

Doctoral Dissertation (Shinshu University)

Study on spider silk genes and transgenic  
silkworm including the spider silk gene.

September 2013

Han Leng

Doctoral Dissertation (Shinshu University)

Study on spider silk genes and transgenic  
silkworm including the spider silk gene

Principal advisor: Professor Masao Nakagaki

September 2013

Han Leng

## Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor, Professor Masao Nakagaki, a respectable, responsible and resourceful scholar, who has provided me with valuable guidance in every stage of the writing of this thesis. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also in my future study.

I shall extend my thanks to Professor Zhonghuai Xiang, Qinyou Xia, Ping Zhao and Aichun Zhao, for providing me the chance to study in Japan.

I would also like to thank all my professors: Tsukada, Hayashida, Kajiura, Shiomi, Shirai, Hochi and Gotoh, who have helped me to develop the fundamental and essential academic competence.

My sincere appreciation also goes to the teachers and students: Tianfu Zhao, Yujun Wang and Lei Zhang, who participated this study with great cooperation.

Last but not least, I'd like to thank my family for their support and faith through the three years.

*Leng Han*

*H25 Ueda*

## Contents

Acknowledgements .....	3
General Introduction .....	5
Chapter 1: The research of the spider silk genes .....	10
Paragraph 1: Analysis of a new type of Major Ampullate Spider Silk gene, <i>MaSp1s</i> .11	
Materials and Methods.....	12
Results .....	14
Discussion.....	25
Paragraph 2: Reanalyzing the <i>AbCySp2</i> gene of Wasp Spider <i>Argiope bruennichi</i> . ....	30
MATERIALS AND METHODS.....	31
RESULTS .....	32
DISCUSSION.....	38
Chapter 2 Expression dragline silk protein in the silkworm using transgenic method	41
Paragraph 1: Generation of transgenic silkworms which including the MaSp.....	42
MATERIALS AND METHODS.....	43
Result .....	46
Paragraph 2: Biological property of the transgenic silkworms which including	
MaSp1s.....	49
MATERIALS AND METHODS.....	49
Results and discussion .....	52
Conclusions.....	58
Paragraph 3: Physical properties and Structure of silk from the transgenic silkworm	
.....	59
MATERIALS AND METHODS.....	59
Results .....	62
Conclusion .....	65
Summary.....	66

## General Introduction

Over the 400 million years of evolution, as of 2012, 43,244 spider species, 3879 genera and 111 families have been recorded by taxonomists (The World Spider Catalog, Version 13.0). Spider silk is a protein fiber spun by spiders. Spidroins are proteins from spider silks that have been studied for over 50 years(1). All spiders produce silks, spiders make their webs and perform a wide range of tasks ranging from three to seven different types of silk fiber. These different fibers allow a comparison of structure with function, because each silk has distinct mechanical properties and is composed of peptide modules that confer those properties(2). Araneoid spiders use specialized abdominal glands to manufacture distinct fibers and glues(3, 4). The first is Major Ampullate silk (MAA)—Dragline silk—used for the web's outer rim and spokes and the lifeline. The second is Minor Ampullate silk (MIA) which is used for temporary scaffolding during web construction. From mechanical property perspective, both major and minor silks have high tensile strength, but MIA silks irreversibly deform when stretched, while MAA silks are more elastic(4). The third is Flag silk which is used for the capturing lines of the web. Flag silk is highly elastic (up to 300%) and perfectly dissipates the impact energy of prey(5). The forth is Egg case silk which is used for make protective egg sacs. The scanning electron micrograph shows that egg case silk is composed of fibers of two different diameters. The larger diameter fibers, which represent the major component of egg case silk, are produced by the tubuliform gland. The diameter of the large fibers was 4–5  $\mu\text{m}$ , whereas the diameters the smaller fibers were on the order of 500 nm. The smaller diameter fibers, which constitute a minor component, are likely synthesized by the aciniform gland(6). The fifth silks are from Aciniform gland. Those silks are used to wrap and secure freshly captured prey(7); used in the male sperm webs(8); used in stabilimenta (7, 9). The sixth is silk glue comes from

Aggregate gland. The spider produces a glue to coat the spiral silk. it is the outer part of sticky silk-droplets of an adhesive substance are deposited along the threads.(10) The seventh is silk from Pyriform gland as the major constituent of attachment discs.(11) I must point out that: all spiders produce silks, not all spiders weave webs.

Most molecular studies of spider silk have focused on fibroins from dragline silk and capture silk, two important silk types involved in the survival of the spider, particularly the dragline silk.(6) Because of its extremely high tensile strength and toughness, dragline (major ampullate) silk has received the most attention of the spider silks(12). The pressure of natural selection has made spider dragline silk one of the most attractive materials available to date. The fiber is stronger than steel and has a tensile strength approaching that of Kevlar (13, 14). Unlike the latter, however, spider silk is also characterized by an extremely high elasticity. That unusual combination of high strength and stretch leads to toughness values never attained in synthetic high performance fibers. Other attractive features of spider silk include excellent compressive properties, ease of dye ability, extensive softening and supercontraction in humid environments, and foremost the ability to easily dissolve in aqueous solutions at room temperature (15). The latter is in contrast to synthetic fibers which, for comparable performance, require processing in very caustic solvents and/or at elevated temperatures. High strength synthetic fibers also require a post spinning draw of several hundred percent, which is totally absent in spider silk dragline(14).

The dragline has a composite nature and is predominantly made up by two proteins, the major ampullate spidroins 1 and 2(16-18), which can be considered natural block copolymers (19, 20). Whereas spidroin 1 is distributed almost uniformly within the fiber core, spidroin 2 is missing in the periphery and is tightly

packed in certain core areas which suggest that the role of spidroin 2 in the spinning process could be to facilitate the formation of fibrils and contribute directly to the elasticity of the silk(20). Short glycine-rich regions (GGX, where X represents a subset of aa) followed by a stretch of multiple alanines (poly-A) characterize both proteins. The ubiquitous poly-A stretches are hypothesized to form hydrophobic crystalline domains that are responsible for the high tensile strength of the fiber (12, 21-24). In contrast, the glycine-rich regions are hydrophilic with runs of the peptide motif GGX conforming to a 3<sub>1</sub>-helix(25, 26). While poly-A and GGX motifs describe almost all of MaSp1, MaSp2 also has a large proportion of GPG motifs(18). These proline-containing repeats likely form type II beta-turns, and such kinks in part explain the reversible extensibility of dragline fiber in part explain the reversible extensibility of dragline fiber. (12, 20, 27-29)

Much of the applied research on spider silks has focused on mass-producing silk fibers for industrial use.(30) However, unlike domesticated silkworm, spiders cannot be readily farmed for silk because they are predatory and cannibalistic(12). So, it is very difficult to obtain large amounts of fibrous spider silk from spiders. For example, Simon Peers and Nicholas Godley, a textile artist and a designer-entrepreneur respectively, have created a shimmering golden cape from spider silk. There is 1.5 kg of silk in the cape. It has taken eight years and more than a million Madagascar Golden Orb spiders to create a work of art "with the quality of a fairy story".

Instead, researchers have created biomimetic silks through transgenic technologies using partial-length silk cDNA sequences overexpression in many different types of organisms (2). Such as: Synthetic genes were designed to encode analogs of the two proteins of dragline (MaSp1, MaSp2). Both analogs were produced efficiently in *Escherichia coli*(31), *methylophilic yeast*(32), mammalian cells(33) and higher plants. (34). However, the products only form soluble proteins, which must be

purified and then spun into fibrous silk artificially (35). And what's more, the yield and homogeneity of higher molecular weight silk proteins were found to be limited by truncated synthesis, probably as a result of ribosome termination errors.

The silkworm *Bombyx mori* also synthesizes large amounts of silk proteins in silk glands and spins them out as a fibrous thread to form a cocoon. Insights into the sequences and molecular structures have shown that silkworm silk is very similar to that of spider dragline silk. For example, both of their genes have a high GC ratio, and the proteins are composed of tandem repeats that always contain Gly-rich and Ala-rich domains in their major parts.(36) In 2000, Tamura et al. reported a method for stable germline transformation in silkworms by using a piggyBac transposon-derived vector, inaugurating a new era of Silkworm Transgenesis (36, 37). Furthermore, to date, great progress has been made in research on Silkworm Transgenesis.(38-51) The Silkworm Transgenesis make that using the silkworm to mass-produce silk fibers including spider dragline protein for industrial use possible (36, 52). In comparison with those former hosts, there was no doubt silkworm had a great merit that there is no need for considering the fiber-spinning procedure. Silkworm can spin them out as a fibrous thread to form a cocoon very easy.

Previously, several novel spider silk genes have been cloned and observed their expression in *B. mori* cell lines and larval bodies in our lab(36, 53-56). In my dissertation, I keep on doing some research on the two topics mentioned above(spider silk genes and biomimetic spider silks). Firstly, a completely new clone encoding the third dragline silk fibroin (*Cm.MaSp1s*) has been isolated. During the cloning process of *Cm.MaSp1s*, the genomic DNA sequencing method was proved to be a good way to retrieve information of the 5' end of silk genes, which usually includes the region encoding the signal peptide of the silk protein. So, we used the genomic DNA



sequencing method to reanalyzed the *AbCySp2* gene which was reported as non-signal peptide protein by our lab (Zhao, 2006). The result shown that *AbCySp2* had an apparent signal peptide sequence; actually, the mistake had happened in a former report resulted in a 349 bp deletion event which was caused by intermolecular homologous recombination. Secondly, I have created biomimetic spider silks through transgenic technologies using partial-length silk cDNA sequence(*Cm.MaSp1s*) overexpression in silkworm. Finally, I analyzed the biological properties of the transgenic silkworm and the physical properties of the transgenic silk including spider silk protein.

## **Chapter 1: The research of the spider silk genes**

## **Paragraph 1: Analysis of a new type of Major Ampullate Spider Silk gene, *Cm.MaSp1s***

### Introduction

Araneoid spiders are unique in their production and use of different silks throughout their lifetime. Up to seven different protein-based silk fibers are produced by Orb web-weaving spiders, including Major Ampullate Dragline, Minor Ampullate, Flagelliform, Aciniform, Tubuliform, Aggregate, and Pyriform (57). Among these silks, the Major Ampullate dragline silk receives the most interest from scientists for its unique combination of high tensile strength and elasticity (58). These remarkable mechanical properties make it attractive for industrial, military and medical applications.

However, due to their territorial carnivorous nature, it is impossible to obtain large amounts of fibrous spider silk from spiders (58). Thus, to obtain large quantities of spider silk, molecular biologists have been attempting to clone silk genes for overexpression in many different types of organisms, such as *Escherichia coli*, yeast, mammalian cells, higher plants and silkworms (31-36, 59).

With the development of sequencing technology, more types of spider major ampullate silk genes have been reported. The main types of major ampullate silk genes include two types: major ampullate spidroin 1 and 2 (*MaSp1* and *MaSp2*), and the sizes of the full-length transcripts are approximately 10,000 bp or larger (7, 12, 60, 61). Here, a new type of *MaSp1*, *MaSp1s* (short *MaSp1*), was presented, which deviates from the large-size rule and is different from the previously reported *MaSp1*.

## Materials and Methods

### *Silk protein extraction and SDS-PAGE electrophoresis analysis*

Mature white spiders were collected in the Independent State of Papua New Guinea (PNG). The dragline silk was reeled from its abdominal spinneret and then transported to Japan. The silk was dissolved in 8 M urea by vortexing for 10 min at room temperature, followed by incubation for 2 h at 4°C. The extracted proteins were mixed with 5X loading buffer (with 5%  $\beta$ -mercaptoethanol) and heated for 5 min on a dry block heater at 95°C. The prepared sample was loaded onto denaturing polyacrylamide gels (5% stacking gel, 8% resolving gel) with running conditions of 10 mA for 24 hours.

### *Spider silk gland preparation*

Some of the collected spiders were dissected after collection under a dissection microscope to harvest the silk glands. The harvested silk glands were rapidly immersed in RNAlater solution (*Ambion*) and transported to Japan to be used for RNA isolation and genome DNA extraction (photographs of a spider and its dragline tensile property is shown in(Figure 1.))

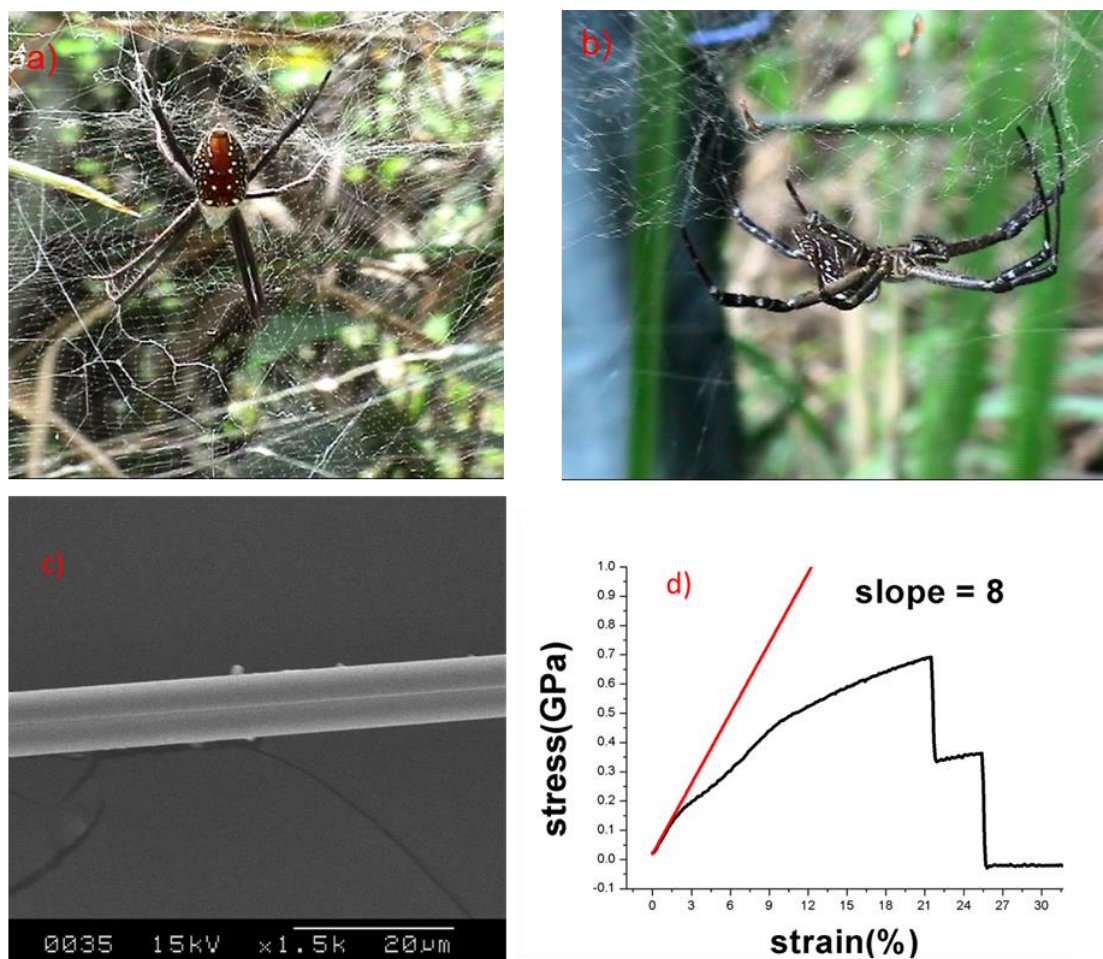


Figure 1. The spider and its dragline tensile property. (a, b) photographs of one of the spiders used in this study; (c) SEM photo of dragline silk at 1500 magnification; (d) tensile property of the dragline.

#### *RNA isolation and cDNA library construction*

The silk glands (200 mg) dissected from *Cyrtophora moluccensis* were completely homogenized in 1 mL lysis buffer. The total RNA and mRNA were extracted from the tissue according to the protocols described in the manual of the Micro-FastTrack™ 2.0 Kit (Invitrogen). The cDNA library construction was accomplished in collaboration with TaKaRa using the Creator™ SMART™ cDNA Library Construction Kit (Clontech, Inc). The strategies for the cDNA library construction are described in detail in the kit manual.

### *DNA Sequencing*

DH10B\* cells (*Invitrogen*) were transformed by electroporation (1.5 kV, 200  $\Omega$ , 25  $\mu$ F, *BIO-RAD*), and 2880 (96X30) recombinant colonies were chosen from the selective medium (LB/Cm plates) for culture. After modified bacterial liquid electrophoresis (62), approximately 1344 (96X14) recombinant plasmids with >1 kb insert fragments (pDNR-LIB/Sif IA and Sif IB) were chosen for plasmid extraction using the Mag-Bind<sup>®</sup> Plasmid 96 Kit (*Omega Bio-Tek*). The chosen colonies were sequenced and analyzed using the ABI Prism genetic analyzer 3100 (*Nippon Gene*). The sequencing primers were as follows:

5'-GCCAAACGAATGGTCTAGAAAGCTTCTC-3' (3' terminal primer) and  
5'-GACGGTACCGGACATATGCC-3' (5' terminal primer).

### *Genomic DNA extraction and inverse PCR (iPCR)*

The genomic DNA was extracted from 25 mg silk gland using the DNeasy Blood & Tissue Kit (*QIAGEN*), digested with *Sau3A* I (*TaKaRa*) and purified using the Wizard<sup>®</sup> SV Gel and PCR Clean-Up System (*Promega*). The products of the restriction enzyme digestion were ligated using the DNA Ligation Kit Ver.2.0 (*TaKaRa*) for iPCR. The primers for iPCR were as follows: I-Right, 5'-AACGAAAGGAGCAGTGAACATAA-3', and

I-Left, 5'-GTAAATGCTCCTGATTGTCCTAA-3'.

## **Results**

### *Protein components in the dragline silk*

To analyze the components of dragline silk, SDS-PAGE was used to separate the proteins and found that, in addition to the two major proteins (MaSp1 and MaSp2), some other types of protein components were present, particularly a small protein of approximately 40 kDa (Figure 2).

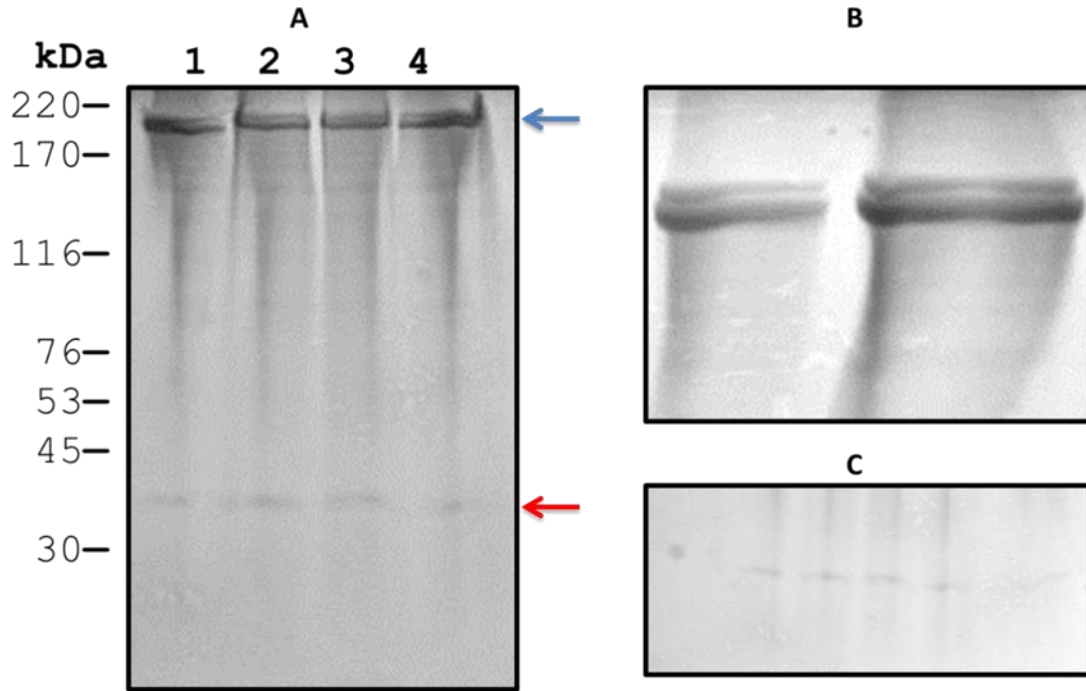


Figure 2. Proteins components separated by SDS-PAGE. The blue arrow indicates the two major components (MaSp1 and MaSp2), which is amplified in panel B. The red arrow indicates the low molecular weight component (approximately 40 kDa), which is shown in panel C. Lanes 1-4 are the same sample.

