Characterization of a Cysteine-Rich Protein Specifically Expressed in the Silk Gland of Caddisfly *Stenopsyche marmorata* (Trichoptera; Stenopsychoidea)

Yujun Wang, Hong Wang, Tianfu Zhao, and Masao Nakagaki

Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda 386-8567, Japan

Received August 13, 2009; Accepted October 21, 2009; Online Publication, January 7, 2010

A novel protein, Smsp-72k, was found to be selectively expressed in the silk gland of aquatic larvae of the Stenopsyched caddisfly (*Stenopsyche marmorata*). The protein was characterized by an abundance of cysteine (13.97%) and charged residues (47.21%). Amino acids with hydroxyl side-chains accounted for an additional 10% of the Smsp-72k protein, with serine at 4.4% and threonine at 5.6%. A cysteine-rich repetitive sequence is common to many potential and known underwater adhesive/cement proteins and cell-cell adhesion molecules. We hypothesized that Smsp-72k is an adhesive/cement protein that increases the adhesiveness of the silk fiber of *S. marmorata*. The hydroxyl groups of Smsp-72k might form a link with the heavy chain fibroin of *S. marmorata*, removing the weak boundary-water layer and allowing the spreading of the silk protein onto the surface of the substratum during the process of adhesion.

**Key words:** cysteine-rich; caddisfly; silk; adhesive protein

The Trichoptera (caddisfly) species, which belongs to one of the major orders of aquatic insects, are found in a wide variety of freshwater and marine habitats. The larvae are similar to those of the sister order Lepidoptera with respect to ability to spin silk and to a pair of labial glands, and the two orders evolved to form the superorder Amphiesmenoptera. Special attention has been placed on the silk of the caddisfly because it is not only able to keep its tensile strength under the long exposure to water, but can also adhere irreversibly to various surfaces in an aquatic environment.

Heavy-chain fibroin (H-fibroin) and light-chain fibroin (L-fibroin), the major components of the lepidopteron filament, have been found to be conserved in caddisfly by a screen of silk gland specific cDNA libraries. Compared to Lepidoptera, the H-fibroin of caddisfly has a higher molecular mass (>400 kDa), is rich in charged amino acids, and is predominantly hydrophilic. On the other hand, the L-fibroins of the caddisfly and Lepidoptera are similar in terms of their sizes, hydrophobicity, and content of charged residues. The amphiphilic nature of H-fibroin and its high content of charged amino acids might facilitate the secretion and storage of a covalently linked L-fibroin/H-fibroin dimer during the process of silk formation within the silk gland. This reflects the close evolutionary relationship between Trichoptera and Lepidoptera. The silk proteins of Trichoptera, however, have not yet been found to be homologous to any proteins from the other aquatic insect orders, such as the midge larvae (Diptera: Chironomus). Considering that aqueous silk proteins, including some adhesive or cement proteins that are secreted by some marine organisms, have similar functions, the aim of our study was to determine if any of the proteins in the silk of Trichoptera have features that resemble other aquatic silk or adhesive/cement.

The larvae of *Stenopsyche marmorata* use their silk to build cases on small pebbles, sand, and other small debris and to create nets that attach securely to the tops of large rocks. In the present study we identified a novel gene, Smsp-72k, that is specifically expressed in the silk gland of *S. marmorata*. Smsp-72k is characterized by a high content of cysteine and charged residues. Our data indicate that Smsp-72k is probably an adhesive protein within the silk thread of the caddisfly.

**Materials and Methods**

**Differential interference contrast (DIC) microscope.** The physical structure of the silk of *S. marmorata* was observed with a Differential Interference Contrast (DIC) microscope (Carl Zeiss, Jena, Germany) at various magnifications. DIC microscopy was performed under ambient temperature and humidity, which ranged from 23 to 28°C and 30–42%.

**Insects and silk gland preparation.** *S. marmorata* (Annulipalpia suborder, Stenopsychidae family) larvae were collected in a shallow area of the Chikuma River in Ueda, Nagano Prefecture, Japan, from April to July on 2007. The majority of the 5th instar larvae were promptly dissected under a dissection microscope to harvest the silk glands. The rest of the harvested silk glands were used in total RNA isolation.

