Bulgarian Chemical Communications, Volume 49, Special Edition K2, (pp. 115 – 120) 2017

Characterization of a glycoprotein from the silk of caddisfly *Stenopsyche marmorata*

Q. Lin[1#], H. Wang[2#], H. Chen[1], Y. Peng[1], B. Kwan[1], M. Nakagaki[3], Y. Wang[1*]

[1] *Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Ocean College, Qinzhou University, Qinzhou-535011, China*
[2] *College of Humanities, Qinzhou University, Qinzhou-535011, China*
[3] *Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda-386-8567, Japan*

Received August 15, 2017, Revised November 15, 2017

By sequencing the N-termini of the luminal proteins and screening the silk gland cDNA library of the caddisfly *Stenopsyche marmorata (S. marmorata*), a major silk component (Smsp-12k) which proved to be a glycoprotein, was identified in this suborder. The primary structure of Smsp-12k consists of three degenerate repeats with Gly (29.7%), Trp (18.7%) and Ser (13.2%) present as the most abundant amino acids. The SXSXSXSX and GGX motifs were repeatedly found in Smsp-12k and are also present in the repetitive region of the heavy chain fibroin (H-fibroin) of *S. marmorata*. The SXSXSXSX and GGX motifs confer toughness and elasticity to the caddisfly silk, respectively. The abundant expression and structural resemblance of Smsp-12k to the repetitive region of H-fibroin indicates that this protein may play an important role in contributing to the physical properties of the aqueous silk.

Keywords: Caddisfly, Silk, *Stenopsyche marmorata*, Glycoprotein

INTRODUCTION

Silk, generally defined as fibrous proteins, is produced by many groups of terrestrial arthropods including spider and insect. Silk production in spider, generally used for prey capture, is secreted from spinnerets at the tip of the abdomen. The domestic silkworm *Bombyx mori* has served as a primary model for *Lepidoptera* insect silk. Silk fibroin is secreted in the lumen of posterior silk gland (PSG) and consists of three main protein components: heavy chain fibroin (H-fibroin; about 350 kDa), light chain fibroin (L-fibroin; 26 kDa) and fibrohexamerin (p25; about 30 kDa) proteins with a molar ratio of 6:6:1 [1]. It was presumed that these three components in the silk of Lepidopteran have been conserved over 150 million years, although the fibroin of *Saturnidae* species was secreted as dimer of H-chain [2]. H-chain and L-chain are linked by a single disulfide bond between Cys-172 of L-chain and Cys-c20 (twentieth residue from the C terminus) of H-chain [3]. It was suggested that in *Lepidoptera* the disulfide linkage of H-fibroin and L-fibroin is not an absolute requirement for secretion of fibroin [1]. But the H-L linkage is essential for the secretion of a vast amount of fibroin, because evidence showed that fibroin is retained in endoplasmic reticulum in the absence of disulfide linkage between the H-and L-chains [4]. P25 is a glycoprotein containing Asn-linked oligosaccharide chains and is associated with the H-L complex by noncovalent interactions, i.e. mainly by hydrophobic interactions with the H-chain moiety [5].

Special attention has been focused on the *Trichoptera* silk for decades due to its stable properties even under long exposure to water. Most caddisfly (*trichoptera*) spin a fine silk-catching net that is anchored to a substrate in fresh water habitats [6,7]. Tensile strain research has shown that the net silk of the caddisfly is a group III silk, with high extensibility and low breaking stress [8]. The physical properties of the silk mainly depend on the molecular design of the proteins which constitute the silk filament. The larvae of *S. marmorata* spin hiding tubes and catching nets (Fig. 1).

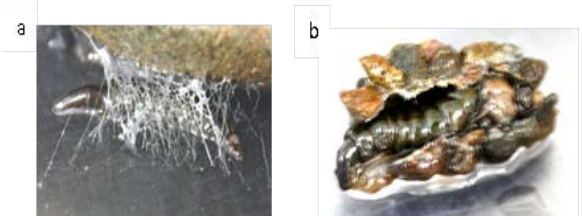

Fig. 1. A catching net (a) and hiding retreat (b) spun by the larva of *S. marmorata*

In a previous study, two filamentous constituents in the caddisfly *S. marmorata* were identified which termed the heavy chain fibroin (H-fibroin) and light chain fibroin (L-fibroin) [9]. Both proteins have counterparts in *lepidoptera*. The H-fibroin of *S. marmorata* has a high molecular weight of >400 kDa and is composed of four blocks (SA, SB, SC and SD) arranged in an order. Each

