eight sections from five individual spiders were carefully mounted on the capture areas (6.5 x 6.5 mm) of the Visium spatial slide (10X genomics) and permeabilized for 30 minutes following the 10X Visium spatial tissue optimization protocol. The libraries were constructed according to Visium spatial gene expression protocol (10X genomics). For cDNA amplification 13-16 PCR cycles were performed and for the indexing 13 PCR cycles were used. The sequencing was performed using a SP-200 flow cell on the Illumina Nova-Seq 6000. All the sections were stained with H&E for histological evaluation.

Genome assembly and polishing

The Falcon and Falcon-Unzip (pb-falcon version 0.2.7) (*54*) *de novo* assemblers were used to assemble the PacBio data. The initial polishing was done using the Falcon-Unzip polishing module. The Chromium 10X linked read data was used to further polish the long-read CLR assembly. Reads were aligned to the PacBio assembly using Long Ranger (version 2.1.4) and three rounds of polishing were done using Pilon (*55*) (version 1.22) using the diploid flag.

Genome size and heterozygosity estimation

The raw Illumina reads from 10X genomic linked sequencing libraries were trimmed using Trimmomatic (*56*), the canonical 20-mer counts were collected using Jellyfish (*57*). With the 20-mer histogram, GenomeScope2 (*58*) was used to estimate the approximate genome size and heterozygosity.

Mitochondrial genome assembly

The mitochondrial genome was identified by mapping the assembly to an existing reference spider species, *Neoscona adianta* (Genbank accession: NC_029756.1) using BLAST. The identified regions from the assembly were extracted using BEDTools (*59*). The extracted mitochondrial contigs were then annotated using MITOS web server (*60*).

PacBio Iso-Seq transcriptome assembly

To generate full-length consensus transcript isoforms from the major ampullate gland, the raw polymerase reads were processed using SMRTlink. The subread BAM file was processed to generate the circular consensus sequence (CSS) reads. These reads were further classified into full length (FL) transcript sequences based on the criteria that they contain 5' primer, 3' primer and polyA tails. The FL transcript sequences were processed using IsoSeq3 platform for generating full length non-chimeric reads (FLNC) which were further clustered using ICE algorithm to produce both high- and low-quality polished full length consensus sequences. The high-quality (HQ) sequences were used for the subsequent analysis. The high-quality transcripts were mapped to the *de novo* assembled *L. sclopetarius* genome using minimap2 (*61*) (version 2.2.4) with parameters *-ax splice -uf --secondary=no -C5*. The alignment in the SAM format were processed into non-redundant full-length transcripts using the "collapse_isoforms_by_sam.py" script from the cDNA-Cupcake tool (https://github.com/Magdoll/cDNA_Cupcake). To assess the completeness of the genome assembly and the annotation, BUSCO (*62*) (version 4.1.4) analysis was performed using the Arachnida_odb10 lineage dataset.

Genome annotation

The annotation of the *de novo* assembled *L. sclopetarius* genome was performed using MAKER (*63*) version 3.01.02. High-confidence protein sequences (561356 proteins) were collected from the Uniprot Swiss-prot database (downloaded in November 2019) and a specific set of spidroin sequences (1051) were downloaded from NCBI (November 2019).

A repeat library was created using the RepeatModeler package (version 1.0.11, https://www.repeatmasker.org/RepeatModeler/). Since the spidroin sequences are highly repetitive in nature, the repeats modelled by the RepeatModeler were vetted against our specific set of spidroin data set. The repeat sequences in the assembled genome were identified using RepeatMasker (version 4.0.9, https://www.repeatmasker.org/) and repeatRunner (https://www.yandell-lab.org/software/repeatrunner.html). The tRNAs were identified using tRNAscan (*64*) version 1.3.1 while the conserved non-coding RNAs were identified using the Infernal package (*65*) and the RNA family database, Rfam version 11 (*66*).

The MAKER package was executed in two runs: a) First MAKER was used to create a profile using the Uniprot Swiss-prot protein sequences, the specific set of spidroin sequences and RNA-seq data from different tissues. An in-house pipeline was used to select a set of genes from this initial evidence-based annotation (first run) and to train Augustus (*67*) and SNAP (*68*), b) MAKER was run a second time using the evidence from the first run and the prediction from Augustus. For the construction of gene models, the prediction from Augustus was used.

Functional annotation of genes and transcripts was performed using the translated CDS features for each of the coding transcripts. The protein sequences were searched against UniProt, Swiss-Prot databases and the specific set of spidroin sequences was scanned using BLAST to retrieve gene names and protein functions. InterProScan (5.30-69.0) (*69*) was used for extracting additional annotations (functional domains and sites) from various other biological databases (20 in total).

Improving 3′ UTR annotation using single-cell data

The aligned BAM files from cell ranger were used for filtering BAM files using UMI-tools FilterBam (*70*) to include reads that were produced with corrected molecular barcode tag by cell ranger counts. The filtered BAM file was processed to remove PCR duplicates using UMI-tools dedup (*70*). The tool findPeaks from Homer was used to identify peaks (-size 50 -fragLength 100 -minDist 1). The peaks file was converted into bed file using BEDTools (*59*). These peaks were then either annotated as 3′ UTR to genes that lacked this feature or reannotated as extended 3′ UTR feature to the nearest genes that were identified within 5000 bps.

Identification of spidroins and manual curation

For identifying the spidroins in *L. sclopetarius* genome, two different approaches were used: a) the translated protein sequences of *L. sclopetarius* were scanned against the PFAM HMM profiles of NT domain, CT domain and Tubuliform egg casing silk strands structural domains. The HMM profile for these domains, Spidroin_N.hmm (NT domain, PF16763), Spidroin_MaSp.hmm (CT domain, PF11260) and RP1-2.hmm (Tubuliform egg casing silk strands domain, PF12042) were downloaded from PFAM database. To minimize the risk of false positive results, the hits with an e-value cutoff below 1e-05 were filtered out; and b) the reference spidroin sequences were downloaded from NCBI database (in November 2019). The redundant sequences were removed using CD-HIT at sequence identity cutoff of 95% and a custom database with full length spidroins (as retrieved from database), NT and CT domain sequences were created using BLAST package. The spidroin sequences of *L. sclopetarius* were identified by homology search using BLASTp with an e-value cutoff of 1e-05 against the full-length reference database. The identified sequences were confirmed for NT and CT domains.