### *Identification of the MaSp gene*

A total of 1344 clones were selected for partial sequencing with 3' and 5' terminal primers. The sequencing results comprised DataSet I. To identify the target *MaSp* gene, all of the reported amino acid sequences for spider silk protein genes were downloaded from DDBJ and NCBI using the following keywords: spidroin, major ampullate spidroin (MaSp), minor ampullate spidroin (MiSp), flagelliform silk protein (Flag), *Araneus diadematus* fibroin (ADF), cylindrical silk protein (CySp), dragline spider silk protein, Aciniform, Tubuliform, Aggregate, and Pyriform. The results were defined as DataSet II. DataSet I was then compared with DataSet II using LocalBlast.exe (C. G. Zhang, 1999) with its Blast X subprogram. The results indicated that one clone, pWB4-1 (clone number), matched well with the previously

reported MaSp1. pWB4-1 was completely sequenced (Figure 3), and the inserted fragment sequence (1233 bp) was analyzed using Blast X and Blast P (Figure 4), revealing that the 5' terminal region appeared to be lacking some upstream sequence (signal peptide region).

```

1 AGCGCAAGTA TGGCAGAAAG CTTTATGACT TACTTTTCTG AAGCTTTAGG ACAATCAGGA
61 GCATTTACGA AAGAACAGAT AGATGATATT GATACCATG CCTCATCTAT AAAACTGGGA
121 GTAGATAAAA TGGAAAGAAG CGGGAAGACT TCAGGAAGCA AACTGCAAGC TATGAATATA
181 GCTTTCGCTT CAGCAGTGGC AGAAATAGCC ACGACAGAA GAGGAGAACA GACCGCTGAA
241 GTCAAACTA AAGCTGTTGC CGACGCTTTG GCATTCGCAT TCTTTCAAAC GAAAGGAGCA
301 GTGAACATAA ATTTTCATAA TGAAATCAAA AATTTAATTT CTATGTTTGC TCAGACAAAT
361 ACCATTTCTT CGTCTTTAGA TTCAGCCGGT GCATCAGCTG CACAACTGT TAGAATAAAT
421 GGATATGGAC AAATAGAGGC TGAAGCAGCT GCCGCAGCAG CCGCTGGATC TGGTGTAGCA
481 AGACGAGGAG GTTATGGTCA AGACGAACT GGAGCAAGGA ATTCAGCTGC AATTGCGACC
541 GCAGCAGCAG CCGCTGGAGC TGGCGGAGCA GGAAGAATAG GTTATGGTCA AAGAGGAGCT
601 GGAACAGGTG ATTCCCCTGC AGCTACAGTC GCAACAGTCG CTGGAGTTGG CGGAGCAGGA
661 AGAGGAGGTT ATGATCAAGG AAGATCAGGA GCCACAGTGG CAGCCGCTAC AGCAGGACGA
721 GGAGGTTATT ATCAAGGAGG AGCAGGATTA GGTGATGCCG CAGCAGCCAC TGGAGCCGGC
781 AGAGCAGAAA GAGGAGGTTA TGGTGAAGGA GGAGCTGCAG CAGGTAATGC CGCAACAGCA
841 GCAGCCGGAG AACAAGGCGG ATATGGTGGT CAAGGACTAA GTGGTTCTTA CGGTGGACAA
901 CAAGGAGCAG CAGCCTTGGC TTCGGCTGCA GCTACACGCT TGTCTTCTCC TAATGCTGCT
961 TCAAGAATTT CAAATTCTGT ATCATATTTA TTATCAAACG GTGGACCATC TAGTTCTATA
1021 GCTCTTTCAA ATACCATTAA TAATGCTGTT TCTCAAATTA GCGCAAGTAA TCCTGGTCTT
1081 TCGACCTGTG ATATTCTTGT TCAAGCTTTA CTTGAACTTA TTTCCGCCCT GATATACATT
1141 CTTGGATCAT CTACTATTGG TCAAGTTAAC TCCAGTTTAG CTGGCCAGTC CGCTTCAATG
1201 GTCGGACAAT CTGTATACCG AGCTTTCCTT TAA

```

Figure 3. The sequence of pWB4-1.

```

Query: 1 MAESFMTYFSEALGQSGAFTKEQIDIDTIASSIKLGVDKMERSGKTSKSLQAMNIAFA 60
      +A++F+ F      G++GAFT +Q+DD+ TI  ++K  +DKM RS K+S SKLQA+N+AFA
Sbjct: 34 LADAFINAFLEAGRTGAFTADQLDDMSTIGDTLKTAMDKMARSNKSSQSKLQALNMAFA 93
      SAVAEIATTEGGEQTAEVKTAVADALAFQTKGAVNINFINEIKNLISMFAQ 115
      S++AEIA E G +   KT A+AD+L AF+QT GAVN+ F+NEI++LISMFAQ
      SSMAEIAAVEQGGLSVAEKTNAIADSLNSAFYQTTGAVNVQFVNEIRSLISMFAQ 148

```

Figure 4. Partial results of blastp. Sequences producing significant alignments.  
Query : pWB4-1; Sbjct: Nephila clavipes MaSp1B.

### *Acquisition of the full-length MaSp1s*

Inverse PCR (iPCR) is a useful method to amplify DNA sequences that flank a region of known sequence (63). Because the above analysis indicated that pWB4-1 (1233 bp) does not have an intact 5' terminus, iPCR was used to analyze the intact 5'



terminal sequence of *MaSIs*. The result showed that the full-length gene (*MaSp1s*) is actually 87 bp longer (1320 bp) than initially determined.

To confirm the existence of the full-length *MaSp1s* gene (1320 bp) in the genome and the presence of the transcript in the total RNA, standard PCR and RT-PCR were performed using primers that included the start codon ATG to the stop codon TAA(Figure 5), respectively. The sequencing results were 100% identical (data not shown) to the sequence that comprised pWB4-1 and the iPCR results(Figure 6).

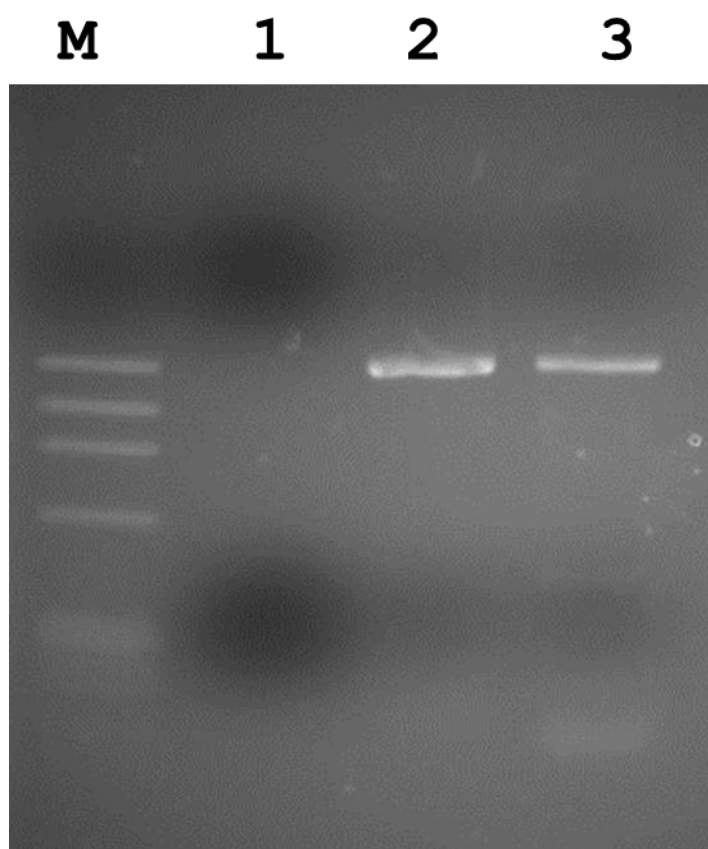


Figure 5. The results of Genomic DNA PCR and RT-PCR. The standard PCR and RT-PCR were performed using forward primer, 5' ATGACTTGGACAACCTCGACTTGCCTTAT 3', and reverse primer, 5' TTAAAAGAAAGCTCGGTATACAGATTGT 3'. Lane M,  $\phi$ X174 Hae III digest; Lane 1, PCR negative control; Lane 2 Genomic DNA PCR product; Lane 3, RT-PCR product.

1	CAC	GAT	GCG	TCA	AAA	AGC	ATT	CGG	TCG	AAA	AGA	ACC	AAT	aGT	aTG	ATG	ACT	TGG	ACA	ACT
	H	D	A	S	K	S	I	R	S	K	R	T	N	S	<b>M</b>	<b>M</b>	T	W	T	T
61	CGA	CTT	GCC	TTA	TCA	TTT	CTC	GTA	GTG	ATC	TGC	TCT	CAG	AGC	CTA	TTT	GCT	CTA	TGT	CAA
	R	L	A	L	S	F	L	V	V	I	C	S	Q	S	L	F	A	L	<b>C</b>	<b>Q</b>
121	TCA	CCA	TGG	CAA	AGC	GCA	AGT	ATG	GCA	GAA	AGC	TTT	ATG	ACT	TAC	TTT	TCT	GAA	GCT	TTA
	S	P	W	Q	S	A	S	M	A	E	S	F	M	T	Y	F	S	E	A	L
181	GGA	CAA	TCA	GGA	GCA	TTT	ACG	AAA	GAA	CAG	ATA	GAT	GAT	ATT	GAT	ACC	ATT	GCC	TCA	TCT
	G	Q	S	G	A	F	T	K	E	Q	I	D	D	I	D	T	I	A	S	S
241	ATA	AAA	CTG	GGA	GTA	GAT	aAA	ATG	GAA	AGA	AGC	GGG	AAG	ACT	TCA	GGA	AGC	AAA	CTG	CAA
	I	K	L	G	V	D	K	M	E	R	S	G	K	T	S	G	S	K	L	Q
301	GCT	ATG	AAT	ATA	GCT	TTC	GCT	TCA	GCA	GTG	GCA	GAA	ATA	GCC	ACG	ACA	GAA	GGA	GGA	GAA
	A	M	N	I	A	F	A	S	A	V	A	E	I	A	T	T	E	G	G	E
361	CAG	ACC	GCT	GAA	GTC	AAA	ACT	AAA	GCT	GTT	GCC	GAC	GCT	TTG	GCA	TTC	GCA	TTC	TTT	CAA
	Q	T	A	E	V	K	T	K	A	V	A	D	A	L	A	F	A	F	F	Q
421	ACG	AAA	GGA	GCA	GTG	AAC	ATA	AAT	TTC	ATA	AAT	GAA	ATC	AAA	AAT	TTA	ATT	TCT	ATG	TTT
	T	K	G	A	V	N	I	N	F	I	N	E	I	K	N	L	I	A	S	M
481	GCT	CAG	ACA	AAT	ACC	ATT	TCT	TCG	TCT	TTA	GAT	TCA	GCC	GGT	GCA	TCA	GCT	GCA	CAA	ACT
	A	Q	T	N	T	I	S	S	S	L	D	S	A	G	A	S	A	A	Q	T
541	GTT	AGA	ATA	AAT	GGA	TAT	GGA	CAA	ATA	GAG	GCT	GAA	GCA	GCT	GCC	GCA	GCA	GCC	GCT	GGA
	V	R	I	N	G	Y	G	Q	I	E	A	E	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>
601	TCT	GGT	GTA	GCA	AGA	CGA	GGA	GGT	TAT	GGT	CAA	GAC	GAA	ACT	GGA	GCA	AGG	AAT	TCA	GCT
	S	G	V	A	R	R	G	G	Y	G	Q	D	E	T	G	A	R	N	S	A
661	GCA	ATT	CCG	ACC	GCA	GCA	GCA	GCC	GCT	GGA	GCT	GGC	GGA	GCA	GGA	AGA	ATA	GGT	TAT	GGT
	A	I	A	T	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	G	A	<b>G</b>	<b>G</b>	<b>A</b>	G	R	I	G	Y	G
721	CAA	AGA	GGA	GCT	GGA	ACA	GGT	GAT	TCC	CCT	GCA	GCT	ACA	GTC	GCA	ACA	GTC	GCT	GGA	GTT
	Q	R	G	A	<b>G</b>	<b>T</b>	G	D	S	P	A	A	T	V	A	T	V	A	G	V
781	GGC	GGA	GCA	GGA	AGA	GGA	GGT	TAT	GAT	CAA	GGA	AGA	TCA	GGA	GCC	ACA	GTG	GCA	GCC	GCT
	G	G	A	G	R	<b>G</b>	<b>G</b>	<b>Y</b>	D	Q	G	R	S	G	A	T	V	A	A	A
841	ACA	GCA	GGA	CGA	GGA	GGT	TAT	TAT	CAA	GGA	GGA	GCA	GGA	TTA	GGT	GAT	GCC	GCA	GCA	GCC
	T	A	G	R	G	G	Y	Y	Q	G	G	A	G	L	G	D	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>
901	ACT	GGA	GCC	GGC	AGA	GCA	GAA	AGA	GGA	GGT	TAT	GGT	GAA	GGA	GGA	GCT	GCA	GCA	GGT	AAT
	T	G	A	G	R	A	E	R	G	G	Y	G	E	G	G	A	A	A	G	N
961	GCC	GCA	ACA	GCA	GCA	GCC	GGA	GAA	CAA	GGC	GGA	TAT	GGT	GGT	CAA	GGA	CTA	AGT	GGT	TCT
	A	A	T	A	A	A	G	E	Q	G	G	Y	<b>G</b>	<b>G</b>	<b>Q</b>	G	L	S	G	S
1021	TAC	GGT	GGA	CAA	CAA	GGA	GCA	GCA	GCC	TTG	GCT	TCG	GCT	GCA	GCT	ACA	CGC	TTG	TCT	TCT
	Y	G	G	Q	Q	G	A	A	A	L	A	S	A	A	A	T	<b>R</b>	<b>L</b>	<b>S</b>	<b>S</b>
1081	CCT	AAT	GCT	GCT	TCA	AGA	ATT	TCA	AAT	TCT	GTA	TCA	TAT	TTA	TTA	TCA	AAC	GGT	GGA	CCA
	P	N	A	A	S	R	I	S	N	S	V	S	Y	L	L	S	N	G	G	P
1141	TCT	AGT	TCT	ATA	GCT	CTT	TCA	AAT	ACC	ATT	AAT	AAT	GCT	GTT	TCT	CAA	ATT	AGC	GCA	AGT
	S	S	S	I	A	L	S	N	T	I	N	N	A	V	S	Q	I	S	A	S
1201	AAT	CCT	GGT	CTT	TCG	ACC	TGT	GAT	ATT	CTT	GTT	CAA	GCT	TTA	CTT	GAA	CTT	ATT	TCC	GCC
	N	P	G	L	S	T	C	D	I	L	V	Q	A	L	L	E	L	I	S	A
1261	CTG	ATA	TAC	ATT	CTT	CGA	TCA	GCT	ACT	ATT	GGT	CAA	GTT	AAC	TCC	AGC	TTA	GCT	GGC	CAG
	L	I	Y	I	L	R	S	A	T	I	G	Q	V	N	S	S	L	A	G	Q
1321	TCC	GCT	TCA	ATG	GTC	GGA	CAA	TCT	GTA	TAC	CGA	GCT	TTC	TTT	TAA					
	S	A	S	M	V	G	Q	S	V	Y	R	A	F	F	*					

Figure 6. Deduced translation of the MaSp1s sequence. The red frame represents the N-terminus, the blue frame denotes the C-terminus, and the region between two termini is the so-called repetitive region. The Met residue representing the first amino acid of the transcript is shaded with a red bold format. Every adenine at the -3 position (with respect to the potential start codon) is in lower case. The characteristic motifs of MaSp1s are highlighted in red.

### *The organization of the new MaSp1s gene*

*MaSp1s* deviates from the rule that spider silk genes are always very large (~10 kb). The full-length transcript is only 1320 bp, which encodes 439 amino acids. According to convention, MaSp1s was divided into 3 sections: the N-terminus, repetitive region and C-terminus. And the boundaries were defined using the Blast P results (Figure 7). The amino acids sequence deduced from the DNA sequence is shown in(Figure 6).

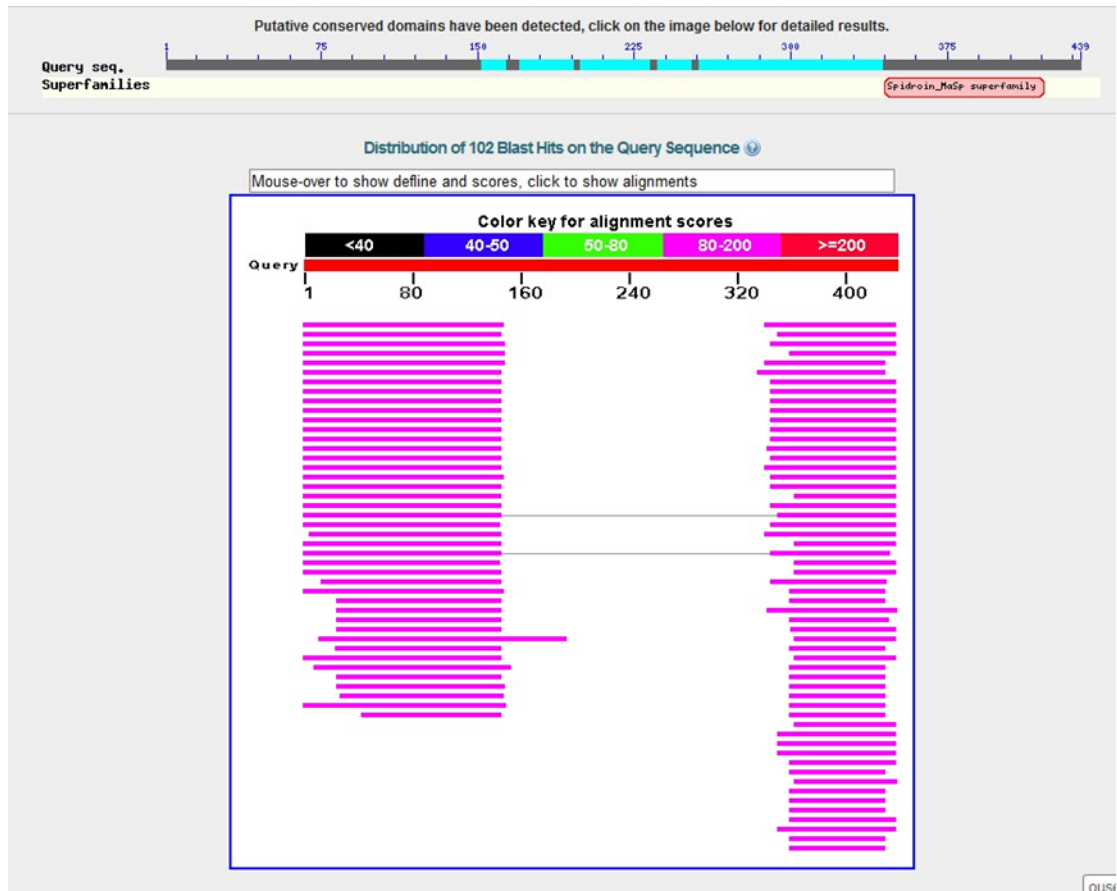


Figure 7. The Blast P result for MaSp1s. The boundaries of the MaSp1s N-terminal region, repetitive region, and C-terminal region were predicted by the Blast P results (NCBI).

#### *Evolutionary relationships of the N-terminal region.*

According to previous reports, the N-terminal regions of spidroins are conserved (64-66), thus the N-terminus of MaSp1s was aligned with the N-termini from other

spidroins(Figure 8). The analysis showed that the N-terminal sequence of MaSp1s is highly homologous with that of Nc.MaSp1a, sharing 56% amino acid identity (83/149) and 74% amino acid similarity (111/149).



Figure 8. Alignment of published N-terminal amino acid sequences and MaSp1s using GENETYX. The amino acids shared by 100% of the proteins are highlighted by boxes. The gaps are represented by dashes. Abbreviations: Cm.MaSp1s, *Cyrtophora moluccensis* MaSp1s (this study); Nc.MaSp1a, *Nephila clavipes* MaSp1 (EU599238); Nc.MaSp1b, *Nephila clavipes* MaSp1 (EU599239); Lh.MaSp1, *Latrodectus hesperus* MaSp1 (EF595246); Lh.MaSp2, *Latrodectus hesperus* MaSp2 (EF595245); Lh.TuSp1, *Latrodectus hesperus* TuSp1 (DQ379383); Ab.TuSp1, *Argiope bruennichi* TuSp1 (AB242144); Lh.MiSp, *Latrodectus hesperus* MiSp (HM752570); Bc.fibroin1, *Bothriocyrtum californicum* fibroin1 (HM752562); and Ni.Flag, *Nephila inaurata* madagascariensis Flag (AF218623S1).

The amino acid sequence motif “TTGXXN” identified by Rising et al also does not appear at position +126 to +131 in MaSp1s (64, 65). Of the 149 aligned residues in the N-terminal region, three residues, the start codon (position +1), aspartic acid

(position +58) and glycine (position +128), as reported by Hayashi et al to be conserved across the spidroin N-termini, also are found in MaSp1s (64).