To whom all correspondence should be sent:
E-mail: wangxiaochen528@hotmail.com
#These authors contributed equally to this work

block is distinguished by unique motifs such as the SXSXSX(SX), $GPXG(X)_{1-3}$ or the triplet GGX sequence. The 24 kDa L-fibroin is similar to its counterpart found in lepidopteran silk. Characteristics found to be similar include molecular weight, charged residues and hydrophobicity, and presumably Cys-links to the H-fibroin to form the H/L-fibroin dimer in the lumen of the silk gland. The homologues of H-fibroin and L-fibroin have also been reported in other caddisfly species [10-11]. A 37 kDa Net forming protein (Nf-1) from one *Hydropsyche* species was assumed to be an essential component of the silk filament [12].

Both the proteins extracted from silk gland and the silk screen of *S. marmorata* were analyzed through SDS-PAGE. Unexpectedly, the results showed that a 16 kDa protein, rather than H-fibroin or L-fibroin, is the most abundant constituent in the lumen or fiber. Identification of the primary sequence of this 16 kDa protein would further contribute to our understanding of the caddisfly silk. In the present study, the protein was cloned and named Smsp-12k. The primary sequence characteristics of Smsp-12k suggest that this protein is an important component of *S. marmorata* silk.

EXPERIMENTAL

Insect care and sample preparation

Fully grown larvae of *S. marmorata* (*Annulipalpia* suborder, *stenopsychidae* family) were gathered from under the stones of the Chikuma River shallow area, Ueda, Japan. For most collected larvae, their silk glands were dissected immediately for total RNA isolation and lumen proteins collection. The freshly dissected silk glands were ruptured in cold distilled water (60 µl per pair of glands), incubated overnight at 4 °C and the luminal proteins that are released from the glands were collected for protein analysis. The remaining larvae were cultivated in a fish jar at room temperature to allow the *S. marmorata* to produce nests. Silk fiber proteins were extracted in 8 M urea and heated for 12 min in a water bath.

Silk protein analysis and cDNA library construction

The lumen proteins of silk glands and fibers were dissolved in a buffer solution containing 2% SDS and 5% β-mercaptoethanol. Denatured electrophoresis was performed by using an 8% polyacrylamide gel under reducing conditions. Part of the gels were stained with Coomassie brilliant blue R-250, Gel Code Glycoprotein Staining Kit (Pierce Biotechnology, Inc., USA), the other were electronically transferred onto a PVDF membrane (polyvinyl fluoride) for the purpose of protein sequencing. The steps of RNA isolation and cDNA library construction were based on the method described by Wang *et al*. (2010) [9].

Northern blotting

About 6 µg of the total RNA were collected from the silk glands of 20 insects. By utilizing the standard procedures of the DIG nucleic acid detection kit (Boehringer Mannheim), total RNA and RNA markers were electrophoresed on a 1% agarose gel (1× MOPS, 0.66 M formaldehyde) and subsequently blotted, as described by Wang *et al.*, (2010) [9]. The forward primer Sm12-FW: 5'-GCGTCAGTTGGGAACAT-3' (corresponding to the Smsp-12k nucleotides 109 to 126) and the reverse primer Sm12-RV: 5'-TCGCACTGTGACT AGCTGA-3' (corresponding to the Smsp-12k nucleotides 330 to 349) were used to amplify the Smsp-12k probe.

RESULTS AND DISCUSSION

The 16 kDa protein separated by SDS-PAGE showed the strongest band (Fig. 2).

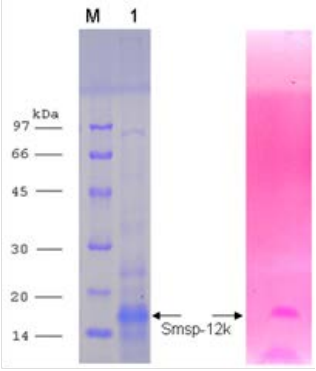

Fig. 2. Silk proteins of *S. marmorata* separated by SDS-PAGE from the silk fibers used to construct webs for food collection (lane 1) and from dope collected from the silk gland lumen (lane 3). Lane 2 shows the molecular mass standards (Bio-Rad) and the numbers on the left represent the molecular masses of the protein bands. The protein band at ~16 kDa (indicated by an arrow) was N-terminally sequenced and corresponds to the mature sequence of Smsp-12k.