Because of the huge size and high repetitiveness of spidroins, identification of exon boundaries by assembling tools can result in inaccuracies. Thus, the identified *L. sclopetarius* spidroin sequence loci and their surrounding regions (extending 5000–10000 bps on either end of the gene) were manually inspected if those were defined by the automated MAKER gene model. Web Apollo (*71*) genome browser was used for viewing gene models by entering the transcript identifier and identifying supporting data from bulk RNA sequencing and PacBio Iso-seq experiments. The extended gene sequences were searched separately against the custom NT and CT domain reference database using BLASTx with an e-value cutoff of 1e-05. An additional 100 bp region upstream of the identified NT domain region was scanned for signal peptide using SignalP version 6 (*72*). Thus, we assume that the boundaries of spidroin genes were properly defined. The sequences for which the CT domains could not be defined were kept unchanged as per the automated MAKER gene model. We also looked for multiple spidroin genes that were collapsed into single locus due to their high sequence similarity and, therefore, were hard to assemble as separate loci. After defining gene boundaries for every identified spidroin, the gene sequences were translated in all six translational frames and manually inspected for repetitive motifs to identify if any mis-annotations (missing exons due to incorrect reading frame) existed in the current gene model using Unipro UGENE software (*73*). Based on the identified mis-annotations, either previously unidentified genes were added, or the existing gene models were replaced with a corrected model. All the corrected sequences were later confirmed by mapping against the reference genome using exonerate. GeneWise (*74*) and Scipio (*75*) were used to generate GFF file for the corrected gene models.

Functional assignment to proteins with hypothetical function

To predict function for proteins assigned as “hypothetical protein” (from Genome annotation), Orthofinder (*76*) (version 2.5.2) was used with default settings to identify gene family clusters between *L. sclopetarius, Trichonephila clavipes* (NCBI accession number PRJDB10126)*, Trichonephila clavata* (PRJDB10007), *Nephila pilipes* (PRJDB10128), *Trichonephila inaurata madagascariensis* (PRJDB10127), *Argiope bruennichi* (PRJNA629526), and *Araneus ventricosus* (PRJDB7092). The protein sequences were downloaded from NCBI. The putative transcript isoforms were removed from the *L. sclopetarius* proteome dataset and the longest canonical sequence were kept for the analysis. Proteins from each of the species were processed using Orthofinder-Diamond to assign proteins into orthogroups. In-house python scripts were used to process equivalent genes that were grouped as an orthogroup to further assign functions to hypothetical proteins based on proteins with known functions within the same orthogroup. The annotation was performed at two levels; a) analyzing homologous cluster assigned to an orthogroup within *L. sclopetarius* and b) analyzing gene clusters from other spider species (*T. clavipes, T. clavata*, *N. pilipes*, *T. inaurata madagascariensis*, *A. bruennichi*, *A. ventricosus*) that were assigned to an orthogroup.

Analysis of bulk RNA-sequencing data

The raw sequence reads were mapped to the *de novo* assembled genome using STAR (*77*) (version 2.7). The read counts were generated using featureCounts (*78*) from Rsubread package (version 2.0.0) (*79*). Transcripts per million (TPM) were used for visualization and comparison of gene expression levels within samples. For PCA and PLS-DA analysis DEseq2 variance stabilization transformation (VST), DESeq2 package (version 1.38.3) (*80*), was used. For differential expression (DE) analysis between samples DEseq2 default normalization was used. Genes with TPM >4 in at least one sample, a total of 15,086 genes, were kept for subsequent analysis.

A four-step filtering approach pinpointed variable major ampullate gland-specific genes (fig. S28). Initially, PCA on head and body samples distinguished the tissue types using the 1st PC, which explained 73.5% of the variation. Genes with a positive 1st PC loading value were retained, totaling 7,777 genes. Subsequently, PCA on major ampullate gland and body samples further differentiated the tissue types, with the 1st PC accounting for 76.8% of the variation. Genes with a positive 1st PC loading value, along with *MaSp4* and *SpiCE-LMa4*, resulted in a set of 2,776 genes. In the third step, PCA was conducted on tail, sac, and duct samples from the major ampullate gland. The first two principal components explained 77.0% of the sample variation. Genes with a Euclidean distance exceeding 0.13 from the origin were classified as the major ampullate gland-specific gene set and annotated based on their PCA coordinates: tail ($PC1 < 0$, $PC2 \geq 0$), sac ($PC1 \geq 0$, $PC2 \geq 0$), and duct ($PC2 < 0$).

Lastly, PLS-DA was performed on the same samples (tail, sac, and duct) and the major ampullate gland gene set using ropls (version 1.34.0) (*81*). This analysis effectively distinguished tail, sac, and duct samples using two components, based on the Q-value and visualised in Fig. 3. The PLS-DA loading of genes in the two principal components facilitated the classification of major-specific genes into tail ($PC1 < 0$), sac ($PC1 \geq 0$, $PC2 \leq 0$), and duct ($PC1 \geq 0$, $PC2 > 0$). DE analysis using DESeq2 with stringent criteria (adjusted $P < 0.001$, fold change > 4) identified DE genes in each major ampullate gland region (maximum 100 per part). DE analysis compared one part of the major ampullate gland, e.g., tail, against the other two, i.e., sac and duct. VST, PCA loadings, PLS-DA loadings, and DESeq2 fold change for all comparisons of the major ampullate gene set can be found in table S14.