### *Signal peptide analysis*

In addition to the conservation of the N-terminus, the 5' terminal sequence of the mRNA is also likely to encode a signal peptide, a key element of silk proteins, such as spider silk and silkworm silk, which must be transported across the endoplasmic reticulum and secreted (28). Therefore, the N-terminal sequence of MaSp1s was analyzed using signalP4.0 (eukaryotic option) (67) to determine whether a signal peptide was present. The analysis suggested that there was an apparent signal sequence in MaSp1s(Figure 9), and the putative cleavage site between amino acids 24 and 25 corresponds to the mature protein starting at position 25 (glutamine, Q), which is highlight by the yellow character in(Figure 6).

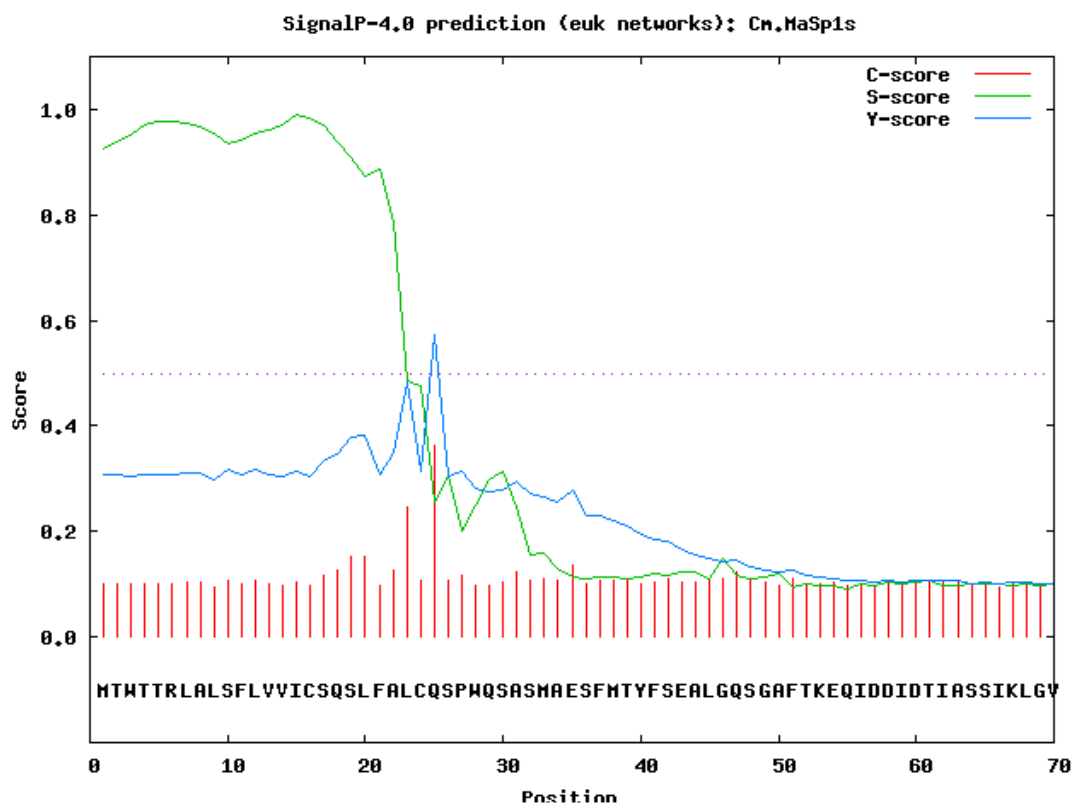


Figure 9. Signal peptide analysis of MaSp1s.

### *So-called Repetitive region*

The entire gene is only 1320 bp, and analyses indicate that MaSp1s does not have obvious repetitive regions, only including 7 similar repeat units (no extreme homogenization). However, the motif analysis suggested that MaSp1s included all of the MaSp1 characteristic motifs, such as GGX (X =A, Q, or Y), GX (X =Q, A, or R), and poly-A, some of which are highlighted with a red color in(Figure 6). However, the sequence did not contain GPG motifs, which is a characteristic motif of MaSp2. Thus, this region was termed as the ‘so-called repetitive region’.

### *C-terminal region*

The C-terminal region of MaSp1s was aligned with the C-termini from other spidroins(Figure 10). The most homologous region of MaSp1s to the MaSp1 proteins from other spider species is the C-terminus. The alignment results demonstrated that MaSp1s is highly homologous with Aa.MaSp1 (identity = 71/97, 73%, similarity = 88/97, 91%), and an NJ tree was obtained. These sequences can be separated into groups, depending on the type of silk, with the C-terminus of MaSp1s clustering robustly with the other MaSp1(Figure 11).

Bc.fibroin1	1	-----LTSEAAKERISIIASLLSAKSSQDFNALLLSNILEPSLISKISQKASG	48
Aa.MaSp1	1	-----RLSSPGAASRVSSAVTSLVSSGGPTNSAALSNTISDVVSQISASNPG	47
Lh.MiSp	1	-----GGVLNSAALPSVSNMMSQVSSASSPG	26
Ab.TuSp1	1	ALAPIVSSSGLGSSAATAKVSSLANSFASAISSGGSLSVPTFLNLLSSVGAQVSSSSS	59
Lh.TuSp1	1	--APVLS PAGLASTAATSINDIAQSLSTLSS--GSLAPDNVLPGLIQLSSITQSGNPD	57
Lh.MaSp1	1	-----RISSHASALLS--NG--PTNPAISI SNVISNAVSQISSSNPG	37
Lh.MaSp2	1	-----PTTHARISSHASTLLS--SG--PTNAAALSNI SNVISNAVSQVSSASNPG	42
Nc.MaSp1a	1	-----RLSSPEASSRVSSAVSNLVS--SG--PTNSAALSSTISNVVSQIGASNPG	46
Nc.MaSp1b	1	-GASAAASAAASRLSSPQASSRVSSAVSNLVA--SG--PTNSAALSSTISNVVSQIGASNPG	56
Ni.Flag	1	-GPGSPGGAYGPGSSGGSYYPSSRVFDMVNGIMSAMQGGGFNYQMFNMLSQYSSGGSGSCN	59
Cm.MaSp1s	1	-----RLSSPNAASRIENSVSYLELS--N--SGPSSSIALSNTINNAVSQISASNPG	47
Bc.fibroin1	49	LSPT EMVTEALLEVLAGCMEILSSFNVGAQSISSTSSSNALVQSI SNQFSGLNAAA	105
Aa.MaSp1	48	LSGC DVLVQALLEIVSALVHILSSANIGQVNSSSASQASLVGQSVYQALS-----	98
Lh.MiSp	27	MSSEEVVIQALLELVSSLIHILSSANIGQVDFNSVNTAAVVGQSLGAAALG-----	77
Ab.TuSp1	60	ESSLEVINEVLLAIAALLQVINGGSIITSVLYVPNAQQDLVNALSG-----	107
Lh.TuSp1	58	LDPA GVLI ESLL EYTSALLALQNAQITTYDAATLPAFNTALVNYYLVPLV-----	107
Lh.MaSp1	38	ASAC DVLVQALLELV TALLTIIGSSNIGSVNYDSGGQYAVVITQSVQNAFA-----	88
Lh.MaSp2	43	SSSC DVLVQALLEIITALLISILDSSSVSQVNYGSSGGQYQIVGQSMQAMG-----	93
Nc.MaSp1a	47	ESGC DVLVQALLEVV SALIHILSSSISGQVNYGSAAGQATQIVGQSIYQALG-----	97
Nc.MaSp1b	57	ESGC DVLVQALLEVV SALIHILSSSISGQVNYGSAAGQATQIVGQSVYQALG-----	107
Ni.Flag	60	FNNVNVLM DALLAALHCLSNHSSSFAPSETPAAMSAYSNSVGRMFAY-----	107
Cm.MaSp1s	48	ESTCDILVQALLELISALIIYILRSATIGQVNSSLASQASMVGQSVYQAF-----	98



Figure 10. Alignment of the published C-terminal amino acid sequences and MaSp1s.

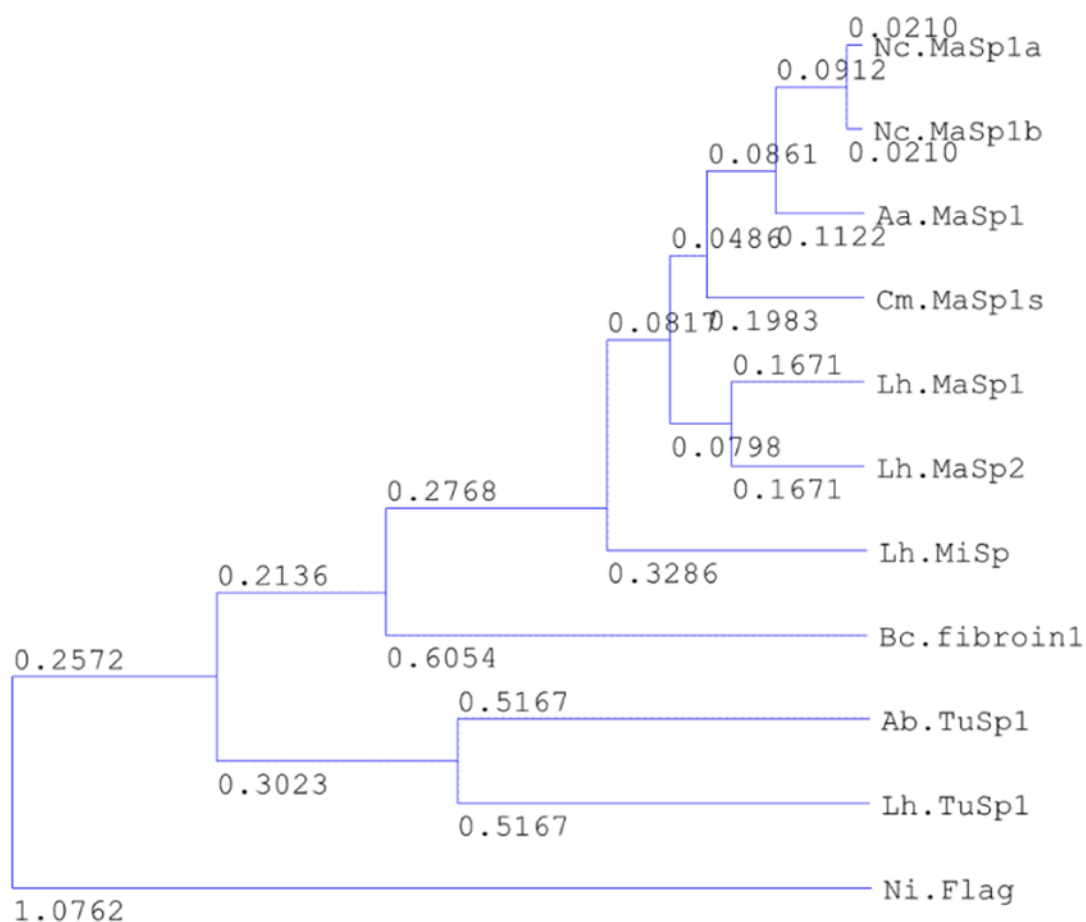


Figure 11. Comparison of the C-termini of spider silk proteins. Schematic diagram of an evolutionary tree of the C-terminal-encoding sequences using GENETYX. Abbreviations: Nc.MaSp1a, Nephila clavipes MaSp1a (AAT75312.1); Nc.MaSp1b, Nephila clavipes MaSp1b (AAT75311.1); Aa.MaSp1, Argiope amoena MaSp1 (AAP88232.1); Cm.MaSp1s, Cyrtophora moluccensis MaSp1s (this study); Lh.MaSp1, Latrodectus hesperus MaSp1 (ABR68856.1); Lh.MaSp2, Latrodectus hesperus MaSp2 (ABD66603.1); Lh.MiSp, Latrodectus hesperus MiSp (ADM14322.1); Bc.fibroin1, Bothriocyrtum californicum fibroin 1 (ABW80565.1); Ab.TuSp1, Argiope bruennichi TuSp1 (BAE86855.1); Lh.TuSp1, Latrodectus hesperus TuSp1 (AAY28931.1); and Ni.Flag, Nephila clavipes flag (AAC38847.1).

### *Amino acid content and codon usage*

The predicted amino acid content and codon usage of MaSp1s were compared with those of the black widow spider MaSp1 (12) (Table 1). The results showed that, unlike the black widow spider MaSp1, there MaSp1s has a higher alanine than glycine content, even though the codon usage is very similar.

### *Hydrophobicity analysis*

Using the method of Kyte and Doolittle, the hydrophobicity of MaSp1s was predicted, which is slightly hydrophobic, as averaged across all of the residues (average =0.103 on the Kyte-Doolittle score). The N- and C-terminal domains are hydrophobic, whereas the C-terminal region displays a higher hydrophilicity (N-terminal max = 3.66, total hydrophobicity score=27.9; C-terminal max = 2.93, total hydrophobicity score= 40.8). The poly-(A) motif ( $n > 4$ ) regions show hydrophobicity, whereas the glycine-rich (GX, GGX) regions present hydrophilicity (12), regions that alternate throughout the so-called repetitive portion. Because there are more glycine-rich regions than poly-A regions, the so-called repetitive portion demonstrates hydrophilicity (average = -0.12, total hydrophobicity score= -23.09). The most hydrophobic region of MaSp1s is found in the signal peptide zone(Figure 12).

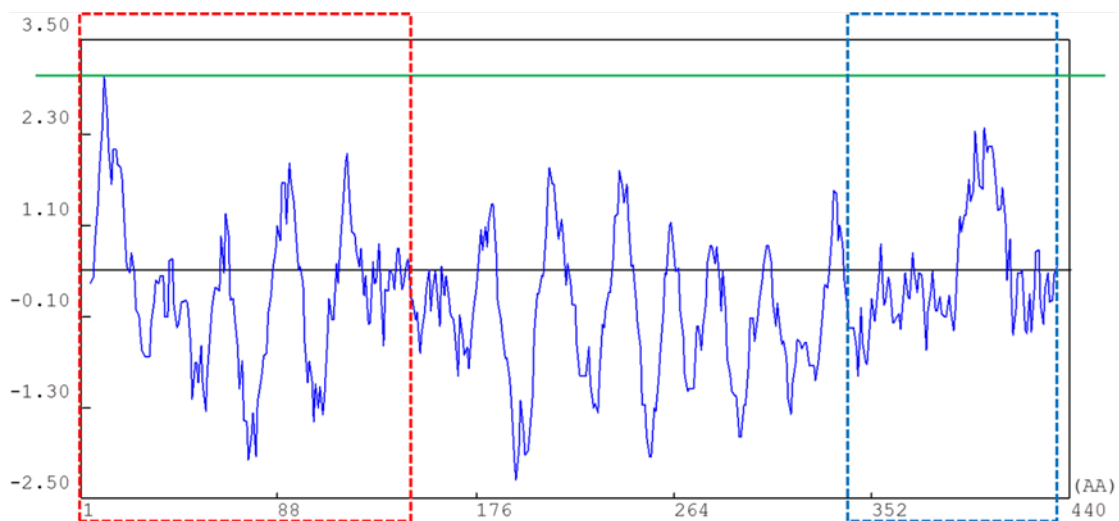


Figure 12. Kyte and Doolittle hydrophilicity plots for MaSp1s. The negative values



indicate hydrophilicity. The region framed in red represents the hydrophobicity profile of the N-terminus, and the region framed in blue represents the hydrophobicity profile of the C-terminus. The green line shows the highest hydrophilicity value.

## Discussion

### *Screening of a cDNA Library*

To obtain as much information from the library, the common screening methods (PCR or hybridization) were not chosen, but used direct sequencing instead. Furthermore, the screening method mentioned in our previous report was also not been employed (54) because a different cDNA library construction method was used in the present study, as the previous method was not suitable for this library screening. According to previous reports, the C-termini of spidroins share considerable conservation and also tend to group by gland type rather than the spider taxon at the protein level (12, 28). Thus, the sequencing reaction was started from the 3'-terminus of the library clones rather than from the 5'-terminus, which has been used by many other researchers. My results showed that the new screening method based on this approach is feasible and can yield at least two pieces of useful information. The first is which genes among the sequenced clones are spider silk genes; secondly, this approach reveals the type of spider silk gene (MaSp, MiSp, Flag, or TuSp). To decrease the cost of the library sequencing, 1344 (96X14) recombinant colonies of 2880 (96X30) with >1 kb insert fragments were selected for partial sequencing by the visualization of the plasmid DNA using the modified Gel Electrophoresis of Bacteria Solution method (68). Unlike the method of Beuken et al (69), the modified Gel Electrophoresis of Bacteria Solution does not use a phenol/chloroform mixture and is much faster and more stable, with results that are more clear(Figure 13).

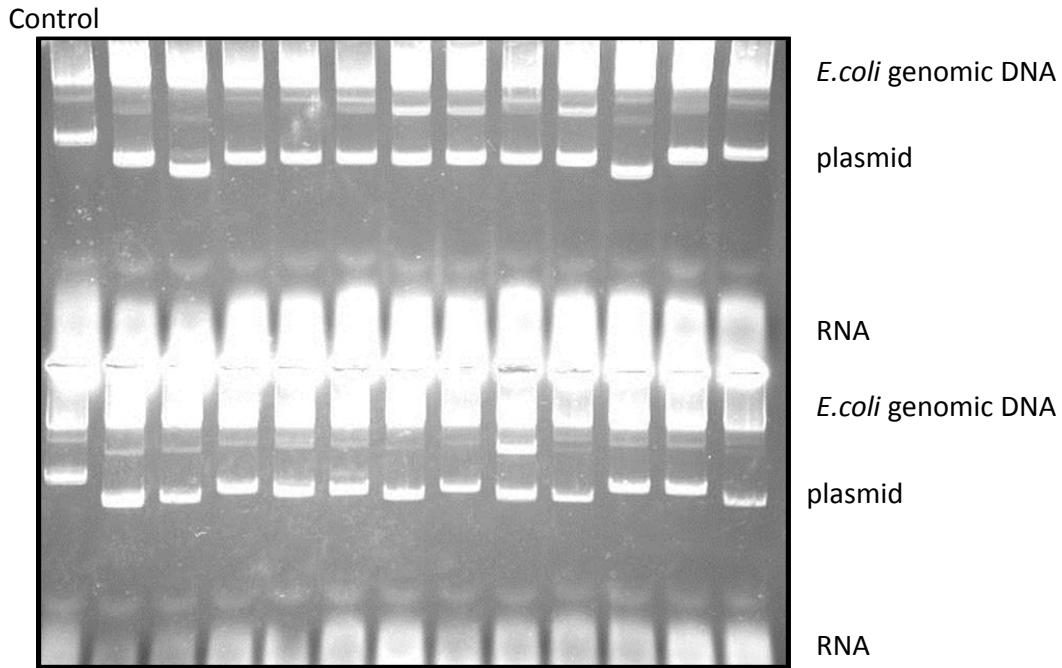


Figure 13. A characteristic result of the Modified Gel Electrophoresis of Bacteria Solution method. The electrophoresis bands are very clear.

### *Signal peptide analysis*

A signal peptide is a key element for such secretory proteins as spidroins. The cDNA library data initially demonstrated that the sequence (pWB4-1) did not include a signal peptide, as predicted by SignalP 4.0 (Figure 14). Similar results have been reported in other spider silk genes (55, 64). It has been reported that the N-terminal Cys residues have a role in linkages between AbCySp1 and AbCySp2, which is similar to the linkage between the H-Chain and L-Chain via one of the Cys residues present in the C-termini of lepidopteran silk. The AbCySp1-AbCySp2 complex, which would possess a signal peptide, would then be transported across the endoplasmic reticulum and secreted to the Golgi (55). However, the Cys residues were not identified in the N-terminal region of pWB4-1. Furthermore, although the

model that the AbCySp1-AbCySp2 complex would possess a signal peptide is an interesting hypothesis, it has not been verified. Additionally, when the similarity of the N-terminus was analyzed, the result shown that the matching region between pWB4-1 and the previously published MaSp1 always began at a position approximately +35 AA with respect to the first Met of the previously published MaSp proteins with obvious signal peptides(Figure 4). A further analysis demonstrated that the amino sequence before the +35 is the signal peptide (data not shown).