According to sequencing results, the 11 amino acids of its N-terminus were read as HVGGYWPXGXG, in which the X amino acid was unidentified. An EST library was constructed from the silk glands of the final instar larvae of *S. marmorata*. Using the EST library, a gene encoding a protein with an N-terminal sequence of HVGGYWPXGXG was identified. Of the 260 ESTs analyzed, 18 fell into one cluster which was represented by the clone TBS2D11. TBS1A02

clone displayed a complete open reading frame (ORF) by sequence analyzing. Residues 22-32 of the ORF had the sequence HVGGYWPVGRG, which closely matched the N-terminal sequencing result for the unknown 16 kDa protein. Furthermore, the molecular weight of the protein deduced from TBS2D11 was approximately14 kDa, and the first 20 residues were predicted to be a signal peptide (Fig. 2). Consequently, taking into account the signal peptide, the gene from TBS2D11 was coded for a 12 kDa mature protein. The similar molecular weight and matching sequence revealed that the 16 kDa protein was encoded by the gene present in the TBS2D11 clone. The gene was named Smsp-12k for the gene encoded a 12 kDa silk protein of *S. marmorata*. However, the molecular weight of Smsp-12k observed in the SDS-PAGE analysis was larger than that determined from the gene sequence. Therefore, Smsp-12k may be a glycosylated protein. The sugar moiety of Smsp-12k was proved to be a glycosylated protein by using a glycoprotein staining kit (Fig. 2).

Such glycosylation has been found in particular *Chironmous* (midge) silk proteins which are also secreted underwater [13,14]. Northern blot analysis revealed that Smsp-12k is expressed specifically in the silk gland (Fig. 3).

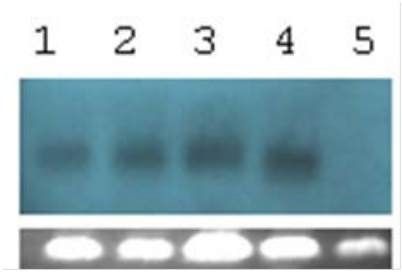

Fig. 3. Identification of Smsp-12k mRNAs on a Northern blot of total RNA (5 μg) prepared from the MSG-1, MSG-2, MSG-3 and PSG sections and the whole larva that were removed from the silk glands (lanes 1, 2, 3, 4 and 5, respectively). RNA was fractioned by electrophoresis and blotted onto Hybond N+ membranes (Amersham Bioscience). The probes were amplified from the appropriate cDNA clone (TBS2D11) using the special primers Sm12-FW and Sm12-RV, and isolated with a QIAquick gel extraction kit. Probes were labeled with the DIG-HIGH-primer reagent. As an internal control, ribosomal RNA was stained with ethidium bromide as presented at the bottom of the photo.

The Smsp-12k gene had an ORF of 396 base-pairs and was predicted to translate to a premature protein of 132 residues. The sequence of Smsp-12k is shown in Figure 3. Amino acid analysis shows that, like most silk proteins, Gly (29.7%) is the predominant amino acid present in the sequence. Trp (18.7%) and Ser (13.2%), unexpectedly, were the second and third most abundant amino acid, respectively. The aromatic residue Trp was an energy rich, probably essential amino acid for insects [15]. But the residue could be supplied by the larva's main food resource, such as algae which contain energy rich amino acids (for example Arg, Leu, His and Trp). Comparison of the Smsp-12k and H-fibroin protein sequences showed that both sequences contained the same motifs (Fig. 4).

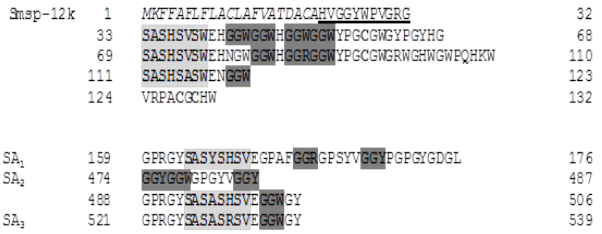

Fig. 4. Comparison of the primary sequences of (A) Smsp-12k with (B) the SA1, SA2 and SA3 blocks from the H-fibroin of *S. marmorata*. The core region of Smsp-12k is loosely divided into three repeats based on the alignments of the motifs. The first 21 residues (italics) were proved to be a signal peptide. Two major motifs SXSXSXSX (highlighted in light gray) and GGX (highlighted in heavy gray) were conserved in Smsp-12k and H-fibroin. The cysteines in Smsp-12k which likely form intermolecular or intramolecular disulfide bonds are printed in bold.