Analysis of single-cell RNA sequencing data

Reads were mapped to transcripts using CellRanger (version 3.0.1, https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger). Initial QC analysis removed all cells with fewer than 300 expressed genes and/or less than 500 total transcripts. Only samples with more than 500 cells were retained for further analysis. All steps to separate the different glands and identify corresponding marker genes were performed in Seurat (version 4.0.3, https://satijalab.org/seurat/) (*82, 83*). Cells were normalized using SCTransform and integrated with Canonical correlation analysis distances between samples.

After initial QC steps, 18539 cells were obtained from 7 samples that clustered into 23 clusters. Among these clusters, a subset was identified where the marker genes overlapped with the major ampullate tail, sac, and duct gene sets. By reevaluating this specific subset and restricting the genes analyzed to the intersection of the top 2000 most variable genes from the scRNAseq data and the major ampullate gene set derived from the bulk RNA analysis, 9700 cells were obtained from 7 samples which clustered into nine groups. The smallest cluster only contained cells from one sample and was removed from further analysis. Marker genes for each cluster were identified using FindAllMarkers from Seurat (version 5.0.3). SCTtransformed values from gene counts, based on mapping using Cell Ranger (version 7.0.1), were used.

Analysis of spatial transcriptomic data and deconvolution

Eight slides were manually annotated as silk glands based on the morphology using the annotation tool in the Loupe browser (version 6.4.1, https://support.10xgenomics.com/spatial-gene-expression/software/visualization/latest/what-is-loupe-browser). Reads were mapped with Space

Ranger (version 1.2.0, https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger) to the reference genome and annotation. Initial QC removed samples with less than 300 genes and 500 transcripts. All steps were carried out in Seurat (version 4.0.3) (*82, 83*). Cells were normalized using SCTransform and integrated with canonical correlation analysis distances between samples. The distance between the samples was visualized using UMAP. Marker genes for different classes were identified using a subset of 400 cells per class.

For analysis of the major ampullate gland, five samples with good annotation of the gland were kept. The image files were imported in QuPath (version 0.4.3) (*46*) and the regions identified as major ampullate gland were further separated into different zones A, B and C based on H&E-staining and cell morphology using brush tool. Average Eosin, average Hematoxylin and perimeter values were determined for each region using QuPath default parameters. Marker genes for the different zones were identified using gene counts from Space Ranger (version 2.0.1) and findAllMarkers for SCTransformed data from Seurat (version 5.0.3).

Pairwise Pearson correlation was performed for eosin, hematoxylin, and perimeter values for the zone A regions. Zone A regions were split into three classes based on their perimeter value. Regions with perimeter less than 500 pixels were assigned proximal, regions with a perimeter larger than 1000 were assigned distal and the rest were assigned to middle. Spots on the spatial transcriptomic sections were assigned to the closest region that overlapped with the zone annotation from the QuPath analysis.

To identify the proportion of different cell types on each spot on the slide we used CARD version 1.0.0 (*84*). Only genes from the single-cell analysis with an average $\log_2$-fold >2 in at least one cluster was kept to deconvolute the spatial spots. Only genes with at least 200 counts and found in at least 50 spots in the spatial transcriptomics data was kept for the deconvolution analysis. Cell type proportions were estimated for each spot in the major ampullate gland. The average proportion for the five classes zone C, zone B and the subclasses proximal, middle, and distal of zone A was calculated by taking the average cell type proportions from all spots that belonged to the five classes.

Sample preparation for proteomics

Major ampullate glands

The spiders were anesthetized and dissected as mentioned earlier. The major ampullate glands were carefully pulled out by holding the duct using micro tweezers. A cut was made in the sac of the gland allowing the dope to flow out for 15 min. The glands were washed three times with PBS and then transferred to a low-binding 1.5 mL microtube (Axygen) containing 60 µL of 8M urea in 20 mM Tris-HCl at pH 8, vortexed and sonicated in an ultrasonic bath sonicator (VWR) for 30 min at room temperature. The samples were stored at -20°C until further use. Three biological replicates were used for the final protein sequencing.

Sample aliquots were supplemented with 0.2% ProteaseMAX (Promega) in 20% ACN/20 mM Tris-HCl, pH 8 to obtain 4M urea concentration before water bath sonication for 5 min. Proteins were reduced with 8 mM DTT incubated at 24°C for 1 h with 550 rpm and alkylated with 20 mM chloroacetamide (CAA) incubated for 1 hour at RT in dark. Digestion was started with addition of 2 µg of LysC (Wako, Japan) incubated at 24°C for 2 h and completed with 2 µg sequencing grade modified trypsin (Promega) incubated at 37°C overnight (ca 16 h). Following centrifugation, the supernatants were collected, and proteolysis was stopped with 5% formic acid, and the samples were cleaned on a C18 Hypersep plate with 40 µL bed volume (Thermo Fisher Scientific) and dried using a vacuum concentrator (Eppendorf).

Major ampullate silk fibers

For collecting the silk, each spider was first anesthetized with CO_2 and gently pinned down to immobilize it, without injuring the animal. By using a Zeiss Stemi 305 stereo microscope, the major ampullate silk was identified, pulled out from the anterior spinneret (*85*) with the help of a tweezer. The silk was collected by rolling it onto a frame attached to a rotating wheel until the spider refused to spin

silk. Several spiders were used to collect enough amount of silk for the experiments. After silking, the spiders were fed with fruit flies, watered, and not used again for the next two weeks.