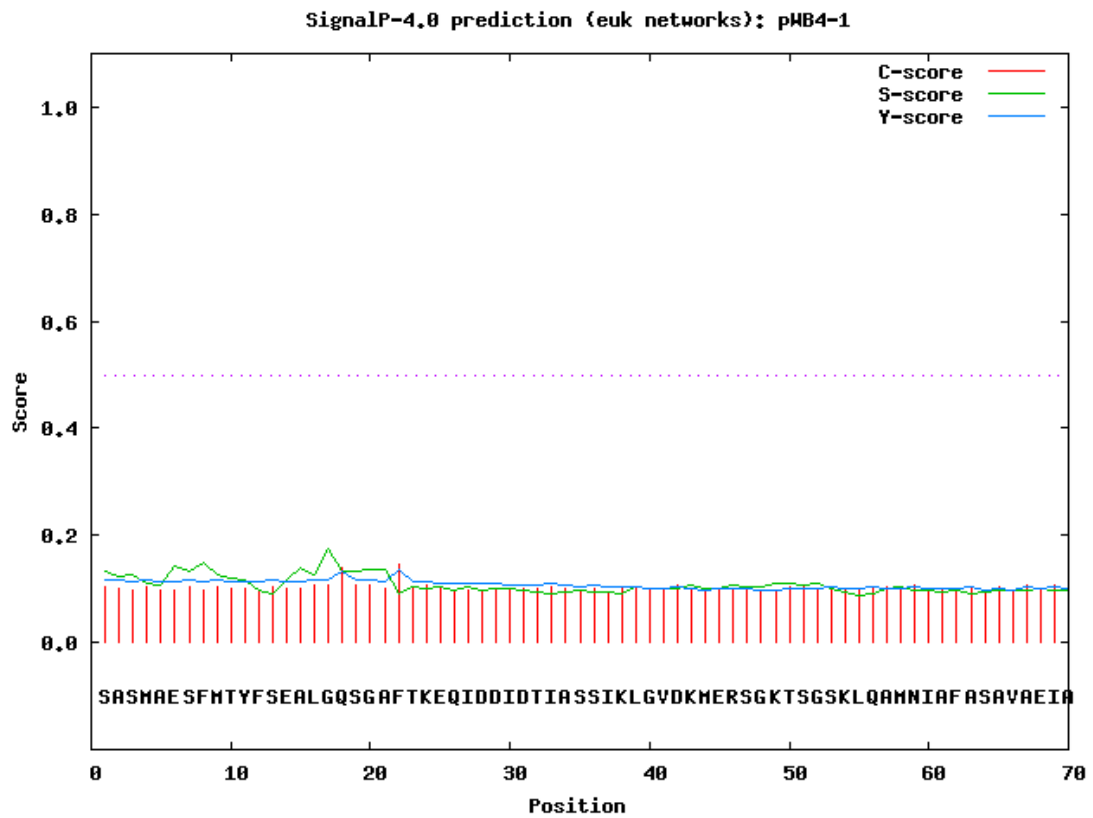


Figure 14. Signal peptide analysis of pWB4-1

Because the complete N-terminal sequence is approximately 150 amino acids (69) and the N-terminal similarity between pWB4-1 and the other MaSp genes began at approximately +35, pWB4-1 may lack some upstream sequences. In other words, unlike most of the other spider silk genes obtained from the cDNA library, which

always have the C-terminal sequences and a small part of the repetitive region sequences, the 5' terminal sequence of pWB4-1 is located in the N-terminal region and not in the repetitive region. For this reason, it was possible to use the iPCR method to obtain the full N-terminal sequence (if the 5' terminal sequence was located in the repetitive region, it would be very difficult to design primers for iPCR, and the PCR products would be a smear by gel electrophoresis). The iPCR method is also much easier than 5' RACE, another reason why the iPCR method was chosen. By using the iPCR method, an additional 87 bp upstream of *MaSp1s* was obtained; these 29 amino acids were predicted to include the signal peptide by signalP4.0. Thus, the full-length gene: an intact non-repetitive N-terminus, a C-terminus and a repetitive region was obtained.

#### *Start codon*

There are seven Met codons at positions 15, 16, 48, 53, 88, 102 and 159 of the N-terminal region, which suggests the possibility of more than one transcription isoform. According to the rule that an adenine at position -3, with respect to the ATG, in mammalian and *Drosophila* sequences enhances the likelihood of the translation start site (66), the Met at positions 15, 16 and 88 exhibit this characteristic (Figure 6). Combined with the result of Blast P (Figure 7), the second Met was chosen as the likely start codon.

#### *MaSp1s: A new type of MaSp gene*

The full-length sequence of *MaSp1s* is only 1320 bp. However, this short isoform is not the same as the short isoforms (mentioned by Motriuk), an isoform resulting from the N-terminus starting at a different position (66). *MaSp1s* is also not one of the alternative splicing variants, which was validated by genomic DNA PCR and RT-PCR (Figure 5). Similar short silk genes (not identical) also exist in other types of

spiders and insects (unpublished data). The reason why MaSp1s is so short that it does not contain as many repetitive units as the other MaSp1 proteins (up to 100 times), which determine the strength and elasticity of dragline silk (70). In other words, MaSp1s would not play a key role in determining the silk's mechanical properties. The SDS-PAGE data are consistent with a former report that the dragline is predominantly composed of two proteins (MaSp1 and MaSp2) and some other proteins (20). The intensity of the small proteins is much weaker than that of MaSp1 and MaSp2, indicating a lower abundance of the small protein. This is circumstantial evidence that MaSp1s would not produce a dominant effect in the determination of the strength and elasticity of silk.

The function of MaSp1s is presently unclear. However, according to the analysis mentioned above, the N and C termini of MaSp1s are highly homologous with those of other MaSp proteins. It is reported that the non-repetitive termini contribute to storage and assembly of spider silk proteins(70). This rationale supports the hypothesis that MaSp1s would be partially responsible for the storage and assembly of spider silk proteins, particularly with regard to the proteins assembling together with MaSp1 and MaSp2. Perhaps MaSp1s plays a role similar to p25 in the silkworm (71). Further research is needed to elucidate the function of MaSp1s.

## Paragraph 2: Reanalyzing the *AbCySp2* gene of Wasp Spider *Argiope bruennichi*.

Due to their unique mechanical properties, spider silks (proteinaceous polymer) have received interest from scientists. Most spider silks are composed of enormous proteins that are encoded by large genes consisting of approximately 10,000 base pairs (bp). Because spider silk genes are usually very large, it is very difficult to retrieve full-length cDNA sequence for these genes.

The development of gene molecular technology has accelerated the identification of novel spider silk genes. Some of the submitted sequences correspond to mRNA sequences, while others correspond to genomic DNA sequences. However, most of these sequences are partial coding sequences (partial cds). To date, five full-length spider silk genes: *Latrodectus hesperus* major ampullate spidroin 1 (*LhMaSp1*, EF595246), *Latrodectus hesperus* major ampullate spidroin 2 (*LhMaSp2*, EF595245), *Cyrtophora moluccensis* major ampullate spidroin 1 short gene (*Cm. MaSp1s*, KF032719), *Argiope bruennichi* cylindrical silk protein 1 (*AbCySp1*, AB242144) and cylindrical silk protein 2 (*AbCySp2*, AB242145) (12, 55, 72) have been identified. To meet the needs of the developing fiber industry, more and more spider silk gene will be applied in the industry. It is important to know the correct information of the genes of interest, such as *CySp* which is famous for its high strain character.

However, *AbCySp2* was reported by our lab as non-signal peptide protein (Zhao, 2006). Zhao explain that the N-termini of *CySp1* and *CySp2* contain several Cys residues, it is possible that N-terminal Cys residues have a role on linkage between *CySp1* and *CySp2*, similar to the function of C-terminal Cys residues seen in other

silks. The CySp1-CySp2 complex, which would possess a signal peptide, would then be transported across the endoplasmic reticulum and secreted to the Golgi(55). But, this explanation was always suspected. Because, secretory signals are key elements for silk proteins, such as spider and silkworm silks. So, we reanalyzed whether the *AbCySp2* gene have signal peptide or not.

## **MATERIALS AND METHODS**

### *Spider Care and Genome DNA extraction*

Mature female *Argiope bruennichi* spiders were collected in Ueda city of Nagano prefecture, Japan during October. Extraction of genome DNA from 25 mg brain was performed using the DNeasy Blood & Tissue Kit (*QIAGEN*).

### *PCR reaction and gel extraction*

PCR was performed on ABI Thermal Cycler 2700 with following cycling parameters: initial denaturation at 94 °C for 5 min, followed by 28 cycles (denaturation: 94 °C for 30 s, annealing: 57 °C for 30 s, extension: 72 °C for 30 s) and a final extension at 72 °C for 10 min. PCR was conducted using the following primer pair: Forward Primer: 5' GGATGCCATTGCCCAGGTGATAC 3' and Reverse Primer: 5' AAACGCCCTGACCGGCTAAGAAT 3'. After electrophoresis, the DNA band of interest was excised from agarose using a clean razor blade and purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega*).

### *DNA Sequencing and analysis*

The purified DNA fragment of interest was cloned into the T-Vector pMD19 (*TaKaRa*) using DNA Ligation Kit Ver.2.1 (*TaKaRa*). The resulting vector was sequenced from both directions using universal sequencing primers, M13 primer RV and M13 primer M4 (*TaKaRa*), and an *ABI* PRISM® 3100 Genetic Analyzer. Computer analyses of DNA and amino acid sequences were conducted using Genetyx ver.10. The new N-terminal sequence for AbCySp2 (AbCySp2N) was analyzed using signalP4.0 (eukaryotic option) (67). Comparison between N-terminal sequences for AbCySp2 (AbCySp2O) as reported by Zhao and AbCySp2N were conducted using Blast P and Blast N on NCBI. An alignment of the transcribed DNA sequence (*AbCySp2O*) with my genomic sequence (*AbCySp2N*) was generated online using SIM4 (73).

## RESULTS

### *Acquisition of the N-terminal genomic sequence of AbCySp2.*

Previously, genomic DNA sequencing was used to retrieve information of the 5' end of silk genes because some spider silk genes lack introns (12); thus, the genomic DNA sequencing method was chosen to retrieve the 5' end of *AbCySp2*. The sequence for the 5' end of *AbCySp2* shows high similarity with *AbCySp1* (Identities = 271/292 (93%), alignment data not shown). Remarkably, *AbCySp2* and *AbCySp1* are primarily composed of homogeneous assemble repeats (55); therefore, a *AbCySp2* specific primer set was designed. The forward primer: 5' GGATGCCATTGCCCAGGTGATAC 3', was located in the 5' non-repetitive region of the gene, while the reverse primer: 5' AAACGCCCTGACCGGCTAAGAAT 3' was located in the first repeat unit of *AbCySp2O*. A 650 bp genomic PCR product was amplified, as indicated by the red arrow in (Figure 15).



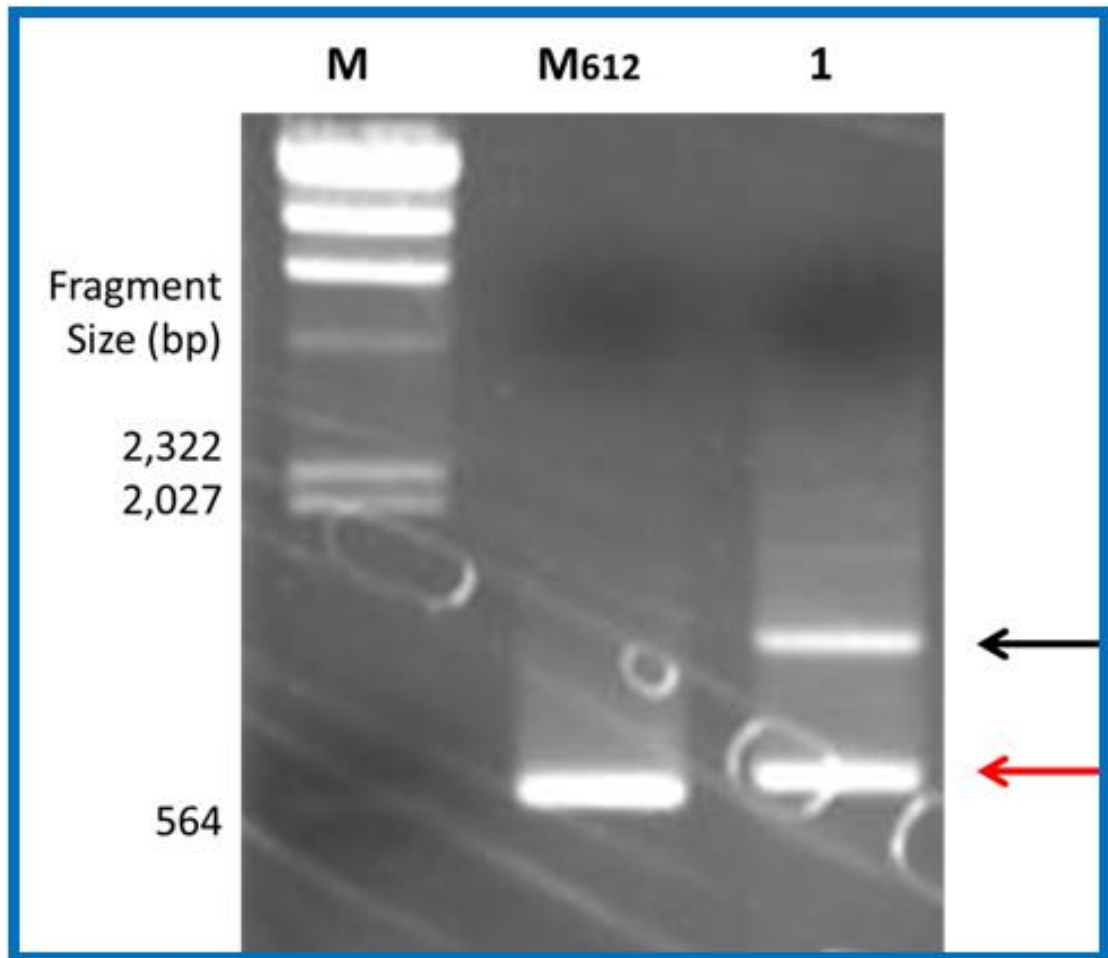


Figure 15. PCR amplification of genomic DNA to retrieve the 5' end of AbCySp2. Lane M is  $\lambda$ -Hind III digest DNA MW Standard Marker (TaKaRa). Lane M612 is a self-made DNA marker (612 bp). Lane 1 is the genomic DNA PCR product. The red arrow indicates the DNA fragment for AbCySp2, spanning from the 5' end non-repetitive region to the first repeat unit. The black arrow shows the DNA fragment that includes the 5' end non-repetitive region and several repeat units (possibly two).

#### *DNA fragment Sequencing and Computer analyses.*

The small DNA fragment (indicated by red arrow in (Figure 15)) was purified from the agarose gel and then ligated into the T-Vector for sequencing. The small fragment was found to be 655 bp.

By combining 655 bp with the sequence reported by A. C. Zhao (2006) , a new genomic DNA sequence was derived for the 5' end of *AbCySp2*, *AbCySp2N*. Sequence analysis for *AbCySp2N* (data not shown) demonstrated 4 sections: the 5' UTR (1 bp to 31 bp), the non-repetitive N-terminal sequence (32 bp to 562 bp), the first “A” type repeat unit (Repeat A0, 563 bp to 853 bp) and the partial sequence for the first “B” type repeat unit (Repeat B1, 854 bp to 872 bp).  
*Comparison AbCySp2O with AbCySp2N.*

*AbCySp2O* was reported as a mRNA (55), therefore, *AbCySp2O* was compared with *AbCySp2N*. The result is schematized in(Figure 16). Compared to *AbCySp2N*, *AbCySp2O* lacks a 349 bp region spanning from 312 bp to 660 bp; this missing region contains a partial non-repetitive N-terminal region (312 bp to 562 bp) and a partial Repeat A sequence (563 bp to 660 bp). Notably, the alignment result indicated 7 bp (312 bp to 318 bp) direct repeats flanked the deleted fragment, suggesting that plasmid pCySp2 used for *AbCySp2O* intermolecular homologous recombination happened in *Escherichia coli* (Figure 17).

```

AbCySp2o  GACAGTGAGGAACCCTTTCAATCAAGCAACTATGATGTGGTTCACTACAG
          |||
AbCySp2n  GACAGTGAGGAACCCTTTCAATCAAGCAACTATGATGTGGTTCACTACAG
          |||
          TAGCGTTCTCTTATGCCTTTTAGGAGCAGTGTCTGCACAATCAGTTGCA
          |||
          TAGCGTTCTCTTATGCCTTTTAGGAGCAGTGTCTGCACAATCAGTTGCA
          |||
          GTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGTTT
          |||
          GTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGTTT
          |||
          CCTTCAGTGTCTCACATTTGGAATCGGAAATTTCTCCCGCATTTCTACTC
          |||
          CCTTCAGTGTCTCACATTTGGAATCGGAAATTTCTCCCGCATTTCTACTC
          |||
          AAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACCAATGCCGTT
          |||
          AAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACCAATGCCGTT
          |||
          TCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAGGCTTTAAGTAC
          |||
          TCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAAGCTTTAAGTAC
          |||
          AGCGCTTGCAATCTTCTCTGACAGATCTGCTCATTGCAGAGTCGGCAAAAA
          |||
          GCAATTACAGCAATCAGCTGTCTGAACTAACAGGAATCCTCTCCGACTGT
          .....
          TTTATCCAACTACTGGATCGGACAATCCAGCATTGTGTCCAGAATTCA
          .....
          ATCTCTCATTTTCAAGTGTCTCCGAGAATGCAGATGTAAATATAATTTCTT
          .....
          CGGCAGGGATACCTTCAGTGTCTAGGAAGAAGTGGAGCTAGAGGAGGATTC
          .....
          TCTAACGCCGCAAGCCAGTCCGCAAGCAGYGCTGCCTCGCAAGCAAGTGC
          .....
          CTCTTCATTCGCACAAGCATCCTCTGCTTCCCTTGCAGCCTCAAGTTCTT
          .....CTTCTCTTCGGCCAATTCCTCTCAGCTCGGCAACGTTGC
          |||
          TCTCCAGTGCCTTCTCTTCGGCCAATTCCTCTCAGCTCGGCAACGTTGC
          |||
          TTATCAATTAGGCTTCAACGTAGCTAACACTCTCGGTATCGGCAACGCTC
          |||
          TTATCAATTAGGCTTCAACGTAGCTAACACTCTCGGTATCGGCAACGCTC
          |||
          CAGGCCTCGGTGCTGCCTTATCTCAAGCTGTCTCTCAGTCGGCGTTGGA
          |||
          CAGGCCTCGGTGCTGCCTTATCTCAAGCTGTCTCTCAGTCGGCGTTGGA
          |||
          GCCAGTTCCAGCACGTACGCTAATGTCGTTTCCAATGCAGTTGGACAATT
          |||
          GCCAGTTCCAGCACGTACGCTAATGTCGTTTCCAATGCAGTTGGACAATT
          |||
          CTTAGCCGGTCAGGGCGTTT
          |||
          CTTAGCCGGTCAGGGCGTTT
    
```

Figure 16. Alignment result of AbCySp2O and AbCySp2N by SIM4. The start codon is highlighted by red frame, and the direct repeats are highlighted by red characters. The PCR primer set is highlighted by red frame too.

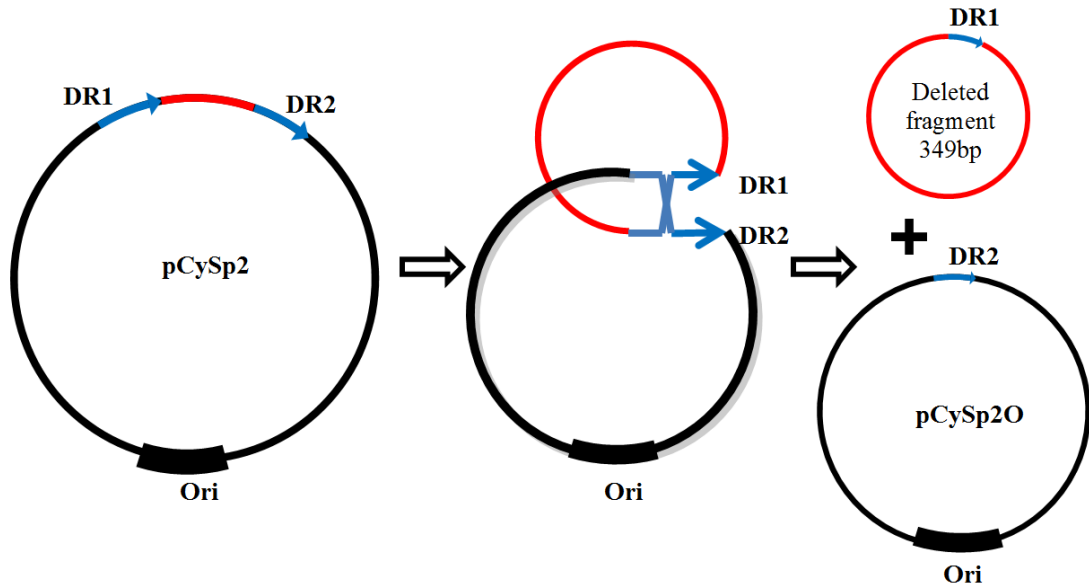


Figure 17. A schematic illustrating the intermolecular homologous recombination event that occurred in pCySp2. The homologous recombination between direct repeats leads to the formation of a smaller plasmid containing the Ori element (origin of replication) and the DNA fragment lacking the Ori element. The blue arrows indicate the DR (direct repeat), and the red part represents the deleted 349 bp region.