Four blocks (SA, SB, SC and SD), tandemly repeated many times, represent the large repetitive regions of *S. marmorata* H-fibroin. The three degenerate SA blocks SA1, SA2 and SA3 which are located at the N-terminus of H-fibroin contain motifs such as SXSXSXSXE and GGX which are motifs present in Smsp-12k. The Ser-rich motifs (SX)n are likely to form a β-sheet and three-dimensional crystallites that reinforce the silk filaments, whereas the GGX triplet provides a stable spiral conformation conferring elasticity to the protein polymer [16-18]. However, DNA analysis revealed that codon bias for the abundant amino acids in Smsp-12k and the three degenerate SA blocks in H-fibroin were different (Table 1), indicating that the genes of Smsp-12k and H-fibroin were not evolutionary related. The research about caddisfly silk gene was very limited. However, the rapid development of science and technology as well as the discovery of caddisfly silk applications has enhanced our understanding of silk protein and its properties.

Yonemura *et al.* (2006) [10] firstly reported two important silk fibroins from Czech caddisfly silk,

Table 1: Codon bias for the three most abundant amino acids in Smap-12k and the SA_1, SA_2 and SA_3 blocks from the H-fibroin of *S. marmorata**

	Smsp-12k		$SA_1 + SA_2 + SA_3$	
	Occurrence (%)	Codon (%)	Occurrence (%)	Codon (%)
Gly	29.7%	gga/40.8% ggc/22.2% ggg/3.7% ggt/33.3%	33.3%	gga/63.3% ggc/6.7% ggt/30%
Ser	13.2%	agc/25% agt/50% tca/8.3% tcg/8.3% tct/8.3%	14.4%	agc/23.1% agt/30.6% tca/23.1% tcc/15.4% tct/7.7%
Pro	4.4%	cca/50% cct/50%	8.9%	cca/100%

*The three blocks of SA_1, SA_2 and SA_3 are regarded as a whole in this analysis and their sequences are presented in Fig.1.

namely caddisfly fibroin heavy chain (Fib-H), a heavy chain protein with a molecular weight greater than 300 kDa, and caddisfly fibroin light chain (Fib-L), a light chain protein with a molecular weight about 25 kDa. As the main components of caddisfly silk, both protein molecules are homologous proteins of fibroin protein of *Bombyx mori* [11]. Nevertheless, the silkworm protein P25 and sericin homologous protein of the caddisfly silk have not been found. For the caddisfly silk protein with small molecular weight, only three novel silk fiber components were determined, in which two of them are active enzyme proteins in the silk, and another is a small molecule structural protein. The two enzyme proteins, which distributed in silk fibroin and peripheral layers of sticky protein, are associated with the cohesive function of the caddisfly silk [19]. The structural protein is a kind of protein rich in proline, similar to the muscle group (titin) PEVK in the field of protein structure, and is specifically distributed in the fibroin layer [19]. All these small molecular weight proteins were identified through the silk gland transcriptome and proteomics; however, their concentrations in the caddisfly silk are very low. In the analysis of SDS-PAGE of caddisfly silk, we found that the amount of Smsp-12 k protein is richer than the light chain protein, after the caddis heavy chain protein, demonstrating the importance of this protein.

The expression of Smsp-12K varied seasonally. It tends to be higher in summer than in winter, suggesting that the protein plays more critical roles in summer. As the caddisfly larvae are more active and spin more silk fibers in summer, it suggests that efficient silk production may be related to the Smsp-12K.

Compared with the silkworm silk (*Bombyx mori*) or spider silk (*Araneida*), the caddisfly silk has many unique properties. Keeping high viscosity in torrent water is the most remarkable feature of caddisfly silk [20-22]. Apart from that, caddisfly silk has excellent physical and mechanical properties. Though caddisfly silk is soft, it remains with high tensile strength and breaking elongation in the water. The fracture strength of some caddisfly silk species can be five times that of the steel. Its breaking elongation is 150% - 200%, which is much higher compared to the silkworm silk [23-25]. Caddisfly silk also has the feature of high-temperature resistance and heavy metal ions adsorption. The maximum decomposition temperature of caddisfly silk is 242 °C, which is higher than that of the silkworm silk (220 °C) and of most spider silk (230 °C) [24]. The highly efficient adsorption feature of caddisfly silk on heavy metal ions, especially Cd, Pb, and Zn, has been confirmed several years ago [26]. In addition to the above-mentioned properties, the caddisfly silk has considerable antibacterial activity and cell growth promotion. After the sterilization process (121°C, 115 kPa, 20 min), the antibacterial property of caddisfly silk was maintained without affecting its physical properties [24].