The collected silk was treated in three different ways for solubilization since it has been reported that the proteins detected might vary depending on the treatment procedure (*22*). About 450 µg of silk was taken for each set of samples and every treatment was done in triplicates. Three different methods were used to prepare silk samples. In the first method, the silk was dissolved by adding 100 µL of HFIP and brief vortexing. The samples were then sonicated in an ultrasonic bath sonicator (VWR) for 30 min at room temperature. The HFIP was evaporated on a Centrivap concentrator system (Labconco). The protein was resuspended in 60 µL of 8M urea in 20 mM Tris-HCl, pH 8 and stored at -20°C until further use. In the second treatment the silk was dissolved in 60 µL of 8M urea in 20 mM Tris-HCl at pH 8, sonicated and stored as mentioned above. For the third method, the silk was dissolved in 60 µL of 9M LiBr in 20 mM Tris-HCl at pH 8, sonicated and stored as mentioned above. Low-binding 1.5 mL microtubes (Axygen) were used throughout the experiments.

An aliquot of 30 µL samples (ca 10 µg) was taken to further preparation. From samples with the HFIP and urea methods, proteins were reduced with 3 µL of 100 mM DTT, incubated at 37°C for 3 h with 1200 rpm and alkylated with 5 µL of 500 mM CAA incubated for 30 min at room temperature in dark. Half of the samples were supplemented with 19 µL of 50 mM Tris-HCl at pH 8.5 and digested with addition of 1 µg of LysC (Wako, Japan) incubated at 24°C for 2 h. Digestion was continued with 1 µg sequencing-grade modified trypsin (Promega) after addition of 56 µL of Tris-HCl and incubated at 37°C overnight (ca 16 h). Samples with the third method (LiBr) were prepared similarly, except for 2 µL of 500 mM DTT was used for reduction incubated at 95°C for 30 min with shaking at 12,500 rpm. Alkylation with 5 µL of 500 mM CAA (as above) was followed by digestion with LysC and trypsin as described above except for that 3 µg trypsin was used. The digestion of all samples was stopped with 6.5 µL concentrated formic acid, the samples were cleaned on a C18 Hypersep plate with 40 µL bed volume (Thermo Fisher Scientific), and dried using a vacuum concentrator (Eppendorf).

Layer-wise dissolution of major ampullate silk fibers

The major ampullate silk was collected by allowing each spider to fall freely from a wooden frame and the extruded silk was rolled on to the same wooden frame. The silk from several individuals was collected in pre-weighed low-binding microtubes, which were measured again to determine the weight of the collected silk. After forceful silking, the spiders were fed with fruit flies, watered, and were not used again for the next two weeks. The collected silk was divided into 3 sets treated with: 1) 2M urea in 50 mM Tris-HCl and 0.5M NaCl, 2) 4M urea in 50 mM Tris-HCl and 0.5M NaCl, and 3) 8M urea in 50 mM Tris-HCl and 0.5M NaCl. Samples were sonicated at room temperature for 2 – 3 h followed by centrifugation at 17,000 *g*. The supernatant was collected and stored at -20°C before using for proteomics analysis. To make sure that the results obtained from sequential solubilization of the silk was not solely due to the relative solubility of the proteins in urea, the following experiment was performed: Two separate samples of the silk weighing 600 mg each were solubilized in 300 mL of formic acid each by vortexing. Each solubilized silk sample was then divided in three low-binding microtubes and subjected to speed vacuum for an hour. The obtained dried samples were further subjected to treatment with: 1) 2M urea in 50 mM Tris-HCl, 2) 4M urea in 50 mM Tris-HCl, and 3) 8M urea in 50 mM Tris-HCl, respectively. The urea-treated samples were sonicated at room temperature for 2 – 3 h followed by centrifugation at 17,000 *g*. The supernatant was collected and used for proteomics analysis.

Liquid Chromatography-Tandem Mass Spectrometry Data Acquisition

Peptides were reconstituted in solvent A and injected on a 50 cm long EASY-Spray C18 column (Thermo Fisher Scientific) connected to an UltiMate 3000 nano-flow UPLC system (Thermo Fisher Scientific) using a 90 min long gradient: 4-26% of solvent B (98% acetonitrile, 0.1% formic acid) in 90 min, 26-95% in 5 min, and 95% of solvent B for 5 min at a flow rate of 300 nL/min. Mass spectra were

acquired on a Q Exactive HF hybrid quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific) ranging from *m/z* 375 to 1800 at a resolution of R=120,000 (at *m/z* 200) targeting $5x10^6$ ions for maximum injection time of 100 ms, followed by data-dependent higher-energy collisional dissociation (HCD) fragmentations of precursor ions with a charge state 2+ to 7+, using 45 s dynamic exclusion. The tandem mass spectra of the top 17 precursor ions were acquired with a resolution of R=30,000, targeting $2x10^5$ ions for maximum injection time of 54 ms, setting quadrupole isolation width to 1.4 Th and normalized collision energy to 28%.

Analysis of proteomics data

Acquired raw data files were converted to Mascot Generic File (*mgf)* format using an in-house developed tool, Raw2MGF (version 2.1.3), and searched with Mascot Daemon version 2.5.1 (Matrix Science Ltd., UK) against a protein database obtained from 22860 protein entries. A maximum of two missed cleavage sites were allowed for full tryptic digestion, while setting the precursor and the fragment ion mass tolerance to 10 ppm and 0.02 Da, respectively. Carbamidomethylation of cysteine was specified as a fixed modification, while oxidation on methionine as well as deamidation of asparagine and glutamine were set as dynamic modifications. The search results were imported into Scaffold version 4.11 (Proteome Software Inc.) to calculate the contribution of each protein to the total sum of spectra (percentage of total spectra). For all proteins where there was no report of a percentage of total spectra for a sample the percentage of total spectra value was set to 0.