#### *Signal peptide analysis.*

The amino acid residue (Figure 18) was deduced from *AbCySp2N* by Genetyx ver.10 with the second reading frame (RF-2), and it was analyzed online using signalP4.0 (eukaryotic option) (67). The analysis highlighted an apparent signal peptide in *AbCySp2N* (Figure 19). The predicted cleavage site is located between amino acids 29 and 30. Corresponding to the mature protein start site at position 30 (Glutamine, Q), which is highlighted using red character in (Figure 18).

```
>CySp2N.
1      MMWFTTVAFL LCLLGAVSSQ
21     SVAVTAVPSV FSSPNLASGF
41     LQCLTFGIGN SPAFPTQEQQ
61     DLDAIAQVIL NAVSSNTGAT
81     ASARAQALST ALASSLTDLL
101    IAESAHSNYS NQLSELTGIL
121    SDCFIQTTGS DNPAFVSRIQ
141    SLISVLSQNA DVNIISSAGI
161    PSVSGRSGAR GGFSNAASQS
181    ASSAASQASA SSFAQASSAS
201    LAASSSFSSA FSSANLSLAL
221    GNVAYQLGFN VANTLGIGNA
241    PGLGAALSQA VSSVGVGASS
261    STYANVSNA VGQFLAGQGV
```

Figure 18. Amino acid residue sequence was deduced from *AbCySp2N*. The initial amino acid sequence (Met) was translated using the start codon highlighted in red, and the predicted signal peptide cleavage site is also highlighted using red character.

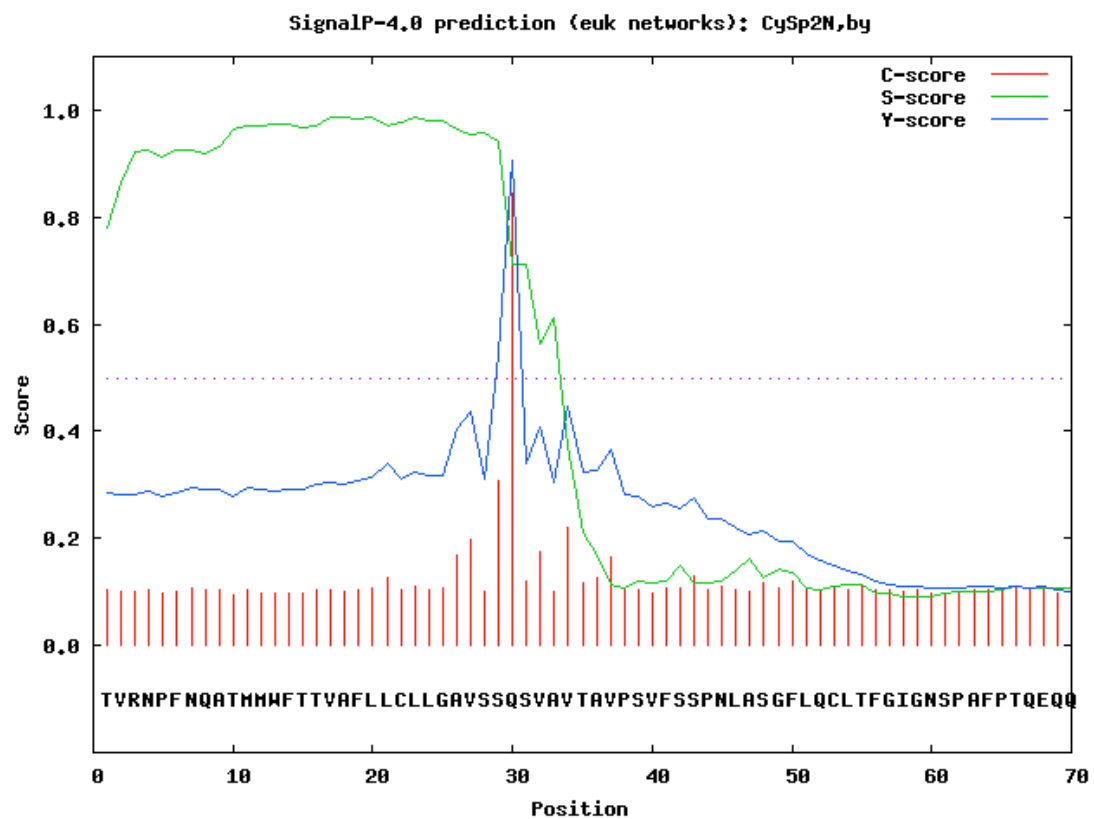


Figure 19. Signal peptide analysis for AbCySp2N using signalP 4.0

*The homologous relationship of N termini between AbCySp1 and AbCySp2N.*

Compared to *AbCySp2O*, the above analysis revealed that *AbCySp2N* had another 349 bp. The N terminal of *AbCySp2N* was aligned with *AbCySp1*. The result indicated that the two non-repetitive N termini shared high homology to each other at the nucleotide sequence level (Identities = 497/534 (93%), and at the amino acid sequence level (Identities = 164/178 (92%),(Figure 20)).

```

Identities = 164/178 (92%), Positives = 169/178 (94%), Gaps = 1/178 (0%)
Query: 1  MVWLTSIAFLVGFLGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ 60
          M+W T++AFL+ LGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ
Sbjct: 1  MMWFTTVAFLLCLLGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ 60

        61  DLDAIAQVILNAVSSNTGATASARAQALSTALASSLTDLLIAESAESNYSNQLSELTGIL 120
          DLDAIAQVILNAVSSNTGATASARAQALSTALASSLTDLLIAESA+SNYSNQLSELTGIL
        61  DLDAIAQVILNAVSSNTGATASARAQALSTALASSLTDLLIAESA+SNYSNQLSELTGIL 120

       121  SDCFIQTTGSDNPAFVSRIQSLISVLSQNADTNISSAGIPSVSGRRGAGGLGFDNTA 178
          SDCFIQTTGSDNPAFVSRIQSLISVLSQNAD NISSAGIPSVSGR GA G GF N A
       121  SDCFIQTTGSDNPAFVSRIQSLISVLSQNADVNISSAGIPSVSGRSGARG-GFSNAA 177
    
```

Figure 20. The non-repetitive N-terminal sequence alignment generated for *AbCySp1* and *AbCySp2N*. Query : *AbCySp1* non-repetitive region DNA sequence; Subject: *AbCySp2N* non-repetitive region DNA sequence.

## DISCUSSION

According to previous reports, spider silk genes are either single exon genes (12) or have an exon-intron structure (66, 74). Compared to *AbCySp2O*, *AbCySp2N* had an additional 349 bp fragment. *AbCySp2N* is a genomic DNA sequence. To determine the odds that the deleted 349 bp fragment would be in an intronic region, SIM4 was used to analyze the two sequences (*AbCySp2O* and *AbCySp2N*). The result demonstrated that the 349 bp fragment is not an intronic sequence but a mRNA

sequence that was transcribed from genomic DNA. In other words, the *AbCySp2O* is not a splice variant (65).

Secretory signals are key elements of spider and silkworm silk proteins and are important in the secretion of these proteins via transport across the endoplasmic reticulum (28). According to Zhao's report, *AbCySp2O* lacked a signal peptide (55). If the intact RF (RF-2) was used for gene analysis, the start codon selected by A. C. Zhao et al. (2006) would encounter numerous stop codons in the downstream sequence of the deleted fragment. Based on this reasoning, *AbCySp2O* would not include a signaling sequence. Similar results from other reports indicate that spider silk genes lack a signal peptide(64); this may be an artifact due to lack of upstream sequences for some of these genes. It is suggested that, if the 5' end sequence containing a non-repetitive N-terminal region, inverse PCR is good method to obtain the full N-terminal sequence.

The hypothesis, which states that: "N-terminal Cys residues have a role on linkage between *AbCySp1* and *AbCySp2*, similar to the function of C-terminal Cys residues seen in other silks. The *AbCySp1-AbCySp2* complex, which would possess a signal peptide, would then be transported across the endoplasmic reticulum and secreted to the Golgi" (55), may also be incorrect. According to recent reports, the N and C termini play key roles during silk protein storage and assembly and do not have a signal peptide function (70). Conversely, the silkworm strains *Nd-s* and *Nd-s<sup>D</sup>* produce extremely small amounts of fibroin, not because the mutated L-chains gene lacking the DNA sequence of the signal peptide region but because the disulfide linkage (located in the 3' end of the genes) between the H and L chains is not formed in *Nd-s* and *Nd-s<sup>D</sup>* mutants because of the lack of Cys-172 in the mutant L chain (75).

The artificial frameshift mutation in the N-terminal region of AbCySp2O completely alters the amino acid sequence encoding this region. Consequently, previous reports incorrectly concluded that AbCySp1 and AbCySp2 contain markedly different N termini.

Why do these types of mistakes persist for long periods of time? In my opinion, there are two major reasons for this. The first is geographical isolation. Some species of spiders only exist in a few countries. Other researchers do not usually have access to materials for their experiments. In order to prevent invasion by alien or exotic organisms, most countries do not allow exotic species such as spiders to pass through Customs. Secondly, the full-length silk genes have been always generated by sequencing cDNA or genomic libraries. It is costly to retrieve full-length spider silk genes.

In summary, the N-terminal region of the AbCySp2 and found that mistakes in a former report resulted in a 349 bp deletion event was reanalyzed. The analysis revealed that the N-terminal region of AbCySp2 had a signal peptide sequence and is highly homologous to AbCySp1 N-terminal region. Furthermore, there is no intronic sequence located between the non-repetitive N-terminal region and the first Repeat A and B of *AbCySp2*. And, correct information of *AbCySp2* obtained here will provide a theoretical basis for the gene's industry application in the future.



## **Chapter 2 Expression dragline silk protein in the silkworm using transgenic method**

## **Paragraph 1: Generation of transgenic silkworms which including the MaSp**

Spider dragline silk is a unique fibrous protein with a combination of tensile strength and elasticity, which more than twice as tough as any previously described silk, and over 10 times tougher than Kevlar(76). However, due to their territorial carnivorous nature, it is very difficult to obtain large amounts of fibrous spider silk from spiders (58). Thus, to obtain large quantities of spider silk, molecular biologists have been attempting to clone silk genes for overexpression in many different types of organisms, such as *Escherichia coli*, yeast, mammalian cells, higher plants and silkworms (31-36, 52, 59). The silkworm, *Bombxy mori*, is a domestic and well-studied silk-secreting arthropod insect in sericulture for several thousand years. Silkworm is a good candidate as a bioreactor for production of foreign proteins of interest, because it can synthesize vast amount of silk protein in its silk glands and spin it into cocoon. Silkworm silk and Spider silk are similar to each other in structures, which indicated that there would be some special advantages using silkworm to express silk protein and form new silkworm silk with similar physical properties of spider fibers(36).

To date, great progress has been made in research on Silkworm Transgenesis. In 2000, Tamura et al. reported a method for stable germline transformation in silkworms by using a *piggyBac* transposon-derived vector, inaugurating a new era of Silkworm Transgenesis.(36, 37)

Previously, we cloned and characterized several novel spider silk genes and observed their expression in *B. mori*(36). In this report, the expression of a spider dragline gene: *Cm.MaSp1s* by germline transformation of the silkworm *Bombyx mori* using the piggyBac transposon-derived vector(37) with the *B. mori* FibH promoter was further attempted.

## MATERIALS AND METHODS

### Material

The *Bombyx mori* N4 (white eye) strain were maintained in my laboratory. Larvae (from the first to the third instar) were reared on artificial diets at 25°C, and then the larvae (the fourth and fifth instar) were reared on mulberry leaves at 25°C.

### Methods

#### Cloning the repetitive unit of *MaSp1s*

The repetitive unit of *MaSp1s* was amplified from cDNA library with the primers:

5' GCTAGCACAAATACCATTCTTCGT 3' (forward) and 5' ACTAGTTGTAGCTGCAGCCGAAGCC 3' (reverse).The amplification program consisted of an initial denaturing step (95°C for 5 min), followed by 28 cycles of denaturing (94°C for 30s), annealing (58°C for 30 s), and extension (72°C for 1min), and a final elongation step at 72°C for 10 min. After electrophoresis, the DNA band of interest was excised from agarose using a clean razor blade and purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega*). The purified DNA fragment of interest was cloned into the T-Vector pMD19 (*TaKaRa*) using DNA Ligation Kit Ver.2.1 (*TaKaRa*). The resulting vector was sequenced from both directions using universal sequencing primers, M13 primer RV and M13 primer M4 (*TaKaRa*), and an ABI PRISM® 3100 Genetic Analyzer.

#### Vector construction

The sequence validated PCR products (*MaSp1s*) were cut from the T-Vector with the restriction enzymes *Nhe* I and *Spe* I and then cloned into a pSLfa1180fa vector to achieve the plasmid pSL-FibH-*MaSp1s*. Then, the gene fragment coding for the carboxyl terminus of the silkworm fibroin H-chain, named the L-chain binding site (LBS), was amplified from silkworm genomic DNA with a pair of primers, 5'-CTAGCTAGCAGTTACGGAGCTGGCAGGGGATACG-3' (forward) and 5'-CGGGATCCTAGTACATTCAAATAAAATGCATAC- 3' (reverse). The PCR product was digested with *Nhe* I and *Bam*H I and then cloned into the plasmid (36). The resulting vector named as pSL-FibH-*MaSp1s*- LBS(Figure 21). The resulting plasmid was then digested with *Asc* I. The intact expression cassette consisting of a FibH promoter, MaSp1S gene, and LBS fragment was inserted into the pBac[3xP3-DsRedaf] vector at the *Asc* I site(77, 78). The final vector was designated as pBac[3xP3-DsRedaf]-FibH-*MaSp1s*-LBS(Figure 21).

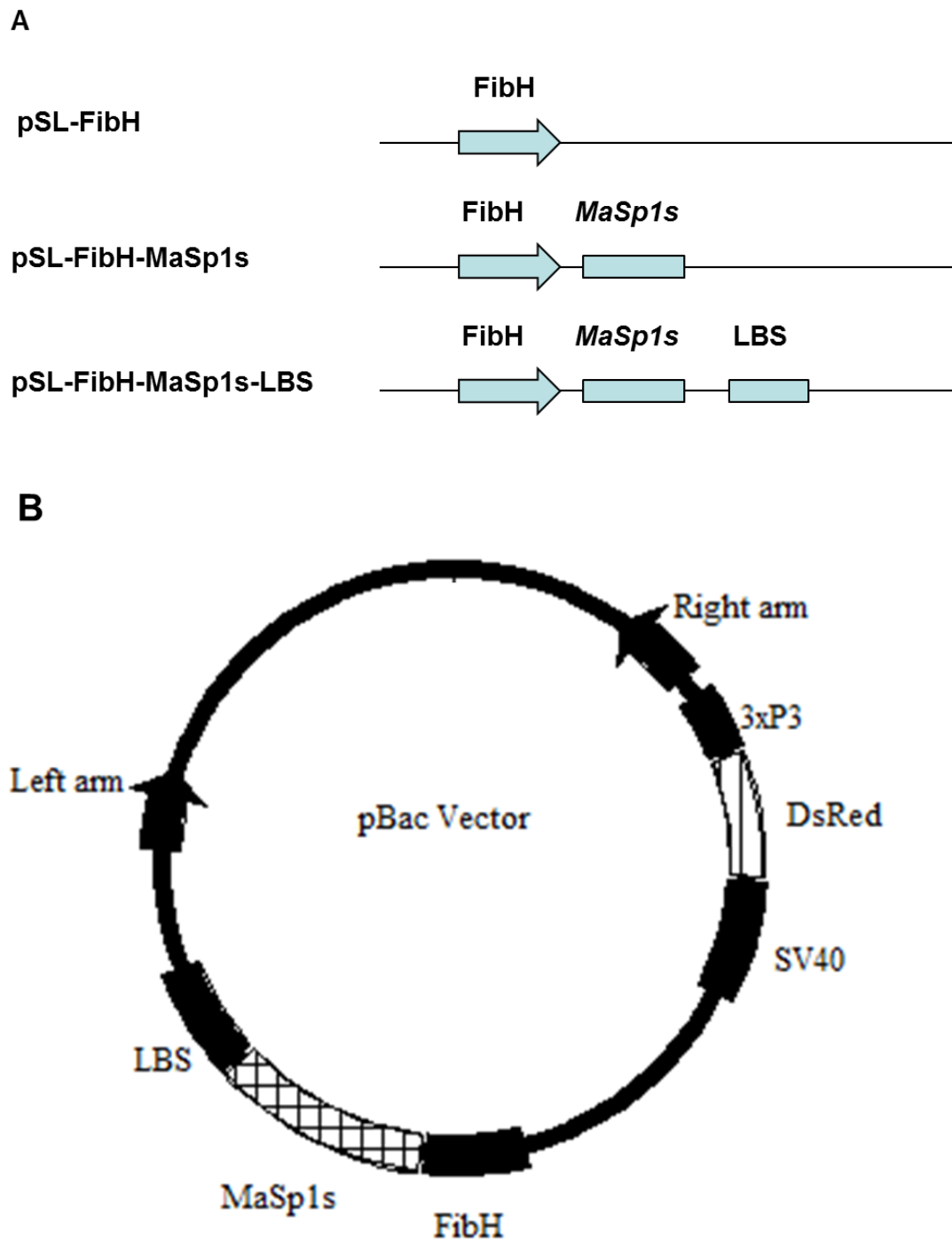


Figure 21. Schematic of the transformation vector structure. (A) Structures of pSL vectors. pSL-FibH-MaSp1s carried silkworm fibroin H-chain promoter (FibH) upstream of repetitive unit of MaSp1s (MaSp1s). pSL-FibH-MaSp1s-LBS was the gene fragment coding for the carboxyl terminus of the silkworm fibroin H-chain been introduced into

pSL-FibH-MaSp1s at the Nhe I and BamH I site. (B) The expression cassette consisting of a FibH promoter, MaSp1S gene, and LBS fragment was inserted into the pBac[3xP3-DsRedaf] vector. The DsRed gene driven by promoter p3xP3 was used as a transformation reporter.

### **Transgenesis and screening of silkworms**

Plasmid pBac[3xP3-DsRedaf]-FibH-*MaSp1s*-LBS was purified with the QIAGEN Plasmid Midi kit (*Qiagen*). A nonautonomous plasmid, pHA3PIG, was used as the helper for the production of transposase (37). Embryo injection was performed as described by Tomita (50). After injection, the embryos were allowed to develop at 25°C. G0 moths were mated randomly, and the G1 embryos were screened by detecting DsRed fluorescence under an Olympus SZX12 fluorescent stereomicroscope (*Olympus*). G1 positive transgenic individuals were mated within the same family to generate the G2 descendents for the following analysis. (36).

### **Result**

It is reported that repetitive units of MaSp1 proteins, which determine the strength and elasticity of dragline silk(70). So, the repetitive unit of *MaSp1s* was cloned. And it was inserted into the piggyBac transposon-derived vector(Figure 21). In order to simplify the screening steps I chose the N4 white eye strain(Figure 22).

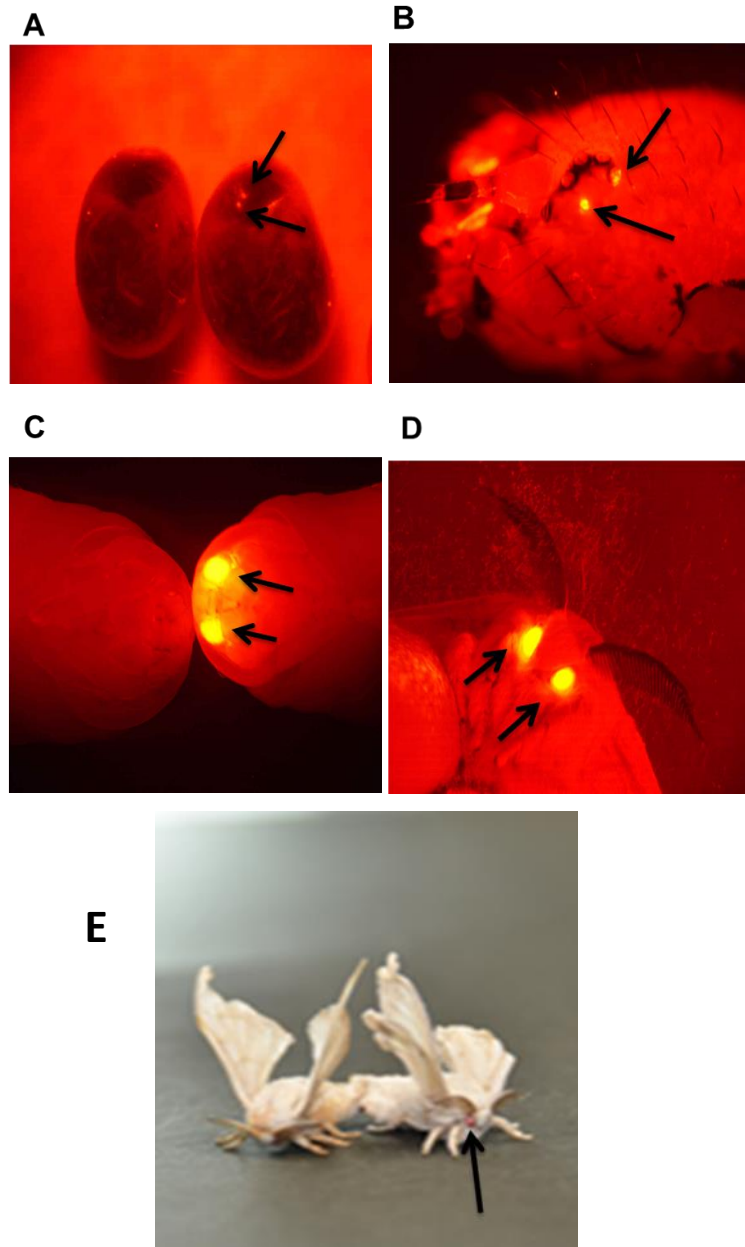


Figure 22. Fluorescence of DsRed in transgenic silkworms bearing pBac[3xP3-DsRedaf]-FibH-MaSp1s-LBS. Panel A shows G1 eggs with DsRed-positive embryos were viewed under DsRed excitation wavelength(left is a control). Panel B shows G1 larva (the fifth day of the fifth instar). Panel C shows G1 Pupae (left is a control). Panel D shows G1 imago. A-D were all viewed under DsRed excitation wavelength. Panel E shows G1 imagos viewed under visible light. The arrows point to the eyes.