It is generally believed that the Fib-H protein is the decisive protein which determines the specific properties of the silk [25, 27]. There is a high homology of the N and C terminals of Fib-H protein between caddisfly and *Bombyx mori*, but their sequence of the repeat region of the silk varied.

The repeat regions of Fib-H protein were connected with four blocks, SA, SB, SC and SD. species. Each repeat consists of at least one (pSX)n motif (pS = phosphoserine, X = arginine or aliphatic amino acid, n =2–6).

The unique structure of SA repeats is believed to be closely related to the properties of silk. It contains two (pSX)4 motifs and 11 residues in between of the two motifs. This is easy to form β-sheet structure [20]. It is believed that the structure forms the molecular basis of the featured mechanical property of the silk in water.

A large number of SA-like structures was found in Smsp-12k. We speculate that the serines in the SA unit were also highly phosphorylated, and form complexes with metal ions such as Ca^{2+} and Mg^{2+}. Together with Fib-H, Smsp-12k plays a key role in the excellent performance of the caddisfly silk.

As observed for H-fibroin, hydrophilicity also prevailed in Smsp-12k (Fig. 5). Smsp-12k most likely forms intermolecular disulfide bonds through the cysteines present in the sequence and such features of Smsp-12k provide protein stability and

insolubility in an aqueous environment. The H-fibroin and L-fibroin, examined in *S. marmorata* have additional cysteines in contrast to the counterpart proteins present in *Lepidoptera* [1, 10].

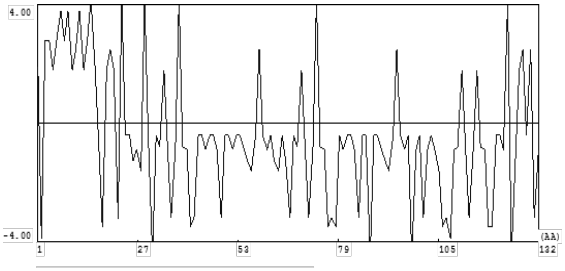

Fig. 5. Kyte-Doolittle hydropathy plots for the determined amino acid sequence of Smsp-12k. The scale bar depicts the number of amino acid residues. The numbers on the axis above the plots indicate amino acid position. The residues above the abscissa indicate hydrophobicity, while those below the abscissa indicate hydrophilicity. Smsp-12k is largely hydrophobic and the total average hydropathicity value is 0.44.

It is generally accepted that the order *Lepidoptera* and its sister order *Trichoptera* diverged from the common ancestor *Amphiesmenoptera* and it was presumed that the polymer components of H-fibroin, L-fibroin and p25 have been conserved over 150 million years in the silk of *Lepidoptera* [2], however, the homologue p25 was not found by studying the filament of caddisfly suborder of *H. angustipennis* and *L. decipiens* [10]. In *S. marmorata*, the homologue of p25 also was not found; this confirmed the suggestion that the lack of p25 is a general feature of *Trichoptera* [10]. On the contrary, Smsp-12k homologue was found in all reported suborders of caddisfly species, and was a major component in their spinning silk. It was hypothesized that the net-spinning caddisfly produce silk-like protein and adhesive protein simultaneously [10]. But no adhesive protein was reported to date.

Considering the glycosylated property of Smsp-12k and its high expressing, we infer that Smsp-12k may be playing an important role in making caddisfly silk sticky underwater. Consequently, Smsp-12k most likely links with the H-fibroin or L-fibroin by formation of disulfide bridges to form an insoluble filament polymer in the lumen of *S. marmorata* silk glands.

CONCLUSION

In summary, data in this study revealed that Smsp-12k is one of the most abundant proteins in the silk of *S. marmorata*. The repeat sequence regions of this protein are mainly composed of two motifs ((SX)n and GGX) which are also present in the H-fibroin of *S. marmorata*.

The (SX)n motif is the only sequence that has the potential to form a crystalline structure in the H-fibroin, raising the intensity of the silk. The combination of the (SX)n and GGX motifs within the Smsp-12k sequence contribute significantly to the physical properties of the *S. marmorata* silk.