To identify the proteins in the silk, we used MS/MS data from both the silk fibers and the glands. To determine their presence in the gland, the average percent of total spectra for each protein was calculated by taking the mean of the three gland samples. All proteins that were reported in at least one of the samples were considered to be present in the gland. The average percent of total spectra for each protein in the silk was calculated similarly by taking the mean of the nine samples, *i.e.,* three biological replicates for the three different solvents (HFIP, LiBr, urea). For a protein to be considered present in the major ampullate silk, four criteria had to be fulfilled: 1) it had to be present in the gland, 2) it had to be found in at least two of the detergents, 3) it had to possess a signal-peptide predicted at the 5' end of its aa sequence, and 4) it had to have an average percent of total spectra greater than 0.15. After applying these criteria, 18 proteins remained. The relative molecular abundance of proteins in the three HFIP silk samples was calculated using three different methods. MaxQuant (version 2.5.2.0) employed intensity-based absolute quantification (iBAQ) and label-free quantification (LFQ) to determine protein relative abundance. Additionally, Normalized Spectral Abundance Factor (NSAF) values, based on spectral counts, were obtained using Scaffold (version 5.3.3). Statistical differences in the percentage of total spectra for a protein between different urea concentrations were determined using an unpaired t-test, focusing on these 18 proteins. Relative percent of total spectra for the 18 silk proteins were estimated for urea extracts from fibers dissolved in formic acid and compared with those obtained from urea extracts of intact silk fibers. Pairwise Pearson correlation (r) based on the relative abundance of the total spectra of the 18 silk proteins was calculated between all urea extract samples, both intact and dissolved silk. Hierarchical clustering of all the samples was carried out based on the Pearson correlation values to identify similarities in expression patterns between the samples.

Tissue processing for histological analysis

Spiders were anesthetized with CO_2, then dissected at the pedicel on an ice-cooled wax plate using 154 mM sodium chloride solution (Fresenius Kabi AG, Germany). Dissections were performed using a Leica M60 stereomicroscope with a Leica IC80 HD camera. Whole opisthosomas were fixed in 2.5% glutaraldehyde in 67 mM phosphate buffer, pH 7.2 at 4°C for 24 h, and later rinsed in 67 mM phosphate buffer. Tissues were dehydrated in graded ethanol (50, 70, 90, and 100%, for 30 minutes each), then infiltrated and embedded in water-soluble glycol methacrylate (Leica Historesin). Sections of 2 µm thickness were obtained using a Leica RM 2165 microtome with glass knives. The sections were stained with hematoxylin and eosin and mounted using Agar 100 resin. Evaluation was performed using a

Nikon Microphot-FXA (Tekno Optik AB) microscope equipped with a Nikon FX-35DX camera. Images were captured and edited using the software Eclipse Net version 1.20.0.

Tensile tests of major ampullate silk fibers

Major ampullate silk fibers were reeled and mounted on cardboard frames with a square window of 1 x 1 cm (gauge length 1 cm). The diameters were measured by means of light microscopy using a Nikon Eclipse Ts2R-FL inverted microscope. The diameter was measured prior to the tensile test at five locations along the fiber and then averaged. The tensile tests were performed with a 5943-Instron machine (USA) equipped with a 5N load cell. Astrain rate of 6 mm/min was used. Load-displacement curves were converted into engineering stress-strain curves assuming a circular cross-section and using the average diameter for each fiber.

Acknowledgements

The authors would like to acknowledge the following individuals and facilities for their valuable contributions to this research: Orlando Contreras-López (NGI Stockholm), for assistance with Bulk RNA data, Elísabet Einarsdóttir (NGI Stockholm) for the technical guidance, assistance with ordering, and help with invoices, and Christian Tellgren-Roth (NGI, SciLifeLab, Uppsala) for their expertise in PacBio Genome and IsoSeq data analysis. We acknowledge the valuable contributions of Karolina Wallenborg and Matilda Eriksson at the single-cell sequencing facility, Science for Life Laboratory (SciLifeLab), Karolinska Institutet (KI), Sweden, for their help in single-cell sample preparations and sample QC. We also thank Remi-Andre Olsen from the Spatial transcriptomics facility at SciLifeLab, Stockholm, for preliminary analysis of the spatial data and Mattias Ormestad from BioNut at KI for technical details relating to the spatial transcriptomics. Protein analyses were carried out by the Proteomics Biomedicum core facility, KI (https://ki.se/en/mbb/proteomics-biomedicum) with kind help from Akos Vegvari.

We thank Maria Lindskog and Christian Broberger for kindly providing lab space and equipment for spider dissections and single-cell dissociation, and Elisabeth Poirier for assistance with collection of silk from spiders. The authors would like to acknowledge Jens Hjerling-Leffler and Björn Nystedt for engaging in fruitful discussions.

The authors appreciate the support of the National Genomics Infrastructure (NGI) / Uppsala Genome Center and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure for transcriptome and PacBio sequencing. Work performed at NGI / Uppsala Genome Center has been funded by RFI / VR and SciLifeLab, Sweden. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by SciLifeLab, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing of 10X Visium spatial transcriptomics and access to the UPPMAX computational Infrastructure. The authors acknowledge the Eukaryotic Single Cell Genomics (ESCG) facility in Stockholm funded by SciLifeLab, KI Core and StratRegen. The authors appreciate the kind support from data management team at SciLifeLab/NBIS (LTS1810 and 6337) and would like to thank Erik Hedman, Yvonne Kallberg, Wolmar Nyberg Åkerström and Stephan Nylinder, for assisting with the submission of raw and processed data obtained in this work from various platforms.

Funding

This work was supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ARTSILK, grant agreement No 815357), FORMAS (2019-00427 and 2023-01313), Olle Engkvist stiftelse (207-0375), Knut and Alice Wallenbergs stiftelse (23.0331) and the Swedish Research Council (2019-01257) to AR. GG acknowledges support from the Wenner-Gren Foundations (prot. UPD2021-0047) and "EPASS" under the HORIZON TMA MSCA Postdoctoral Fellowships - European Fellowships (project number 101103616).