Four distinct stages of development complete one generation of the silkworm: ova, larvae, pupa and imago. The silkworm larval life is divided into five instars, separated

by four molts. The G1 eggs were screened by detecting DsRed fluorescence under the fluorescent stereomicroscope. And the other three stages were also detected with DsRed fluorescence and visible light(Figure 22).



## **Paragraph 2: Biological property of the transgenic silkworms which including MaSp1s**

To date, great progress has been made in research on Silkworm Transgenesis; a lot of transgenic method was used for the purpose.(38-51). Tamura et al.(2000) reported a method for stable germline transformation in silkworms by using a piggyBac transposon-derived vector, inaugurating a new era of Silkworm Transgenesis(36, 37) . In my research, the piggyBac transposon-derived vector system was chosen to obtain the transgenic silkworm expressing the spider dragline silk gene MaSp1s.

Previously, researchers have cloned and characterized several spider silk genes and observed their expression in *B. mori* (36, 52). But, there is limited molecular information on the biological properties of transgenic silkworms carrying a foreign silk gene. The purpose of the present work was to provide insight into the biological properties and structure of transgenic silkworm fibroin genes and proteins.

The results showed the following: 1 At the DNA level, *MaSp1s* (major ampullate spidroin 1) was integrated into the *B. mori* chromosome at scaffold 868(79); 2 At the RNA level, the transcription of *MaSp1s* did not affect the transcription of *Fib-H*, *Fib-L* or *p25*, and the transcription level of *MaSp1s* was higher than *Fib-H*; and 3 At the protein level, MaSp1s, the foreign silk protein, had low abundance among the total silk fibroin proteins (approximately 2%). Additionally, it had a protein-protein interaction with Fib-L through its LBS (L-chain-binding site) region through intermolecular disulfide bonds.

## **MATERIALS AND METHODS**

### **Material**

G1 positive transgenic individuals (mentioned in Chapter 2 Paragraph 1) were mated within the same family to generate the G2 descendents (test cross) for the following analysis.

## **Methods**

### *Genomic DNA extraction and inverse PCR (iPCR)*

The genomic DNA was extracted from 25 mg moth head using the DNeasy Blood & Tissue Kit (*QIAGEN*), digested with *Hae* III (*TaKaRa*) for 20 h, and purified using the Wizard® SV Gel and PCR Clean-Up System (*Promega*). The products of the restriction enzyme digestion were ligated using the DNA Ligation Kit Ver.2.0 (*TaKaRa*) for iPCR. The primers for iPCR were as follows: The left primer pair was 5'-ATCAGTGACACTTACCGCATTGACA-3' and 5'-TGACGAGCTTGTTGGTGAGGATTCT-3'. Amplification was carried out with a 2 min denaturing cycle at 96°C followed by 40 cycles of 1 min at 96 °C, 30 sec at 65 °C, 4 min at 72 °C, and a final extension at 72 °C for 10 min. Amplification using the right primer pair (5'-TACGCATGATTATCTTTAACGTA-3' and 5'-GGGGTCCGTCAAAACAAAACATC-3') was carried out under the same conditions except that annealing was performed at 58 °C (37, 50). The purified DNA fragment of interest was cloned into the T-Vector pMD19 (*TaKaRa*) using DNA Ligation Kit Ver.2.1 (*TaKaRa*). The resulting vector was sequenced from both directions using universal sequencing primers, M13 primer RV and M13 primer M4 (*TaKaRa*). The sequencing results was blast on the Silkworm Genome Database (80).

### *RNA Preparation and Relative Quantitation of Gene Expression*

PSG was excised from larvae of transgenic and wild type silkworms at the three day of the fifth instar. Total RNA was extracted with *TRIzol* reagent (*Invitrogen*)

according to the manufacturer's protocol. cDNA was synthesized using PrimeScript™ RT reagent Kit (Perfect Real Time) (*TaKaRa*), following the manufacturer's instructions. mRNAs of *FibH*, *FibL*, *p25* and *LBS* were quantified using SYBR® Premix DimerEraser™ on Thermal Cycler Dice™ Real Time System (*TaKaRa*). In this study, house keeping gene actin A3 were used as internal standards. Relative quantitative real-time PCR was performed in triplicate. The primer sets used for actin A3 were 5' - ATCACCATCGGAAACGAAAGAT-3' and 5' - GGTGTTGGCGTACAAGTCCTTA-3'; for *p25* were 5' - ACCCTAATGACCAACGCACAG-3' and 5' - GCAACCGAAGTCGCAGAGTG-3'; for *fibH* were 5' - TGCTGGTGGTAGCGTCA-3' and 5' - GGACGTTACGACGAGAATAG-3'; for *fibL* were 5' - AACTCGTCATCAACCCTG-3' and 5' - AAGTCATAGATTCTTCCACCT-3'; for *LBS* were 5' - ACGGACAAGGTGCAGGAAGT-3' and 5' - GCGGCAGTGCTCTGAATTAAAC-3'; for *TG* were 5' - GCAGCAAGCGGAACTAG-3' and 5' - GGACGTTACGACGAGAATAG-3'.

#### *Preparation MaSp proteins for Polyclonal Antibody*

The sample of MaSp proteins were separated from the lumen of spider major ampullate silk gland and send to *Kitayama Labes* Company for making the polyclonal antibody: anti-MaSp.

#### *SDS-PAGE and immunoblot analysis of fibroin proteins*

One part of PSGs were dissected out from the larvae at the 5th day of the fifth instar, and the fibroin secreted into the PSG lumen was manually separated from surrounding tissues. The coagulated fibroin was immersed in 99.5% ethanol, washed three times in ether, and air-dried(71). The air dried coagulated fibroin protein was dissolved in 60% lithium thiocyanate (LiSCN) used for the SDS-PAGE and western

blotting analysis. The separated FibH, FibL protein band was excised from from SDS-PAGE gel for producing antibodies: anti-FibH and anti-FibL.

PSG was ground in liquid nitrogen. Then, total protein was extracted from the homogenate using the Dynabeads® Co-Immunoprecipitation Kit (*Invitrogen*) according to the manufacturer's protocol. One portion of the extracted proteins was directly used for co-IP. The other part of the sample was mixed with an equal volume of 26 mM DTT (dithiothreitol) under a N<sub>2</sub> atmosphere(71) and used for co-IP.

## Results and discussion

### *Genotyping analysis of the transgenic silkworms*

In order to detect the foreign gene *MaSp1s* inserted position, iPCR was used to analyze the flanking chromosome sequence of the expression cassette (Figure 23). The Blast N result showed that the inserted site located in scaffold868 (79).

```

Query: 65      ttaa aactcaaaactttactaaatgacgtaaaccgagccgaccgatctgtcaatacaacc 124
               |||||
Sbjct: 16507   ttaa aactcaaaactttactaattaacgtaaaccgagccgaccgatctgtcaatacaacc 16566

               gaattcaataaccgacgatcttttcaatagtcttcaagtcttcaataaa-atctttaata 183
               |||||
               gaattcaataaccgacgatcttttcaatggtcttcaagtcttcaataaataatcttcgata 16626

               tttcaataaaattcttcgaggcgagatgtctttactcgtgcgttttttggtgatcttcg 243
               |||||
               tttcaat-aaattcttcgaggcgagatgtctttactcgtgcgttttttggtgatcttcg 16685

               ttcgagggtcggcggcc 261
               |||||
               ttcgagggtcggcggcc 16703

```

Figure 23. The foreign gene *MaSp1s* inserted site detection. The query sequence is the iPCR product, and the Sbjct sequence is part of the silkworm genome sequence (scaffold868, Chr.25). The characteristic tetranucleotide sequence 'ttaa' highlighted by the yellow is the piggyBac transposon's the target site.

*Transcriptional analysis of key RNAs*

Silk fibroin produced by the silkworm *B. mori* consists of three components: a heavy chain (*Fib-H*), a light chain (*Fib-L*), and the glycoprotein P25(71). Quantitative real-time PCR was used to analyze the transcriptional level of these three genes after transgene insertion. There was no significant difference between the wild-type silkworm (WT) and transgenic silkworm (TG) in the transcriptional level of the three genes (Figure 24). Specifically, *Fib-H* had the same transcriptional level in both WT and TG silkworms (Figure 24). However, Figure 24 shows that the *LBS* expression level in TG silkworms was 2.8-fold higher than WT. Thus, the expression level of *MaSp1s* was 1.8-fold higher than *Fib-H*. Both *Fib-H* and *MaSp1s* contain *LBS* (Figure 24). However, after transgene insertion, the expression level of *Fib-H* did not change (Figure 24). Therefore, of the 2.8-fold difference in *LBS* expression, 1-fold is accounted for by *Fib-H*, whereas the remaining 1.8-fold is accounted for by *MaSp1s*. The expression of the foreign gene *MaSp1s* did not affect the expression of the other three genes at the RNA level, even though *MaSp1s* used the same promoter and *LBS* region as *Fib-H* for expression (Figure 24). Additionally, the expression level of *MaSp1s* was higher than *Fib-H*, which suggested the following: 1, the cell transcription system is equally responsive to *Fib-H* and *MaSp1s*, although *MaSp1s* is a foreign inserted gene; and 2, the transcriptional level of the foreign gene *MaSp1s* being higher than *Fib-H* may be because the expression cassette of *MaSp1s* is shorter than *Fib-H* (Figure 24). The foreign gene *MaSp1s* expression cassette may have a higher transcriptional efficiency than *Fib-H*.

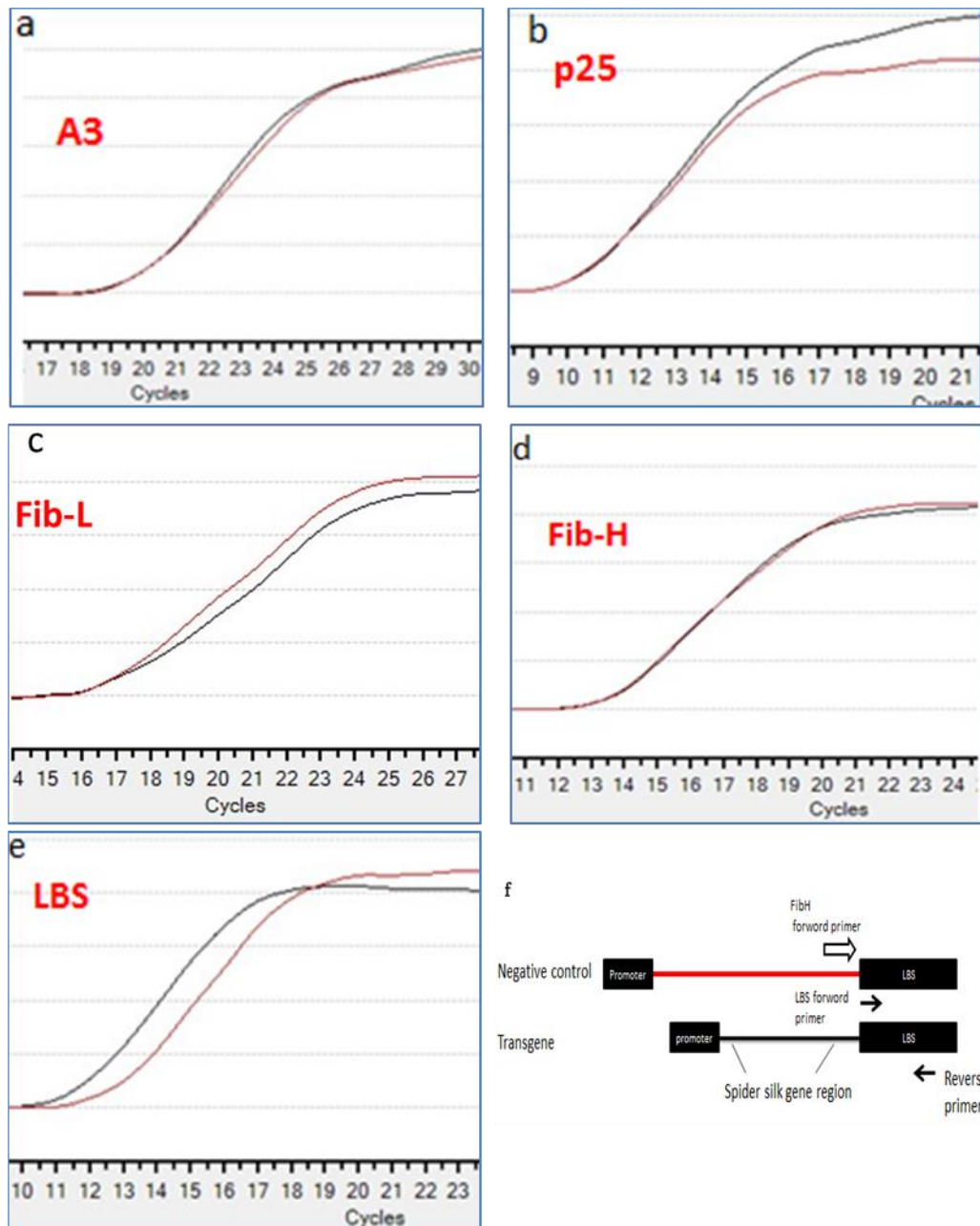


Figure 24. Typical relative quantitative PCR results. b-d show the transcriptional level of genes (A3, p25, FibL and FibH). e shows the significant difference between the TG and WT. The red line stands for WT, and the black line is the TG. f: Schematic of the primer design strategy for Fib-H and the foreign gene MaSp1s.

### *Protein expression analysis*

The fluorescence had been detected from the G1 moth, which is just the evidence that *DsRed* gene had been expressed in the silkworm. Whether, the foreign protein

MaSp1s had been expressed into the silk PSG lumen is still unknown. To validate the foreign protein MaSp1s had been expressed, SDS-PAGE was used to separate the fibroin proteins of G2, in addition to the three proteins (FibH, FibL and p25), the other type of protein was present(Figure 25). The abundance the foreign protein (about 2% among the total fibroin proteins) is much lower than that of FibH and FibL, but richer than p25. The expression of the foreign gene *MaSp1s* at protein level was confirmed by Western blotting by the anti-MaSp(Figure 26).

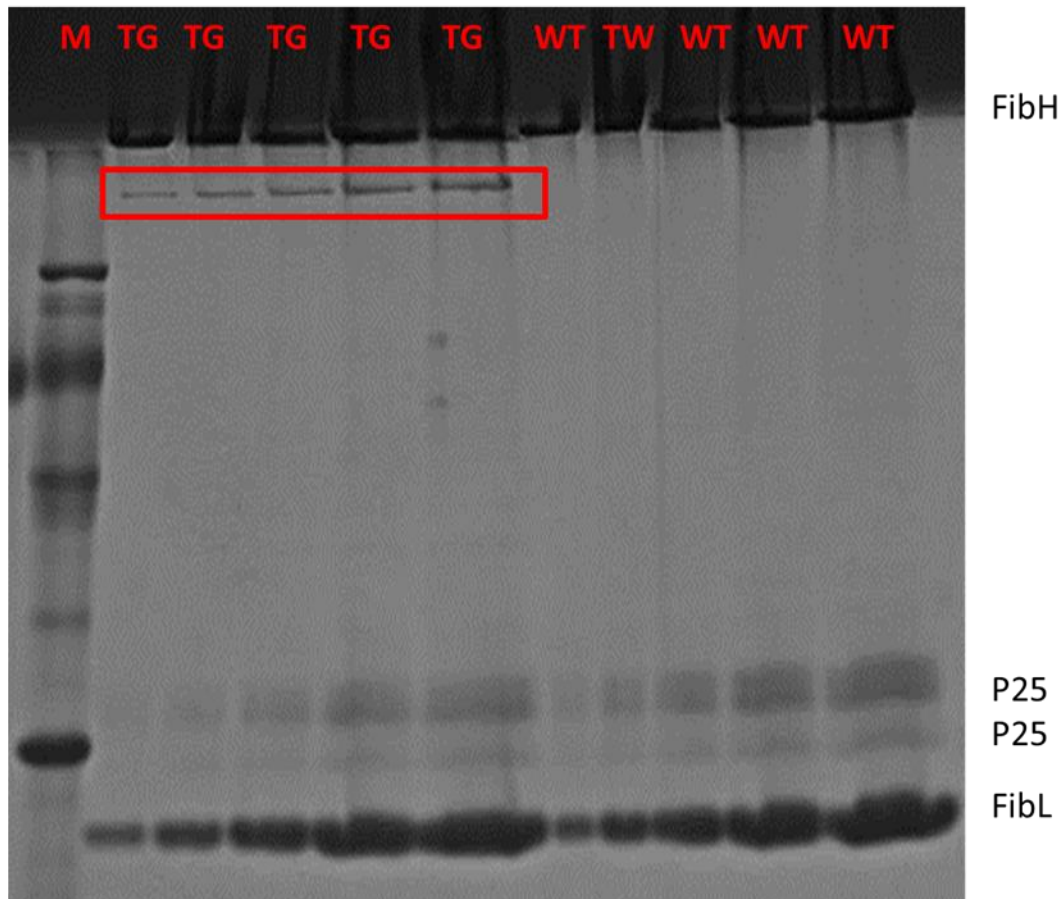


Figure 25. SDS-PAGE of luminal fibroin proteins. The image shows an additional protein band in TG compared to WT. M indicates the molecular mass marker (GE). The abundance of the foreign protein (approximately 2% of the total fibroin proteins) is lower than Fib-H and Fib-L but higher than p25. TG is the transgenic silkworm. WT is the wild-type silkworm.



Figure 26. Validation of the MaSp1s protein existing in the silkworm fibroin proteins by Western blotting using the antibody : anti-MaSp.

Although, the transcription system is equally responsive to both of them, yet the translation system is not. The abundance of MaSp1s (2%) was much lower than Fib-H (92%). The problem can probably be attributed to the difference of the repetitive elements of the two silks (spider and silkworm silk). It is reported that both organisms appear to produce gland specific pools of tRNAs(18). Since the repetitive element of spider is composed of GGX (X=A, Q or Y) and poly-A (4-10 alanines), while the *B. mori* is (Gly-Ala-Gly-Ala-Gly-Ser)<sub>n</sub>, it may be necessary to have a larger pool of alanine tRNA during translation of spider silk. What's more, the amino acid composition (mol %) is also different from each other, particularly the glutamine (Q). The spider silk gene contains 11.3%, while just 0.19% in Fib-H.

#### *The interaction relationship between MaSp1s, FibL and FibH*

It is reported that the fibroin protein complex of approximately 2.3 MDa, designated an elementary unit of fibroin having 6:6:1 molar ratios of the heavy chain, light chain, and P25, existed in posterior silk gland cells(71, 81). And the disulfide linkage between Cys-172 of L-chain and Cys c20 (20th residue from the C terminus) of H-chain has been demonstrated to be essential for the secretion of a large amount of fibroin from PSG cells (82). The L-chain binding site (LBS), which consists of the fibroin H-chain gene 3' terminal sequence and the 3' flanking sequence including Cysc20 and the putative polyadenylation signal sequence of H-chain(51). Based on those theories, LBS sequence was cloned to the end of *MaSp1s* to ensure the foreign protein can secreted from PSG cells.