Acknowledgements: *The authors acknowledge the financial support of the Natural Science Foundation of Guangxi Province (No: 2016GXNSFBA380156, 2016GXNSFCA380007), the Natural Science Foundation of china (No:31760713), the Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi (No: KY2016YB498) and the Foundation of Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Qinzhou University (No: 2018ZB08).*

REFERENCES

1. S. Inoue, K. Tanaka, F. Arisaka, S. Kimura, K. Ohtomo, S. Mizuno, *J. Biol. Chem.*, **275**, 40517 (2000).
2. N. Yonemura, F. Sehnal, *J. Mol. EVol.*, **63**, 42 (2006).
3. K. Tanaka, N. Kajiyama, K. Ishikura, S. Waga, A. Kikuchi, K. Ohtomo, T. Takagi, S. Mizuno, *Biochim. Biophys. Acta*, **1432**, 92 (1999).
4. M.L. Casem, L.P. Tran, A.M. Moore, *Tissue Cell.*, **34**, 427 (2002).
5. K. Mori, K. Tanaka, Y. Kikuchi, M. Waga, S. Waga, S. Mizuno, *J. Mol. Biol*., **251**, 217 (1995).
6. P.S. Giller, B. Malmqvist, *The biology of rivers and streams*, Oxford University Press, 119, 1998.
7. R.W. Holzenthal, R.J. Blahnik, A.P. Prather, *Zootaxa*, **1668**, 639 (2007).
8. S.A. Brown, D.R. Graeme, H. Stuart, *J. N. Am. Benthol. Soc.*, **23**, 771 (2004).
9. Y.J. Wang, K. Sanai, H.X. Wen, T.F. Zhao, M. Nakagaki, *Mol. Biol. Rep.*, **37**, 2885 (2010).
10. N. Yonemura, F. Sehnal, K. Mita, T. Tamura, *Biomacromolecules*, **7**, 3370 (2006).
11. N. Yonemura, T. Tamura, K. Mita, F. Sehnal, *J. Mov. Evol*., **68**, 641 (2009).
12. J.H. Eum, S.M. Yoe, Y.R. Seo, S.W. Kang, S.S. Han, *Insect. Biochem. Mol. Biol.*, **35**, 435 (2005).
13. S.E. Wellman, S.T. Case, *J. Biol. Chem.*, **264**, 10978 (1989).
14. S.V. Smith, J.J. Correia, S.T .Case, *Protein Sci.*, **4**, 945 (1995).
15. R. Fedic, M. Zurovec, F.J. Sehnal, *J. Biol. Chem.*, **278**, 35255 (2003).
16. E. Bini, D.P. Knight, D.L. Kaplan, *J. Mol. Biol.*, **335**, 27 (2004).
17. J. Ashida, K. Ohgo, K. Komatsu, A .Kubota, T Asakura, *J. Biomol. NMR.*, **25**, 91 (2003).
18. R. Fedic, M. Zurovec, F. Sehnal, *J. Insect Biotech. Sericol.*, **71**, 1 (2002).

19. N. Ashton, R. Stewart, D. Ferré, A. Amaré, G. Prendergast, E. Ziff, S. Burley, *Soft Matter*, **11**, 1667 (2015).
20. R. Stewart, C. Wang, *Biomacromolecules*, **11**, 969 (2010).
21. C. Wang, N. Ashton, R. Weiss, R. Stewart, *Insect Biochem. Mol. Biol.*, **54**, 69 (2014).
22. C. Wang, H. Pan, G. Weerasekare, R. Stewart, *Journal of The Royal Society Interface*, **12**, 112 (2015).
23. S. Brown, G. Ruxton, S. Humphries, *J.N. Am. Benthol.Soc.*, **23**, 771 (2004).
24. M. Tszydel, A. Zabłotni, D. Wojciechowska, *J. Mech. Behav. Biomed. Mater.*, **45**, 142 (2015).
25. L. Albertson, M. Daniels, *Freshwater Biology*, **61**, 670 (2016).
26. M. Tsukada, M.M. Khan, E. Inoue, G. Kimura, J.Y. Hun, M. Mishima, K. Hirabayashi, *Int. J. Biol. Macromol.*, **46**, 54 (2010).
27. Y. Wang, K. Sanai, M. Nakagaki, *Adv. Mater. Res.*, **79**, 1631 (2009).