Author contributions
AR, JR, and JJ conceptualized and designed the project. AR acquired funds for the project. Resources were provided by AR, OVP, UC, AM, LH, NF, IB, GG, and SS. Sample preparation for genome sequencing was performed by NF. OVP provided experimental design for reference genome and long-read RNA sequencing. Sample preparations for bulk RNA sequencing were performed by SH and SS. Single-cell experiments were optimized and performed by UC and SS. For spatial transcriptomics, sample preparation was done by SS, and library preparation and sequencing were performed by AM. UC optimized and prepared major ampullate gland and silk fiber samples for proteomic studies. Mechanical testing of the fibers was performed by GG. LH assisted in sample preparation and analysis of histological studies performed by SS. AC and HL assembled the genome. LS, IB, AC, and HL annotated the genome. IB performed PacBio ISOseq data analysis. JR, SS, and SH analyzed the bulk RNA data. JR and SH analyzed the single-cell RNA data. JR analyzed the proteomic data. JR and SS performed the spatial transcriptomics and QuPath analysis. Data curation was done by JR, SH, LS, and NF. AR, JR, JJ, and HL supervised the project. Software was developed by AC, JR, SH, NF, and SS. The original draft and figures of the manuscript were prepared by SH, SS, JR, and AR with input from HL, LH, UC, and NF. All authors contributed to revision and editing.

Competing interests
S.S., J.R., and A.R. are listed as inventors on a patent application (serial number SE 2450315-3, filed on 21 March 2024). All other authors declare that they have no competing interests.

Data and materials availability
All data needed to evaluate the conclusions in the paper are available in the paper and/or the Supplementary Materials. Raw sequence reads used for genome assembly and annotation have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the following accession numbers: PacBio genome data (PRJEB74310), 10x Chromium genome data (PRJEB74311), genome annotation (PRJEB74312), IsoSeq PacBio sequencing of major ampullate glands (PRJEB74494). Short reads from bulk RNA sequencing of silk glands, head and abdomen are available in ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-14073 and E-MTAB-14077. The raw data generated from single-cell sequencing of major ampullate glands and spatial transcriptomics of abdomen sections are available in the ArrayExpress under E-MTAB-14078 and E-MTAB-14082, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*86*) partner repository with the dataset identifier PXD052116. The code used for data analysis in this study is described in Materials and Methods and is publicly available in Zenodo (https://doi.org/10.5281/zenodo.11184961) and GitHub (https://github.com/AnnaRisingLab/Rising-Lab/tree/v1.0.0).

Supplementary materials
The supplementary data related to this article can be found at https://www.science.org/doi/10.1126/sciadv.adn0597.

References

1. A. Rising, M. J. Harrington, Biological Materials Processing: Time-Tested Tricks for Sustainable Fiber Fabrication. *Chem Rev* **123**, 2155-2199 (2023).
2. K. Bourzac, Spiders: Web of intrigue. *Nature* **519**, S4-6 (2015).
3. L. Brunetta, C. L. Craig, *Spider silk: evolution and 400 million years of spinning, waiting, snagging, and mating* (Yale University Press, 2010).
4. D. B. Peakall, Synthesis of silk, mechanism and location. *American Zoologist* **9**, 71-79 (1969).
5. F. Vollrath, Strength and structure of spiders' silks. *J Biotechnol* **74**, 67-83 (2000).
6. N. A. Ayoub, J. E. Garb, R. M. Tinghitella, M. A. Collin, C. Y. Hayashi, Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2**, e514 (2007).
7. P. L. Babb, N. F. Lahens, S. M. Correa-Garhwal, D. N. Nicholson, E. J. Kim, J. B. Hogenesch, M. Kuntner, L. Higgins, C. Y. Hayashi, I. Agnarsson, B. F. Voight, The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nature Genetics* **49**, 895-903 (2017).
8. N. Kono, H. Nakamura, R. Ohtoshi, D. A. P. Moran, A. Shinohara, Y. Yoshida, M. Fujiwara, M. Mori, M. Tomita, K. Arakawa, Orb-weaving spider Araneus ventricosus genome elucidates the spidroin gene catalogue. *Scientific reports* **9**, 8380 (2019).
9. N. Kono, H. Nakamura, M. Mori, Y. Yoshida, R. Ohtoshi, A. D. Malay, D. A. Pedrazzoli Moran, M. Tomita, K. Numata, K. Arakawa, Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proc Natl Acad Sci U S A* **118**, 1-10 (2021).
10. N. Kono, R. Ohtoshi, A. D. Malay, M. Mori, H. Masunaga, Y. Yoshida, H. Nakamura, K. Numata, K. Arakawa, Darwin's bark spider shares a spidroin repertoire with Caerostris extrusa but achieves extraordinary silk toughness through gene expression. *Open Biol* **11**, 210242 (2021).
11. P. L. Babb, M. Gregoric, N. F. Lahens, D. N. Nicholson, C. Y. Hayashi, L. Higgins, M. Kuntner, I. Agnarsson, B. F. Voight, Characterization of the genome and silk-gland transcriptomes of Darwin's bark spider (Caerostris darwini). *PLoS One* **17**, e0268660 (2022).
12. K. Arakawa, N. Kono, A. D. Malay, A. Tateishi, N. Ifuku, H. Masunaga, R. Sato, K. Tsuchiya, R. Ohtoshi, D. Pedrazzoli, A. Shinohara, Y. Ito, H. Nakamura, A. Tanikawa, Y. Suzuki, T. Ichikawa, S. Fujita, M. Fujiwara, M. Tomita, ..., K. Numata, 1000 spider silkomes: Linking sequences to silk physical properties. *Sci Adv* **8**, (2022).
13. W. Hu, A. Jia, S. Ma, G. Zhang, Z. Wei, F. Lu, Y. Luo, Z. Zhang, J. Sun, T. Yang, T. Xia, Q. Li, T. Yao, J. Zheng, Z. Jiang, Z. Xu, Q. Xia, Y. Wang, A molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk. *Nat Commun* **14**, 837 (2023).
14. J. E. Garb, N. A. Ayoub, C. Y. Hayashi, Untangling spider silk evolution with spidroin terminal domains. *BMC Evol Biol* **10**, 243 (2010).
15. M. Xu, R. V. Lewis, Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci U S A* **87**, 7120-7124 (1990).
16. M. B. Hinman, R. V. Lewis, Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes dragline silk is a two-protein fiber. *Journal of Biological Chemistry*, (1992).
17. M. A. Collin, T. H. Clarke, 3rd, N. A. Ayoub, C. Y. Hayashi, Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. *International Journal of Biological Macromolecules* **113**, 829-840 (2018).
18. J. E. Garb, R. A. Haney, E. E. Schwager, M. Gregoric, M. Kuntner, I. Agnarsson, T. A. Blackledge, The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness. *Communications Biology* **2**, 275 (2019).
19. M. Saric, L. Eisoldt, V. Doring, T. Scheibel, Interplay of Different Major Ampullate Spidroins during Assembly and Implications for Fiber Mechanics. *Adv Mater* **33**, e2006499 (2021).
20. C. Y. Hayashi, R. V. Lewis, Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Journal of Molecular Biology* **275**, 773-784 (1998).
21. K. W. Sanggaard, J. S. Bechsgaard, X. Fang, J. Duan, T. F. Dyrlund, V. Gupta, X. Jiang, L. Cheng, D. Fan, Y. Feng, L. Han, Z. Huang, Z. Wu, L. Liao, V. Settepani, I. B. Thogersen, B.