**SDS–PAGE and protein sequence.** The extracted proteins were dissolved in sample buffer containing 2% SDS and 5% β-mercaptoethanol, heated for 5 min in a boiling water bath, and centrifuged for 3 min at 10,000 rpm. The prepared samples were loaded onto polyacrylamide gels (4% stacking gel, 8% separating gel). Electrophoresis was performed under denaturing conditions. The gel was stained with Coomassie brilliant blue R-250 or electronically transferred onto a polyvinylidene difluoride (PVDF) membrane (Millipore, Massachusetts, America) for protein sequencing. The transfer was

---

1 To whom correspondence should be addressed. Tel: +81-268-21-5335; Fax: +81-268-21-5338; E-mail: nakagak@iptc.shinshu-u.ac.jp
RNA isolation and cDNA library construction. Total RNA was extracted from the silk glands using TRIzol following manufacturer’s instructions (Invitrogen, Tokyo). Ten pairs of silk glands were dissected from euthanized *S. marmorata* and then rapidly homogenized in 1 ml of TRIzol. Genomic DNA and proteins were removed by partitioning with 200 µl of chloroform, and RNA was precipitated from the remaining aqueous phase with an equal volume of isopropanol. The precipitate was rinsed with 75% ethanol, dried at room temperature, and dissolved in 0.5% SDS with 20 mm sodium acetate (pH 5.3). The RNAs were used to create a silk gland-specific cDNA library, which was constructed commercially by Hitachi Instrument Service (Tokyo). Electro-MAX DH12S cells (Invitrogen) were transformed by electroporation, and more than 500 recombinant plasmids were randomly chosen and purified by alkaline lysis with SDS. The inserts were sequenced using an ABI Prism Genetic Analyzer 3100 and T7 primer. Computer analyses of DNA and amino acid sequences were done using the Genetyx package (Windows version, Genetyx, Tokyo) and Sequencer 4.1.4 (Demo version) respectively. Comparisons were performed by BLASTX homology search against protein databases that were available on the ExPASy Proteomics server (http://www.expasy.org).

**Northern blotting.** Aliquots of about 5 µg of total RNA were prepared from the silk glands of 15 insects. Total RNA and RNA markers were electrophoresed on a 1% agarose gel (1 × MOPS, 0.66 m formaldehyde) and then blotted. Using the hybridization conditions from the manual for the DIG nucleic acid detection kit (Boehringer Mannheim, Mannheim, Germany), we probed the Northern blots with a Digoxygenin-labeled antisense RNA probe that corresponded to 240 nucleotides of double-stranded flat ribbons, each composed of filaments with diameters of 6–10 µm, that were often organized into twisted bundles (Fig. 1). The terminus of the silk, which functions to adhere to a substance, was transformed (Fig. 1C) in order to enhance the interface area.

**Isolation and sequencing of silk gland cDNA clones**

A silk gland-specific cDNA library was constructed from final instar larvae of *S. marmorata*, and ESTs were obtained by sequencing the 5′-termini of insert fragments (>1 kb) of individual clones. Of a total of 260 tested ESTs, 72 (27.69%) represented single occurrences, while the remaining 148 ESTs were grouped into dozens of clusters, under the assumption that each cluster of overlapping ESTs represented a single gene. Inserts with open reading frames of more than 300 base pairs were subjected to a BLASTX homology search against protein databases that were available on the ExPASy Proteomics server (http://www.expasy.org). The largest cluster, which contained 39 ESTs, was further confirmed to be the major filament component, a homolog of H-fibroin cDNA. A cluster of five ESTs was confirmed to be a homolog of L-fibroin cDNA (data not show).

We then focused on a cluster containing eight ESTs, which are represented within clone TBS1A02. Sequence analysis of the entire TBS1A02 clone revealed the presence of a gene with a complete ORF. The gene was designated *SmSp-72k*, for *S. marmorata* silk protein, and had a molecular mass of 72 kDa.