According to the above vector construction strategy, if the MaSp1s can secrete from PSG cells into PSG lumen, the foreign protein would interact with FibL protein. To test this hypothesis, we detected the interaction between MaSp1s and Fib-L using co-IP (co-immunoprecipitation). Cross-validation results showed that MaSp1s was detected by Western blotting in the complex pulled down from total protein by anti-Fib-L(Figure 27). Furthermore, Fib-L was detected by Western blotting in the complex that was pulled down from total protein with anti-MaSp(Figure 27). Additionally, Fib-H was detected in both complexes mentioned above (data not shown). The results demonstrated that the Fib-L protein and the foreign protein MaSp1s did form a protein complex, as expected. Furthermore, Fib-H also participates in the assembly of the fibroin complex. In addition, to confirm the interaction between Fib-L and the foreign protein MaSp1s through disulfide linkage, the total protein was treated with DTT at 13 mM(71) before co-IP. The Western blotting results showed that the foreign protein MaSp1s was not detected in the complex that was pulled down from DTT-treated total protein with anti-Fib-L

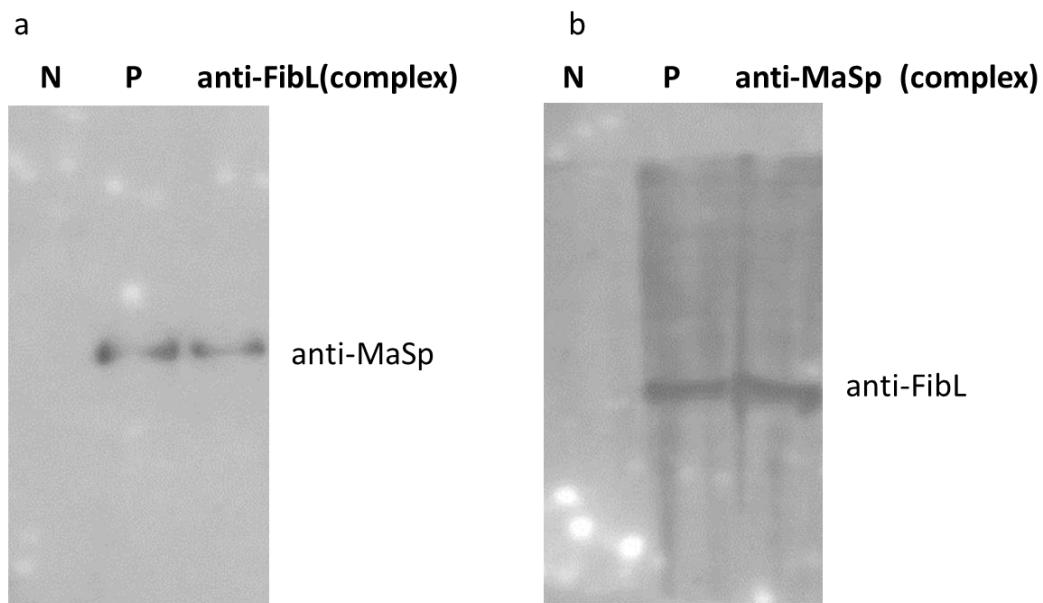


Figure 27. Cross-validation results of the protein-protein interaction between MaSp1s

and Fib-L. (A) The complex pulled down from total protein with anti-Fib-L antiserum. The captured complex reacted with anti-MaSp. Lane N is the negative control (wild-type silkworm). Lane P is the positive control (the antigen protein used for antibody produce). Lane anti-Fib-L(complex) is the anti-Fib-L-captured complex from transgenic silkworm reacting with anti-MaSp. (B) The cross-validation results in contrast to (A). Lane anti-MaSp(complex) is the anti-MaSp-captured complex from transgenic silkworms reacting with anti-Fib-L.

The fibroin protein complex is composed of an elementary unit of fibroin, with 6:6:1 molar ratios of Fib-H, Fib-L and P25, in posterior silk gland cells(71, 81, 83). Based on the present biochemical analysis, a new fibroin elementary unit model is presented. The transgenic silkworm fibroin consists of four proteins, Fib-H, MaSp1s, Fib-L and P25, at  $n:(6-n):6:1$  molar ratios.

## Conclusions

In this study, the biological properties and structure of fibroin genes and proteins of transgenic silkworm were analyzed by biochemical methods. The results demonstrated that the inserted foreign gene *MaSp1s* did not affect the transcriptional level of the other three genes. The MaSp1s protein can interact with Fib-L through disulfide linkage to form a fibroin complex whose composition is  $H_nM_{(6-n)}L_6P25_1$ . We hope that the results presented here will be helpful to this research field.

## **Paragraph 3: Physical properties and Structure of silk from the transgenic silkworm**

### **Introduction**

Humans have utilized insect silk for their benefit and comfort for thousands of years. The most famous example is the using of reeled silkworm silk from *Bombyx mori* to produce textiles(84). Spider dragline silk is a unique fibrous protein with a combination of tensile strength and elasticity, which more than twice as tough as any previously described silk, and over 10 times tougher than Kevlar(76). Unlike the silkworms, spiders aren't easy to domesticate. Since farming the spiders is not commercially viable due to their highly territorial and cannibalistic nature(84). So it is very difficult to obtain large amounts of silk from spiders. Silkworm is a good candidate as a bioreactor for production of foreign proteins of interest (37, 49, 51). Because it can synthesize vast amount of silk protein in its silk glands and then spin into cocoon. Previously, we cloned and characterized several novel spider silk genes and observed their expression in the silkworms larval (36). But, there has been very limited research on the physical properties and structure of the transgenic silks including spider dragline silk protein.

In this paragraph, use the methods such as WAXD, FT-IR and DSC to detect the physico-chemical and ultrastructural characteristics of the silk. I hope that those results presented here would be a little helpful to this research field.

### **MATERIALS AND METHODS**

#### **Material**

The transgenic silkworm strain (CmMA) was maintained in our laboratory(Figure 28), which contained a foreign spider dragline silk gene (*MaSp1s*). Larvae (from the first to the third instar) were reared on artificial diets at 25°C, and then the larvae (the

fourth and fifth instar) were reared on mulberry leaves at 25°C. Silk fibers were produced by artificial reeling from spinneret of the transgenic silkworm at the speed of 8 mm/s at the initial stage of mounting the cocooning frame(85).

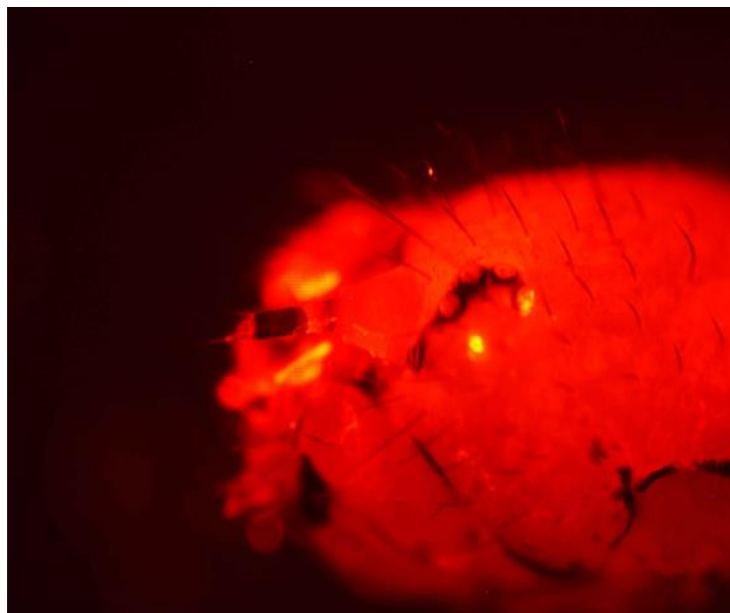


Figure 28. The Fluorescence graph of DsRed in transgenic silkworms bearing pBac[3xP3-DsRedaf]-FibH-MaSp1s-LBS.

## Methods

The morphologies of the fibers were examined with a JSM-6010LA InTouchScope at 10 kV of acceleration voltage. Before placing the samples in the SEM chamber, the samples were mounted onto an aluminum stud and sputter-coated with gold/palladium for 60 s (E-1010 ION SPUTTER, Hitachi, Japan) to prevent charging.

The Tensile properties were measured with a Tensilon Model UTM-II-20 (Orientec Corporation, Japan) using standard technique at 22 °C and 65% RH at a gauge length of 20 mm. The measurements were done with load stress of 5 N at a force rate of 5mm/min.

Fourier transform infrared (FT-IR) spectroscopy was measured with a Shimadzu FT-IR-8400S infrared spectrometer by the ATR method in the region of 4000~400  $\text{cm}^{-1}$  at room temperature.

Wide angle X-ray diffraction (WAXD) profile was obtained by a Rigaku Rotorflex RU-200B diffractometer using Ni-filtered Cu  $K\alpha$  radiation generated at 40 kV and 150 mA ( $5^\circ \sim 40^\circ$ ), at a scanning speed of  $1^\circ/\text{min}$ . XRD photographs were developed using an imaging plate.

Differential scanning calorimetry (DSC) measurement was performed by a Rigaku Denki Co., Ltd. Instrumental (model DSC-8230) at a heating rate of  $10^\circ\text{C}/\text{min}$  under  $\text{N}_2$  gas atmosphere.

The thermal stability of silks was carried out with a TG/DTA apparatus (TG/DTA6200, Seiko Instruments Inc., Japan) by heating from 27 to  $450^\circ\text{C}$  under a continuous nitrogen purge of 20 ml/min. The heating rate was  $20^\circ\text{C}/\text{min}$ .(86)

## Results

### *Surface morphology*

SEM was used to analyze the surface morphology of silk fibers produced by the 5th instar at the initial stage of mounting the cocooning frame. The typical longitudinal micrograph is shown in(Figure 29). It is observed that double-stranded silk fibers linkage to each with thin layer sericin. It is of interesting to note that silks have the uniform diameter around (18.75  $\mu\text{m}$ ), which maybe the result of artificial reeling.

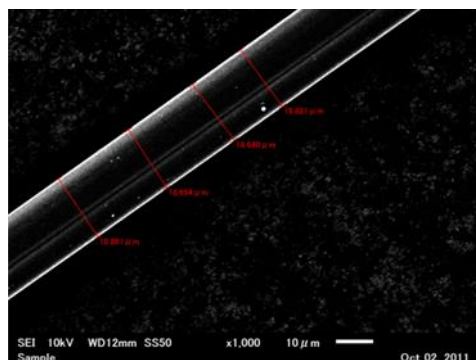


Figure 29. SEM images of silk fiber

### *Tensile properties*

The mechanical performance of transgenic silk fibers was elucidated in(Figure 30). The result showed that the average maximum breaking stress was around 340 MPa, and the breaking strain was about 45% (Because, there was no difference between them, the data of wild type silk is not shown ).

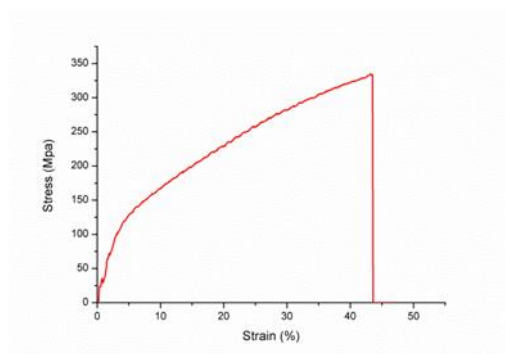


Figure 30. Typical tensile strength and elongation curves of silk fibers

### *Crystalline structure*

FT-IR spectroscopy is a powerful technique to study structure at the molecular level, and reveals typical absorption bands sensitive to molecular conformation (87). (Figure 31) shows FT-IR spectra of transgenic silks. The absorption bands at  $1620\text{ cm}^{-1}$  (Amide I) and  $1227\text{ cm}^{-1}$  (amide III) attributed to anti-parallel  $\beta$ -sheet conformation, and  $1520\text{ cm}^{-1}$  (amide II) assigned to  $\alpha$ -helix conformation (Because, there was no difference between them, the data of wild type silk is not shown ).

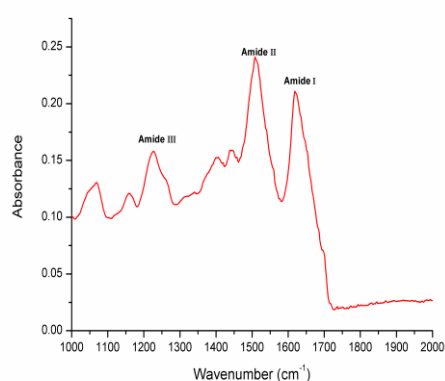


Figure 31. FT-IR profile of transgenic silk.

In order to elucidate the crystalline structure of aquatic silk fiber, WAXD profiles of silk fiber were recorded as shown in(Figure 32). X-ray diffraction curves showed that the major diffraction peak at scattering angle ( $2\theta$ ) of  $23.5^\circ$  (Because, there was no difference between them, the data of wild type silk is not shown ), and the smoothly diffraction peak suggested that the fiber exhibited a low degree of crystallinity (Figure 32).

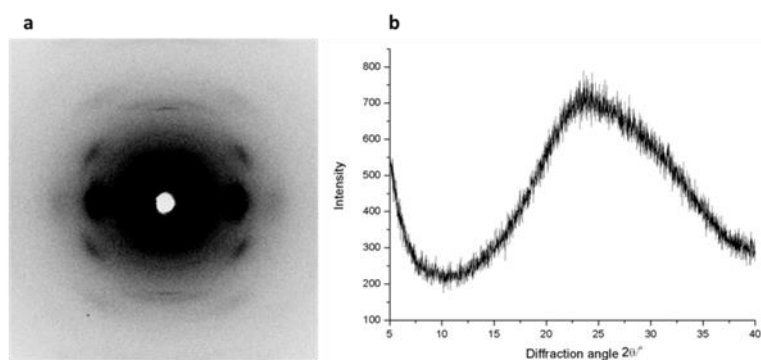


Figure 32. WAXD profiles of silk fibers: (a) X-ray pattern of silk fibers; (b) WAXD diffraction curves.

Depending on XRD photographs(Figure 32), the molecular orientation of the fibers was determined quantitatively by Herman's orientation factor:  $f = (3\langle \cos^2\phi \rangle - 1)/2$ . The parameter  $f$  is 0 for no preferred orientation in fibers and 1 if all crystals are perfectly aligned with respect to each other(88). The transgenic silk with with high orientation, the  $f$  is 0.9403(there was no significant difference between them, the data of wild type silk is 0.9425 ).

#### Thermal behavior

The thermal behavior of silk fiber from transgenic silkworm was examined on the basis of DSC measurement(Figure 33). Silk fiber showed a broad endothermic peak centered around  $89^\circ\text{C}$ , which is attributed to the evaporation of water contained



in the silk sample. It is elucidated that the endothermic peak around 326 °C is due to the thermal decomposition of silk fiber (Because, there was no difference between them, the data of wild type silk is not shown ). The TG/DTA curves showed that there was at 5 wt% weight-loss at 235.5 °C. (Figure not shown)

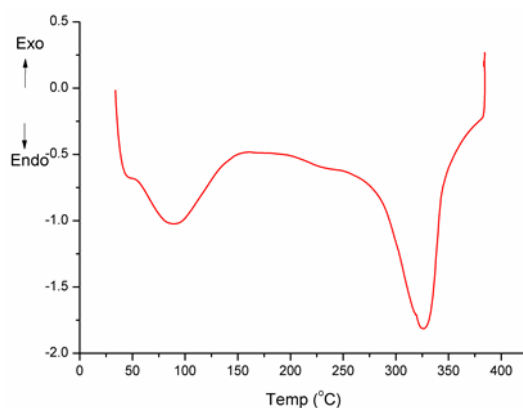


Figure 33. DSC curves of silk fiber.

## Conclusion

In this study, the physical properties and structure of silk from the transgenic silkworm and wild type silkworm were analyzed by many physical methods. The results demonstrated that there was no significant difference between the wild type silk and transgenic silk of their Surface morphology, tensile properties, crystalline structure, thermal properties. This result is in accord with the the former function hypothesis that MaSp1s would not produce a dominant effect in the determination of the strength and elasticity of silk.

## Summary

Spider silk receives the most interest from scientists for its unique combination of high tensile strength and elasticity. Previously, our lab had cloned and characterized several novel spider silk genes and observed their expression in *B. mori* cell lines and larval bodies.

In this study, two research topics were focused: one is that keeping on analyzing those novel spider silk genes (Chapter I). The other is that analyzing the transgenic silkworm and the silk from them (Chapter II).

Firstly, a unique *MaSp* gene (Chapter I paragraph I) had been isolated, *Cyrtophora moluccensis MaSp1s*. The full-length gene is only 1320 base pairs (bp), which encodes 439 amino acids that includes the intact non-repetitive N-terminal (149 residues), C-terminal (98 residues) and so-called repetitive regions (192 residues); the deduced molecular weight is approximately 40 kDa. It deviates from the large-size rule and is different from the previously reported MaSp1 (10 kb). The sequence analysis demonstrated that the two termini are highly homologous to the other characterized dragline silk genes but that the so-called repetitive region is different. The motif analysis suggested that MaSp1s included all of the MaSp1 characteristic motifs, such as GGX (X =A, Q, or Y), GX (X =Q, A, or R), and poly-A. My results suggest that *MaSp1s* is a possible new characteristic dragline gene; the discovery of this gene should enhance my understanding of the Major Ampullate spider silk genes. The function of MaSp1s is presently unclear. But, the N and C termini of MaSp1s are highly homologous with those of other MaSp proteins. It is reported that the

non-repetitive termini contribute to storage and assembly of spider silk proteins(70). This rationale supports the hypothesis that MaSp1s would be partially responsible for the storage and assembly of spider silk proteins, particularly with regard to the proteins assembling together with MaSp1 and MaSp2. Perhaps MaSp1s plays a role similar to p25 in the silkworm. Further research is needed to elucidate the function of MaSp1s.

Secondly, in the Chapter I paragraph II, the *CySp2* gene of Wasp Spider *Argiope bruennichi* was reanalyzed. The reason is that to gain further understanding of egg case silk proteins gene family, A. C. Zhao et al. (2006) isolated two full-length cDNAs for egg case silk proteins, cylindrical silk protein 1 (CySp1) and cylindrical silk protein 2 (CySp2), from the wasp spider, *Argiope bruennichi*. AbCySp2 was reported to contain no apparent signal peptide sequences, and the AbCySp1-AbCySp2 complex, which would possess a signal peptide, would be transported across the endoplasmic reticulum and secreted to the Golgi. According to a report by Hayashi, genomic DNA sequencing is one approach that can be successfully utilized to retrieve 5' ends of silk genes (66); using this method, the 5' end of *CySp2* was retrieved. The result shown that CySp2 contained a typical signal peptide similar to that found in CySp1; thus, due to technical limitations, an artificial error had occurred in the CySp2 sequence reported by Zhao et al. (2006). And what's more, the mechanism which lead to the error had happen have been elucidated. The excision event occurred during cDNA library construction. When the double-strand DNA fragments produced by RT-PCR were linked into a plasmid and transformed into *E.coli*, an intermolecular homologous recombination occurred during the plasmid self-replicating process. And, correct information of AbCySp2 obtained here will provide a theoretical basis for the gene's industry application in the future.

Chapter II was divided into three sections. In the first section (Chapter II paragraph I), the expression of a spider dragline gene: *Cm.MaSp1s* by germline transformation of the silkworm *Bombyx mori* using the *piggyBac* transposon-derived vector with the *B. mori* FibH promoter was further attempted. the repetitive unit of *MaSp1s* was cloned. And it was inserted into the *piggyBac* transposon-derived vector. In order to simplify the screening steps, the N4 white eye strain was chosen. The G1 eggs were screened by detecting DsRed fluorescence under the fluorescent stereomicroscope. And the other three stages (larvae, pupa and imago.) were also viewed with DsRed fluorescence and visible light. G1 positive transgenic individuals were mated within the same family to generate the G2 descendents for the following analysis.

Although, many other researchers have done a lot work about the transgenic silkworm, there has been very limited information on the biological properties of silkworm carried the foreign silk protein. So in the second section (Chapter II paragraph II), the relationship between the foreign gene and three major components of silk fibroin (FibH, FibL and p25) was analyzed, systematically. At the DNA level, the iPCR result showed that foreign gene *MaSp1s* inserted site located in scaffold868. At the RNA level, the Real-time PCR results showed that there was not significant difference between the wild type silkworm (WT) and transgenic silkworm (TG) on the transcriptional level of the three genes (*FibH*, *FibL* and *p25*), particularly the *FibH*. It means that the expression of the foreign gene *MaSp1s* didn't affect three major genes expression at RNA level. At the protein level, the SDS-PAGE data showed that the abundance the foreign protein (about 2% among the total fibroin proteins) is much lower than that of FibH and FibL, but richer than p25. At last, the Co-IP (Co-immunoprecipitation) was used for detecting the interaction between

MaSp1s and FibL. The results showed that the MaSp1s and FibL did form the protein complex as expected. And the fibroin complex also contains the FibH.