Vanthournout, T. Wang, Y. Zhu, ..., J. Wang, Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun* **5**, 3765 (2014).
22. C. Larracas, R. Hekman, S. Dyrness, A. Arata, C. Williams, T. Crawford, C. A. Vierra, Comprehensive Proteomic Analysis of Spider Dragline Silk from Black Widows: A Recipe to Build Synthetic Silk Fibers. *Int J Mol Sci* **17**, 1-16 (2016).
23. T. Pham, T. Chuang, A. Lin, H. Joo, J. Tsai, T. Crawford, L. Zhao, C. Williams, Y. Hsia, C. Vierra, Dragline silk: a fiber assembled with low-molecular-weight cysteine-rich proteins. *Biomacromolecules* **15**, 4073-4081 (2014).
24. J. Kovoor, "Comparative Structure and Histochemistry of Silk-Producing Organs in Arachnids" in *Ecophysiology of Spiders* (1987), pp. 160-186.
25. F. Vollrath, D. P. Knight, Structure and function of the silk production pathway in the spider Nephila edulis. *Int J Biol Macromol* **24**, 243-249 (1999).
26. M. Andersson, L. Holm, Y. Ridderstrale, J. Johansson, A. Rising, Morphology and composition of the spider major ampullate gland and dragline silk. *Biomacromolecules* **14**, 2945-2952 (2013).
27. S. F. Li, A. J. McGhie, S. L. Tang, New Internal Structure of Spider Dragline Silk Revealed by Atomic Force Microscopy. *Biophysical journal* **66**, 1209-1212 (1994).
28. F. Vollrath, T. Holtet, H. C. Thøgersen, S. Frische, Structural organisation of spider silk. *Proceedings of the Royal Society of London* **263**, 147-151 (1996).
29. S. Frische, A. B. Maunsbach, F. Vollrath, Elongate cavities and skin-core structure in Nephila spider silk observed by electron microscopy. *Journal of Microscopy* **189**, 64-70 (1998).
30. K. Augsten, P. Muhlig, C. Herrmann, Glycoproteins and skin-core structure in Nephila clavipes spider silk observed by light and electron microscopy. *Scanning* **22**, 12-15 (2000).
31. P. Poza, J. Pérez-Rigueiro, M. Elices, J. Llorca, Fractographic analysis of silkworm and spider silk. *Engineering Fracture Mechanics* **69**, 1035-1048 (2002).
32. A. Sponner, W. Vater, S. Monajembashi, E. Unger, F. Grosse, K. Weisshart, Composition and hierarchical organisation of a spider silk. *PLoS One* **2**, e998 (2007).
33. I. Iachina, J. Fiutowski, H. G. Rubahn, F. Vollrath, J. R. Brewer, Nanoscale imaging of major and minor ampullate silk from the orb-web spider Nephila Madagascariensis. *Sci Rep* **13**, 6695 (2023).
34. A. Sponner, E. Unger, F. Grosse, K. Weisshart, Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat Mater* **4**, 772-775 (2005).
35. D. H. Hijirida, K. G. Do, C. Michal, S. Wong, D. Zax, L. W. Jelinski, 13C NMR of Nephila clavipes major ampullate silk gland. *Biophysical journal* **71**, 3442-3447 (1996).
36. M. Andersson, G. Chen, M. Otikovs, M. Landreh, K. Nordling, N. Kronqvist, P. Westermark, H. Jörnvall, S. Knight, Y. Ridderstråle, L. Holm, Q. Meng, K. Jaudzems, M. Chesler, J. Johansson, A. Rising, Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biology* **12**, 1-14 (2014).
37. F. Hagn, L. Eisoldt, J. G. Hardy, C. Vendrely, M. Coles, T. Scheibel, H. Kessler, A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* **465**, 239-242 (2010).
38. N. Kronqvist, M. Otikovs, V. Chmyrov, G. Chen, M. Andersson, K. Nordling, M. Landreh, M. Sarr, H. Jornvall, S. Wennmalm, J. Widengren, Q. Meng, A. Rising, D. Otzen, S. D. Knight, K. Jaudzems, J. Johansson, Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nat Commun* **5**, 3254 (2014).
39. M. Landreh, G. Askarieh, K. Nordling, M. Hedhammar, A. Rising, C. Casals, J. Astorga-Wells, G. Alvelius, S. D. Knight, J. Johansson, H. Jörnvall, T. Bergman, A pH-Dependent Dimer Lock in Spider Silk Protein. *Journal of Molecular Biology* **404**, 328-336 (2010).
40. A. Rising, J. Johansson, Toward spinning artificial spider silk. *Nat Chem Biol* **11**, 309-315 (2015).
41. J. Sparkes, C. Holland, Analysis of the pressure requirements for silk spinning reveals a pultrusion dominated process. *Nature Communications* **8**, 1-10 (2017).
42. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A.

Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, ..., D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
43. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zidek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, ..., S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444 (2022).
44. P. L. Stahl, F. Salmen, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, A. Borg, F. Ponten, P. I. Costea, P. Sahlen, J. Mulder, O. Bergmann, ..., J. Frisen, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).
45. L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, (2018).
46. P. Bankhead, M. B. Loughrey, J. A. Fernandez, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, P. W. Hamilton, QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
47. R. C. Chaw, S. M. Correa-Garhwal, T. H. Clarke, N. A. Ayoub, C. Y. Hayashi, Proteomic Evidence for Components of Spider Silk Synthesis from Black Widow Silk Glands and Fibers. *Journal of proteome research* **14**, 4223-4231 (2015).
48. K. Yazawa, A. D. Malay, H. Masunaga, K. Numata, Role of Skin Layers on Mechanical Properties and Supercontraction of Spider Dragline Silk Fiber. *Macromolecular Bioscience* **19**, (2019).
49. B. Madsen, Z. Z. Shao, F. Vollrath, Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int J Biol Macromol* **24**, 301-306 (1999).
50. S. Sonavane, P. Westermark, A. Rising, L. Holm, Regionalization of cell types in silk glands of Larinioides sclopetarius suggest that spider silk fibers are complex layered structures. *Sci Rep* **13**, 22273 (2023).
51. K. Jaudzems, G. Askarieh, M. Landreh, K. Nordling, M. Hedhammar, H. Jornvall, A. Rising, S. D. Knight, J. Johansson, pH-dependent dimerization of spider silk N-terminal domain requires relocation of a wedged tryptophan side chain. *Journal of Molecular Biology* **422**, 477-487 (2012).
52. S. Keten, M. J. Buehler, Atomistic model of the spider silk nanostructure. *Applied Physics Letters* **96**, (2010).
53. W. Lu, D. L. Kaplan, M. J. Buehler, Generative Modeling, Design, and Analysis of Spider Silk Protein Sequences for Enhanced Mechanical Properties. *Advanced Functional Materials* **34**, (2024).
54. C. S. Chin, P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, M. C. Schatz, Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**, 1050-1054 (2016).
55. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
56. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
57. G. Marcais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770 (2011).
58. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**, (2020).
59. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

60. M. Bernt, A. Donath, F. Juhling, F. Externbrink, C. Florentz, G. Fritzsch, J. Putz, M. Middendorf, P. F. Stadler, MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* **69**, 313-319 (2013).
61. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
62. F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
63. B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado, M. Yandell, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188-196 (2008).
64. T. M. Lowe, S. R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964 (1997).
65. E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).
66. S. W. Burge, J. Daub, R. Eberhardt, J. Tate, L. Barquist, E. P. Nawrocki, S. R. Eddy, P. P. Gardner, A. Bateman, Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41**, D226-232 (2013).
67. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**, W309-312 (2004).
68. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
69. T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B. L. Pinto, G. A. Salazar, M. L. Bileschi, P. Bork, A. Bridge, L. Colwell, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D. A. Natale, C. A. Orengo, A. P. Pandurangan, C. Rivoire, ..., A. Bateman, InterPro in 2022. *Nucleic Acids Res* **51**, D418-D427 (2023).
70. T. Smith, A. Heger, I. Sudbery, UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491-499 (2017).
71. E. Lee, G. A. Helt, J. T. Reese, M. C. Munoz-Torres, C. P. Childers, R. M. Buels, L. Stein, I. H. Holmes, C. G. Elsik, S. E. Lewis, Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* **14**, R93 (2013).
72. F. Teufel, J. J. A. Armenteros, A. R. Johansen, M. H. Gíslason, S. I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology* **40**, 1023-+ (2022).
73. K. Okonechnikov, O. Golosova, M. Fursov, U. team, Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166-1167 (2012).
74. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res* **14**, 988-995 (2004).
75. O. Keller, F. Odronitz, M. Stanke, M. Kollmar, S. Waack, Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *Bmc Bioinformatics* **9**, (2008).
76. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).
77. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
78. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
79. Y. Liao, G. K. Smyth, W. Shi, The R package *Rsubread* is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* **47**, (2019).
80. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, (2014).
81. E. A. Thévenot, A. Roux, Y. Xu, E. Ezan, C. Junot, Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *Journal of Proteome Research* **14**, 3322-3335 (2015).
82. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* **33**, 495-U206 (2015).

83. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, ..., R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).
84. Y. Ma, X. Zhou, Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol* **40**, 1349-1359 (2022).
85. R. Foelix, Biology of Spiders: Oxford University Press. *New York* **330**, (1996).
86. Y. Perez-Riverol, J. Bai, C. Bandla, D. Garcia-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J. A. Vizcaino, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552 (2022).

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:56

Spider silk is a high-performance fiber with potential for diverse applications. Particularly, major ampullate silk has high tensile strength and toughness, while minor ampullate silk displays higher extensibility and water resistance. Replicating these properties in artificial silks necessitates a comprehensive understanding of spider silk biology. In this thesis, the molecular basis of silk production in the major and minor ampullate glands of the bridge spider, *Larinioides sclopetarius*, were investigated using histological and multi-omics approaches.

Sumalata Sonavane received her graduate education at the Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala. Sumalata obtained her undergraduate degree from University of Mumbai, India.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-052-1
ISBN (electronic version) 978-91-8046-053-8