*SmSp-72k* was selectively expressed in the silk gland

The expression pattern of a particular gene is important in order to speculate as to its function. According to morphology and functionality, previous studies have divided the silk gland (SG) of *B. mori* into three sections: the anterior section (ASG), the middle section (MSG), and the posterior section (PSG). Fibroins are known to be secreted from the PSG, while sericins are known to be selectively produced by the MSG. By analogy with Lepidoptera, the SG of *S. marmorata* was classified into the same three sections based upon similar criteria, although its posterior silk glands were not well developed. Northern Blotting
analysis, performed to examine the mRNA expression profile of Smsp-72k, revealed that Smsp-72k is selectively expressed in MSG and PSG as a single transcript of about 2–3 kb (Fig. 2). An approximately 70 kDa protein separated by SDS–PAGE from the silk gland lumen of S. marmorata was transferred onto a PVDF membrane for N-terminal sequencing. The presence of the approximately 70 kDa protein appeared as a strong band on SDS–PAGE (Fig. 3). N-terminal sequencing of the isolated protein gave a sequence of SSXGXKKEXA, where X represents an unidentified amino acid. The sequence was closely matched to the 20–29 residues of the deduced protein, Smsp-72k, in which the first 19 residues were predicted to be a signal peptide, as described below (Fig. 4). The matching sequence and similar molecular weight indicates that the approximately 70 kDa protein isolated from the silk gland lumen was encoded by the gene of Smsp-72k.

Characterization of Smsp-72k

The Smsp-72k gene, which is comprised of 2,040 base pairs, encodes a 680 amino acid open reading frame with a signal sequence that encompasses the first 19 residues (Fig. 2). A BLASTX homology search revealed that Smsp-72k has 23% identity and 34% similarity to the notch-like protein of B. mori (Genbank accession no. DQ311297), indicating that it contains an epidermal growth factor (EGF)-like domain. Structurally, Smsp-72k is composed of a repetitive sequence element of 53–55 amino acids that is tandemly repeated 12 times. Each element contains 6–8 conserved cysteines and several other charged residues, which results in a high cysteine content for the entire protein (Fig. 4). Secretory signals are important elements of silk-associated proteins, such as those found in the silk of spiders and silkworms, since these proteins must be transported across the endoplasmic reticulum and secreted. SignaLP software (euKaryotic option) analysis predicted a cleavage site between positions 19 and 20 (TFA-S), which indicated that the hydrophobic N-terminal region might function as a signal sequence. This suggests that Smsp-72k can be secreted into the gland lumen.

Discussion

The existence of adhesive proteins in the silk filament of the caddisfly is predicted by the fact that the filament strongly adheres to the surfaces of many different materials, including rock, wood, leaf, and glass, underwater. Hatano (2006) observed that the filament of S. marmorata can generally be divided into a major inner filamentous region and a peripheral layer, based on the electronic density, as determined by Transmission Electron Microscope. The repetitive region contains 12 repeated units, each of which is composed of 53–55 residues and includes 6–8 conserved cysteines. The conserved residues in each repeat are highlighted in gray, and additional asterisks indicate the conserved cysteines. The italicized sequence was found to be the signal peptide. Residues are numbered from the initiating Met (M).

The repetitive structure and specific expression in the silk gland of Smsp-72k indicates that it is a component of caddisfly silk thread. We assume that Smsp-72k exists in the peripheral layer and serves an adhesive function, considering that the high content of cysteine residues and charged amino acids in this protein is shared by many potential and known adhesive proteins that are secreted underwater (Table 1). The Chironomus species (a midge) also spins silken threads under water to construct feeding and pupation tubes. Secretory proteins (SPs), such as the protein Sp220, have been identified as silk components in the midge larvae of Chironomus tentans and are known to be composed of repetitive elements that are rich in cysteines and/or charged residues. Although the precise nature of SPs remains uncertain, especially with
regard to which of the proteins have the function of causing the tube-constructing materials to adhere, some SPs have been predicted to act as a glue-like substance that confers adhesiveness on the silk. The unusually high cysteine content of *C. tentans* has been to confer on proteins with ability to form intermolecular disulfide bonds that contribute to fiber stability and insolubility underwater.\(^{10,11}\) A similar observation holds for Tcp (Thermzyon Cocoon Protein), the major component of the cocoon of the aquatic leech *Thermzyon rube*.