In the last section (Chapter II paragraph III), the physical properties and structure of silk were analyzed from the transgenic silkworm. The reason is that, to date, there is little report about the physical properties of the transgenic silk including spider dragline silk protein. To study the properties and structure of the silkworm silk including spider silk protein, the silk was characterized by Micro tensile testing, Fourier transform infrared spectroscopy (FT-IR), Wide angle X-ray diffraction (WAXD), and Differential scanning calorimetry (DSC). The tensile testing data showed that the silk's average tensile strength is 340 MPa and its strain is about 45%. FT-IR measurements suggested the existence of binary structure containing random coil conformation and  $\beta$ -sheet structure. DSC result indicated that the melting temperature of the silk is around 326 °C. The molecular orientation calculated from WAXD pattern is 0.9403. The results obtained are providing some clue for understanding the functional role of the small silk gene (*Cm.MaSp1s*).

In conclusion, in this study, I focus on two research topics, one is spider silk genes, and the other is silkworm including spider silk gene. I hope that those results presented here would be a little helpful to this research field.

## Reference

1. Prosdocimi, F., Bittencourt, D., da Silva, F. R., Kirst, M., Motta, P. C., and Rech, E. L. (2011) Spinning gland transcriptomics from two main clades of spiders (order: Araneae)--insights on their molecular, anatomical and behavioral evolution, *PloS one* 6, e21634.
2. Hinman, M. B., Jones, J. A., and Lewis, R. V. (2000) Synthetic spider silk: a modular fiber, *Trends in biotechnology* 18, 374-379.
3. Shultz, J. W. (1987) THE ORIGIN OF THE SPINNING APPARATUS IN SPIDERS, *Biological Reviews* 62, 89-113.
4. La Mattina, C., Reza, R., Hu, X., Falick, A. M., Vasanthavada, K., McNary, S., Yee, R., and Vierra, C. A. (2008) Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver *Latrodectus hesperus*, *Biochemistry* 47, 4692-4700.
5. Romer, L., and Scheibel, T. (2008) The elaborate structure of spider silk: structure and function of a natural high performance fiber, *Prion* 2, 154-161.
6. Hu, X., Kohler, K., Falick, A. M., Moore, A. M., Jones, P. R., Sparkman, O. D., and Vierra, C. (2005) Egg case protein-1. A new class of silk proteins with fibroin-like properties from the spider *Latrodectus hesperus*, *The Journal of biological chemistry* 280, 21220-21230.
7. Hayashi, C. Y., Blackledge, T. A., and Lewis, R. V. (2004) Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family, *Mol Biol Evol* 21, 1950-1959.
8. Knoflach, B. (1998) Mating in Theridion varians Hahn and related species (Araneae: Theridiidae), *Journal of Natural History* 32, 545-604.
9. Seah, W. K., and Li, D. (2001) Stabilimenta attract unwelcome predators to orb-webs, *Proceedings. Biological sciences / The Royal Society* 268, 1553-1558.
10. Gosline, J. M., DeMont, M. E., and Denny, M. W. (1986) The structure and properties of spider silk, *Endeavour* 10, 37-43.
11. Geurts, P., Zhao, L., Hsia, Y., Gnesa, E., Tang, S., Jeffery, F., Mattina, C. L., Franz, A., Larkin, L., and Vierra, C. (2010) Synthetic Spider Silk Fibers Spun

- from Pyriform Spidroin 2, A Glue Silk Protein Discovered in Orb-Weaving Spider Attachment Discs<sup>†</sup>, *Biomacromolecules* 11, 3495-3503.
12. Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A., and Hayashi, C. Y. (2007) Blueprint for a high-performance biomaterial: full-length spider dragline silk genes, *PLoS One* 2, e514.
13. Griffiths, J. R., and Salanitri, V. R. (1980) The strength of spider silk, *J Mater Sci* 15, 491-496.
14. Termonia, Y. (1994) Molecular Modeling of Spider Silk Elasticity, *Macromolecules* 27, 7378-7381.
15. Holland, G. P., Lewis, R. V., and Yarger, J. L. (2004) WISE NMR characterization of nanoscale heterogeneity and mobility in supercontracted *Nephila clavipes* spider dragline silk, *Journal of the American Chemical Society* 126, 5867-5872.
16. Guerette, P. A., Ginzinger, D. G., Weber, B. H., and Gosline, J. M. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science* 272, 112-115.
17. Xu, M., and Lewis, R. V. (1990) Structure of a protein superfiber: spider dragline silk, *Proceedings of the National Academy of Sciences of the United States of America* 87, 7120-7124.
18. Hinman, M. B., and Lewis, R. V. (1992) Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber, *The Journal of biological chemistry* 267, 19320-19324.
19. Jelinski, L. W., Blye, A., Liivak, O., Michal, C., LaVerde, G., Seidel, A., Shah, N., and Yang, Z. (1999) Orientation, structure, wet-spinning, and molecular basis for supercontraction of spider dragline silk, *International journal of biological macromolecules* 24, 197-201.
20. Sponner, A., Unger, E., Grosse, F., and Weisshart, K. (2005) Differential polymerization of the two main protein components of dragline silk during fibre spinning, *Nature materials* 4, 772-775.
21. Simmons, A., Ray, E., and Jelinski, L. W. (1994) Solid-State <sup>13</sup>C NMR of *Nephila clavipes* Dragline Silk Establishes Structure and Identity of Crystalline Regions, *Macromolecules* 27, 5235-5237.
22. Craig, C. L., Riekel, C., Herberstein, M. E., Weber, R. S., Kaplan, D., and Pierce, N. E. (2000) Evidence for diet effects on the composition of silk proteins produced by spiders, *Molecular biology and evolution* 17, 1904-1913.
23. Simmons, A. H., Michal, C. A., and Jelinski, L. W. (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk, *Science* 271, 84-87.

24. Ulrich, S., Glisovic, A., Salditt, T., and Zippelius, A. (2008) Diffraction from the beta-sheet crystallites in spider silk, *The European physical journal. E, Soft matter* 27, 229-242.
25. Dong, Z., Lewis, R. V., and Middaugh, C. R. (1991) Molecular mechanism of spider silk elasticity, *Archives of Biochemistry and Biophysics* 284, 53-57.
26. Kümmerlen, J., van Beek, J. D., Vollrath, F., and Meier, B. H. (1996) Local Structure in Spider Dragline Silk Investigated by Two-Dimensional Spin-Diffusion Nuclear Magnetic Resonance†, *Macromolecules* 29, 2920-2928.
27. Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins, *International journal of biological macromolecules* 24, 271-275.
28. Hayashi, C. Y., and Lewis, R. V. (1998) Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks, *J Mol Biol* 275, 773-784.
29. van Beek, J. D., Hess, S., Vollrath, F., and 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, 10266-10271.
30. O'Brien, J. P., Fahnestock, S. R., Termonia, Y., and Gardner, K. H. (1998) Nylons from Nature: Synthetic Analogs to Spider Silk, *Advanced Materials* 10, 1185-1195.
31. Fahnestock, S. R., and Irwin, S. L. (1997) Synthetic spider dragline silk proteins and their production in Escherichia coli, *Appl Microbiol Biotechnol* 47, 23-32.
32. Fahnestock, S. R., and Bedzyk, L. A. (1997) Production of synthetic spider dragline silk protein in Pichia pastoris, *Appl Microbiol Biotechnol* 47, 33-39.
33. Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J. F., Duguay, F., Chretien, N., Welsh, E. A., Soares, J. W., and Karatzas, C. N. (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells, *Science* 295, 472-476.
34. Scheller, J., Guhrs, K. H., Grosse, F., and Conrad, U. (2001) Production of spider silk proteins in tobacco and potato, *Nat Biotechnol* 19, 573-577.
35. Teule, F., Cooper, A. R., Furin, W. A., Bittencourt, D., Rech, E. L., Brooks, A., and Lewis, R. V. (2009) A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning, *Nat Protoc* 4, 341-355.



36. Wen, H., Lan, X., Zhang, Y., Zhao, T., Wang, Y., Kajiura, Z., and Nakagaki, M. (2010) Transgenic silkworms (*Bombyx mori*) produce recombinant spider dragline silk in cocoons, *Mol Biol Rep* 37, 1815-1821.
37. Tamura, T., Thibert, C., Royer, C., Kanda, T., Eappen, A., Kamba, M., Komoto, N., Thomas, J.-L., Mauchamp, B., Chavancy, G., Shirk, P., Fraser, M., Prudhomme, J.-C., and Couble, P. (2000) Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector, *Nat Biotech* 18, 81-84.
38. Adachi, T., Tomita, M., Shimizu, K., Ogawa, S., and Yoshizato, K. (2006) Generation of hybrid transgenic silkworms that express *Bombyx mori* prolyl-hydroxylase alpha-subunits and human collagens in posterior silk glands: Production of cocoons that contained collagens with hydroxylated proline residues, *Journal of biotechnology* 126, 205-219.
39. Fujii, T., Daimon, T., Uchino, K., Banno, Y., Katsuma, S., Sezutsu, H., Tamura, T., and Shimada, T. (2010) Transgenic analysis of the BmBLOS2 gene that governs the translucency of the larval integument of the silkworm, *Bombyx mori*, *Insect molecular biology* 19, 659-667.
40. Guo, X. Y., Dong, L., Wang, S. P., Guo, T. Q., Wang, J. Y., and Lu, C. D. (2004) Introduction of foreign genes into silkworm eggs by electroporation and its application in transgenic vector test, *Acta biochimica et biophysica Sinica* 36, 323-330.
41. Iizuka, M., Ogawa, S., Takeuchi, A., Nakakita, S., Kubo, Y., Miyawaki, Y., Hirabayashi, J., and Tomita, M. (2009) Production of a recombinant mouse monoclonal antibody in transgenic silkworm cocoons, *The FEBS journal* 276, 5806-5820.
42. Imamura, M., Nakahara, Y., Kanda, T., Tamura, T., and Taniai, K. (2006) A transgenic silkworm expressing the immune-inducible cecropin B-GFP reporter gene, *Insect biochemistry and molecular biology* 36, 429-434.
43. Imamura, M., Nakai, J., Inoue, S., Quan, G. X., Kanda, T., and Tamura, T. (2003) Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*, *Genetics* 165, 1329-1340.
44. Inoue, S., Kanda, T., Imamura, M., Quan, G. X., Kojima, K., Tanaka, H., Tomita, M., Hino, R., Yoshizato, K., Mizuno, S., and Tamura, T. (2005) A fibroin secretion-deficient silkworm mutant, Nd-sD, provides an efficient system for producing recombinant proteins, *Insect biochemistry and molecular biology* 35, 51-59.

45. Kato, T., Kajikawa, M., Maenaka, K., and Park, E. Y. (2010) Silkworm expression system as a platform technology in life science, *Applied microbiology and biotechnology* 85, 459-470.
46. Kojima, K., Kuwana, Y., Sezutsu, H., Kobayashi, I., Uchino, K., Tamura, T., and Tamada, Y. (2007) A new method for the modification of fibroin heavy chain protein in the transgenic silkworm, *Bioscience, biotechnology, and biochemistry* 71, 2943-2951.
47. Tomarev, S. I., Callaerts, P., Kos, L., Zinovieva, R., Halder, G., Gehring, W., and Piatigorsky, J. (1997) Squid Pax-6 and eye development, *Proceedings of the National Academy of Sciences of the United States of America* 94, 2421-2426.
48. Tomita, M. (2011) Transgenic silkworms that weave recombinant proteins into silk cocoons, *Biotechnology letters* 33, 645-654.
49. Tomita, M., Hino, R., Ogawa, S., Iizuka, M., Adachi, T., Shimizu, K., Sotoshiro, H., and Yoshizato, K. (2007) A germline transgenic silkworm that secretes recombinant proteins in the sericin layer of cocoon, *Transgenic research* 16, 449-465.
50. Tomita, M., Munetsuna, H., Sato, T., Adachi, T., Hino, R., Hayashi, M., Shimizu, K., Nakamura, N., Tamura, T., and Yoshizato, K. (2003) Transgenic silkworms produce recombinant human type III procollagen in cocoons, *Nature biotechnology* 21, 52-56.
51. Zhao, A., Zhao, T., Zhang, Y., Xia, Q., Lu, C., Zhou, Z., Xiang, Z., and Nakagaki, M. (2010) New and highly efficient expression systems for expressing selectively foreign protein in the silk glands of transgenic silkworm, *Transgenic research* 19, 29-44.
52. Teule, F., Miao, Y. G., Sohn, B. H., Kim, Y. S., Hull, J. J., Fraser, M. J., Jr., Lewis, R. V., and Jarvis, D. L. (2012) Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties, *Proceedings of the National Academy of Sciences of the United States of America* 109, 923-928.
53. Zhang, Y., Hu, J., Miao, Y., Zhao, A., Zhao, T., Wu, D., Liang, L., Miikura, A., Shiomi, K., Kajiura, Z., and Nakagaki, M. (2008) Expression of EGFP-spider dragline silk fusion protein in BmN cells and larvae of silkworm showed the solubility is primary limit for dragline proteins yield, *Molecular biology reports* 35, 329-335.
54. Zhao, A., Zhao, T., Sima, Y., Zhang, Y., Nakagaki, K., Miao, Y., Shiomi, K., Kajiura, Z., Nagata, Y., and Nakagaki, M. (2005) Unique molecular

- architecture of egg case silk protein in a spider, *Nephila clavata*, *J Biochem* 138, 593-604.
55. Zhao, A. C., Zhao, T. F., Nakagaki, K., Zhang, Y. S., Sima, Y. H., Miao, Y. G., Shiomi, K., Kajiura, Z., Nagata, Y., Takadera, M., and Nakagaki, M. (2006) Novel molecular and mechanical properties of egg case silk from wasp spider, *Argiope bruennichi*, *Biochemistry* 45, 3348-3356.
56. Miao, Y., Zhang, Y., Nakagaki, K., Zhao, T., Zhao, A., Meng, Y., Nakagaki, M., Park, E. Y., and Maenaka, K. (2006) Expression of spider flagelliform silk protein in *Bombyx mori* cell line by a novel Bac-to-Bac/BmNPV baculovirus expression system, *Applied microbiology and biotechnology* 71, 192-199.
57. Rising, A., Nimmervoll, H., Grip, S., Fernandez-Arias, A., Storckenfeldt, E., Knight, D. P., Vollrath, F., and Engstrom, W. (2005) Spider silk proteins--mechanical property and gene sequence, *Zoolog Sci* 22, 273-281.
58. Hu, X., Vasanthavada, K., Kohler, K., McNary, S., Moore, A. M., and Vierra, C. A. (2006) Molecular mechanisms of spider silk, *Cell Mol Life Sci* 63, 1986-1999.
59. Lewis, R. V., Hinman, M., Kothakota, S., and Fournier, M. J. (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins, *Protein Expr Purif* 7, 400-406.
60. Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins, *Int J Biol Macromol* 24, 271-275.
61. Sponner, A., Schlott, B., Vollrath, F., Unger, E., Grosse, F., and Weisshart, K. (2005) Characterization of the protein components of *Nephila clavipes* dragline silk, *Biochemistry* 44, 4727-4736.
62. Beuken, E., Vink, C., and Bruggeman, C. A. (1998) One-step procedure for screening recombinant plasmids by size, *Biotechniques* 24, 748-+.
63. Ochman, H., Gerber, A. S., and Hartl, D. L. (1988) Genetic applications of an inverse polymerase chain reaction, *Genetics* 120, 621-623.
64. Garb, J. E., Ayoub, N. A., and Hayashi, C. Y. (2010) Untangling spider silk evolution with spidroin terminal domains, *Bmc Evolutionary Biology* 10.
65. Rising, A., Hjalms, G., Engstrom, W., and Johansson, J. (2006) N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins, *Biomacromolecules* 7, 3120-3124.
66. Motriuk-Smith, D., Smith, A., Hayashi, C. Y., and Lewis, R. V. (2005) Analysis of the conserved N-terminal domains in major ampullate spider silk proteins, *Biomacromolecules* 6, 3152-3159.

67. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature methods* 8, 785-786.
68. Elias, S., Delestre, C., Courel, M., Anouar, Y., and Montero-Hadjadje, M. (2010) Chromogranin A as a crucial factor in the sorting of peptide hormones to secretory granules, *Cellular and molecular neurobiology* 30, 1189-1195.
69. Garb, J. E., Ayoub, N. A., and Hayashi, C. Y. (2010) Untangling spider silk evolution with spidroin terminal domains, *BMC evolutionary biology* 10, 243.
70. Eisoldt, L., Thamm, C., and Scheibel, T. (2012) Review the role of terminal domains during storage and assembly of spider silk proteins, *Biopolymers* 97, 355-361.
71. Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K., and Mizuno, S. (2000) Silk Fibroin of Bombyx mori Is Secreted, Assembling a High Molecular Mass Elementary Unit Consisting of H-chain, L-chain, and P25, with a 6:6:1 Molar Ratio, *Journal of Biological Chemistry* 275, 40517-40528.
72. Han, L., Zhang, L., Zhao, T., Wang, Y., and Nakagaki, M. (2013) Analysis of a new type of major ampullate spider silk gene, MaSp1s, *Int J Biol Macromol* 56, 156-161.
73. Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M., and Miller, W. (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence, *Genome research* 8, 967-974.
74. Hayashi, C. Y., and Lewis, R. V. (2000) Molecular architecture and evolution of a modular spider silk protein gene, *Science* 287, 1477-1479.
75. Inoue, S., Tanaka, K., Tanaka, H., Ohtomo, K., Kanda, T., Imamura, M., Quan, G. X., Kojima, K., Yamashita, T., Nakajima, T., Taira, H., Tamura, T., and Mizuno, S. (2004) Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of alpha1,2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25, *European journal of biochemistry / FEBS* 271, 356-366.
76. Agnarsson, I., Kuntner, M., and Blackledge, T. A. (2010) Bioprospecting finds the toughest biological material: extraordinary silk from a giant riverine orb spider, *PloS one* 5, e11234.
77. Callaerts, P., Halder, G., and Gehring, W. J. (1997) PAX-6 in development and evolution, *Annual review of neuroscience* 20, 483-532.
78. Thomas, J. L., Da Rocha, M., Besse, A., Mauchamp, B., and Chavancy, G. (2002) 3xP3-EGFP marker facilitates screening for transgenic silkworm Bombyx mori L. from the embryonic stage onwards, *Insect biochemistry and molecular biology* 32, 247-253.

79. Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., Shen, Y., Lan, X., Yuan, L., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen, W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R., Shi, J., Li, H., Su, J., Wang, X., Zhang, Z., Wu, Q., Li, J., Zhang, Q., Wei, N., Sun, H., Dong, L., Liu, D., Zhao, S., Zhao, X., Meng, Q., Lan, F., Huang, X., Li, Y., Fang, L., Li, D., Sun, Y., Yang, Z., Huang, Y., Xi, Y., Qi, Q., He, D., Huang, H., Zhang, X., Wang, Z., Li, W., Cao, Y., Yu, Y., Yu, H., Ye, J., Chen, H., Zhou, Y., Liu, B., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Wang, W., Ye, C., Wong, G. K., and Yang, H. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*), *Science* 306, 1937-1940.
80. Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z., and Xia, Q. (2010) SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology, *Nucleic acids research* 38, D453-456.
81. Tanaka, K., Inoue, S., and Mizuno, S. (1999) Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori*, *Insect biochemistry and molecular biology* 29, 269-276.
82. Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kikuchi, A., Ohtomo, K., Takagi, T., and Mizuno, S. (1999) Determination of the site of disulfide linkage between heavy and light chains of silk fibroin produced by *Bombyx mori*, *Biochimica et biophysica acta* 1432, 92-103.
83. Inoue, S., Tanaka, K., Tanaka, H., Ohtomo, K., Kanda, T., Imamura, M., Quan, G.-X., Kojima, K., Yamashita, T., Nakajima, T., Taira, H., Tamura, T., and Mizuno, S. (2004) Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of  $\alpha$ 1,2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25, *European Journal of Biochemistry* 271, 356-366.
84. Scheibel, T. (2004) Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins, *Microbial cell factories* 3, 14.
85. Shao, Z., and Vollrath, F. (2002) Surprising strength of silkworm silk, *Nature* 418, 741.
86. Wei, K., Ohta, T., Kim, B.-S., Kim, K.-W., Lee, K.-H., Khil, M. S., Kim, H.-Y., and Kim, I.-S. (2010) Development of electrospun metallic hybrid nanofibers via metallization, *Polymers for Advanced Technologies* 21, 746-751.

87. Tsukada, M., Khan, M. M., Inoue, E., Kimura, G., Hun, J. Y., Mishima, M., and Hirabayashi, K. (2010) Physical properties and structure of aquatic silk fiber from *Stenopsyche marmorata*, *International journal of biological macromolecules* 46, 54-58.
88. Du, N., Liu, X. Y., Narayanan, J., Li, L., Lim, M. L., and Li, D. (2006) Design of superior spider silk: from nanostructure to mechanical properties, *Biophysical journal* 91, 4528-4535.