Tcp is an antistan-notch-like fusion protein that contains six tandem repeats, each comprised of 12 ordered cysteine residues in an approximately 62 amino acid repeating unit.\(^{12}\) The notch-like domains of Tcp play critical roles in development, signaling, and adhesion throughout Animalia.\(^{12}\) Nevertheless, a fibropellin-based fiber, which forms a protective layer for the sea urchin (*Strongylocentrotus purpuratus*) embryo, bears a structural resemblance to the leech cocoon membrane.\(^{12-15}\) Furthermore, *Mytilus galloprovincialis* foot protein 2 (Mgfp2), a component of the adhesive plaque of the mussel, is a cysteine-rich protein that contains EGF-like domains.\(^{16}\) Similarly, the *Megabalanus rose* cement protein (Mrcp-20k) of the barnacle, which also has a high content of cysteines and charged amino acids, has a specific role in attachments to rocks and other surfaces underwater.\(^{17,18}\) A comparison of the sequences mentioned above indicates that these proteins are related mainly by their cysteine-rich nature and that they show little sequence similarity overall. However, all of these proteins, from a variety of species, possess a potential adhesive function, which suggests that cysteine is an inherently favorable amino acid for the property of adhesiveness. Further evidence of the adhesive function of cysteine has been found in many cell adhesion proteins, since they can promote cell adhesion through their cysteine-rich domains.\(^{19-23}\)

Cysteine is assumed to play a role in maintaining the topology of charged amino acids on the molecular surface by the formation of intramolecular disulfide bonds.\(^{17}\) The unusually high cysteine content of SmSp-72k implies the might formation of intermolecular disulfide cross-links both within itself and to other proteins. The fact that SmSp-72k is predominantly hydrophilic poses a potential problem with regard to the secretion of this protein in water, for it risks dilution in an aqueous environment. We hypothesize that the inter-molecular cross-links formed by the disulfide bonds render hydrophilic SmSp-72k insoluble. The L-chain fibroin of *S. marmorata* was found to contain nine cysteines in its 247 amino acids (data not shown) and there were possible inter-molecular cross-links between SmSp-72k and L-fibroin by their cysteines, although this hypothesis requires further examination.

Amino acids with hydroxyl side-chains account for 10% of SmSp-72k protein, serine accounting for 4.4% and threonine for 5.74%. Chen and Cyr (1970) noted that hydrogen-bonding is especially important in wet adhesives,\(^{24}\) and Waite (1987) suggested that the hydroxyl groups of the *Mytilus byssus* protein play a role in removing the weak boundary-water layer so that it can be spread onto the surface of the substratum during the process of adhesion.\(^{25}\) The hydroxylated side-chains of SmSp-72k might play a role in underwater adhesion similar to those of the *Mytilus byssus* protein. The filament region of the silk is composed mainly of H-fibroin and L-fibroin, like the silk of Lepidoptera. All the examined H-fibroins had Arg as one of their most abundant residues, differing from their counterparts in Lepidoptera. Engster (1976) indicated that the hydroxyl groups of Ser and Thr could hydrogen-bond with the guanidium groups of Arg, and that the hydroxyl groups of Ser could also form ester links with the Asp or Glu of fibroin.\(^{26}\) Therefore, these bonds provide a potential mechanism through which the peripheral layer protein SmSp-72k can be linked directly to the filament region protein H-fibroin, which explains the physical properties of the aqueous silk of the caddisfly.

In conclusion, the major characteristics of the SmSp-72k protein, including the high content of cysteines and of charged residues and the repetitive structure, are shared with many underwater adhesive proteins and cell-cell adhesion molecules. We suggest that SmSp-72k might functions as an adhesive protein in the aqueous silk of the caddisfly and that there might be inter-molecular links that are established between SmSp-72k and the fibroin molecules, L-fibroin and H-fibroin.

**Acknowledgments**

This work was supported by a Grant-in-Aid in Global COE Program by the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

**References**