

# Biosynthesis of a potentially functional polypeptide derived from silk fibroin

Huanrong Zhao, Yunxing Yang, Honggen Yi, Gaoqiang Yang and Jiannan Wang\*  
*National Engineering Laboratory for Modern Silk, College of Textile and Clothing Engineering,  
Soochow University, Suzhou 215123, China*

**Abstract.** In order to understand the relationship between sequences and biological functions of RGD-containing wild silkworm silk fibroin, it is important to purify the basic RGD-containing motif in large quantities. In this study, a gene monomer encoding RGD-contained motif GSGAGGRGDGGYGS<sub>n</sub>SS (-RGD-) derived from *Antheraea pernyi* (the same in *Antheraea yamamai*) was designed and cloned. (-RGD)<sub>n</sub> in various degrees of polymerizations was obtained by gene monomer doubling-extension and expression. Two glutathione-S-transferase (GST)-tagged fusion proteins GST-(-RGD)<sub>12</sub> and GST-(-RGD)<sub>24</sub> were successfully expressed in *Escherichiacoli* (*E. coli*) BL21. The fusion proteins were isolated and purified by GST affinity chromatography, and the polypeptides (-RGD)<sub>12</sub> and (-RGD)<sub>24</sub> were cleaved from GST fusion proteins by thrombin digestion. Two-dimensional electrophoresis and amino acid composition analysis were performed to confirm the identity of the engineered polypeptides. Results indicated that this technology reliably obtained expected polypeptides (-RGD)<sub>n</sub> for future research on structure and functions.

Keywords: Silk fibroin, RGD, recombinant expression, amino acid composition, isoelectric point

## 1. Introduction

Silk fibroin is a natural protein synthesized and secreted by silkworms, including the domesticated silkworm (*Bombyx mori*) and wild silkworm species (e.g., *Antheraea yamamai*, *Antheraea pernyi*, *Antheraea mylitta*, *Antheraea roylei*, *Antheraea frithi*, *Samia ricini*,) [1]. *B. mori* silk fibroin is useful for the development of biomaterials because of its biodegradability, biocompatibility and low immunogenicity. Moreover, since silk fibroin is the natural product of the organism, this ensures cell affinity and tissue compatibility. There are numerous reports demonstrating that cells such as fibroblasts, vascular cells, mesenchymal stem cells and glial cells exhibit good cell adhesion, spread and growth on *B. mori* silk fibroin materials. Given that rapid cell adhesion is a key factor for promoting the repair of defective tissues, provision of cell adhesion recognition interface is vital. It is well known that the RGD (Arg-Gly-Asp) tripeptide is a recognition sequence for promoting cell adhesion. Indeed, the RGD tripeptide is present in many extracellular matrices including fibronectin

---

\*Corresponding author: Jiannan Wang, College of Textile and Clothing Engineering, Soochow University, No. 199 Ren-ai Road, Suzhou Industrial Park, Suzhou, 215123, P.R. China. Tel.: +86 512-67061153; Fax: +86 512-67061126; E-mail: wangjn@suda.edu.cn.

(FN), laminin (VN), fibrin (FB) and collagen (CA), where it can associate with the  $\beta 1$  methylene group of cell surface integrins [2].

The RGD tripeptide is present in silk fibroin produced by wild silkworm species (e.g., *A. yamamai* [2,3], *A. pernyi* [4], *A. mylitta* [5]), but is absent from the domesticated *B. mori* silk fibroin [6]. On this basis, wild silkworm silk fibroin should share some of the unique properties of biomaterials (i.e. biodegradability, biocompatibility and low immunogenicity), as well as possessing excellent cell affinity and other potential biological functions. It has been previously reported that regenerated silk fibroin materials from *A. yamamai* [7],

*A. pernyi* [8] or *A. mylitta* [9,10] are potentially useful biomaterials with good cell adhesion and cytocompatibility. Unfortunately, the habit of the wild silkworm species renders them unsuitable for domestication, which limits large-scale application. Furthermore, it was reported recently that transgenic technology can be used to induce the synthesis and secretion of *A. yamamai* silk fibroin from the silk glands of *B. mori* at a low level [11]. Clearly, detailed biochemical understanding of the wild silkworm silk fibroin is required before its successful transformation into biomaterials.

In this article, we analyzed and studied *A. yamamai* silk fibroin and *A. pernyi* silk fibroin because its similarities in amino acid sequences. The protein sequence is composed of several repetitive conserved modules. In comparison with other modules, we chose to design a gene monomer encoding the short consensus sequence GSGAGGRGDGGYGS GSS, which is repeated in the two silk fibroins. The gene monomer was cloned and extended up to 24 repeats, followed by induced expression in *E. coli* BL21. We expect our work to provide the key base materials needed for the future development of biomaterials from silk fibroin.

## 2. Materials and methods

### 2.1. Gene design, multimerization and expression vector construction

Double-stranded oligonucleotides (AYR1): 5'CCGGTAGATCTGGTGCTGGAGGACGAGGTGACGGCGGTTATGGTTCAGGCTCTTCAG3' and 5'GATCCTGAAGAGCCTGAACCATAACCGCCGTCACCTCGTCCTCCAGCACCAGATCTA3' were designed to encode an RGD-containing motif SGAGGRGDGGYGS GSSG. The synthetic oligonucleotides (Invitrogen) were flanked by two restriction enzyme sites *AgeI*...*Bg**II* in the N terminus and a *Bam**HI* in the C terminus. AYR1 was cloned into *AgeI*/*Bam**HI* digested plasmid pSLFA1180FA (kept in our Lab) to construct plasmid pSL-AYR1 containing a single copy of the coding gene. The gene motif AYR1 was doubled and extended to 2-, 4-, 8-, 12-, 16- and 24-mer using anisocaudamer (*Bg**II*/*Bam**HI*) strategy. The resulting

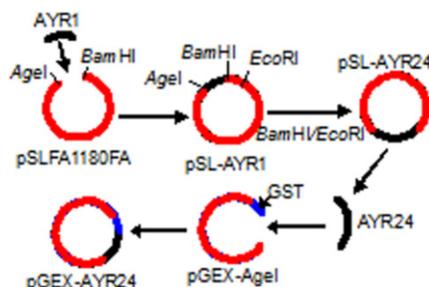


Fig. 1. Abridged general view of plasmid construction.

plasmids were designated pSL-AYR2, pSL-AYR4, pSL-AYR8, pSL-AYR12, pSL-AYR16 and pSL-AYR24, respectively. Finally, the extended genes (AYR)<sub>n</sub> were inserted into pGEX-AgeI [12] to construct expression vector pGEX-(AYR)<sub>n</sub> (Figure 1).

## 2.2. Protein expression

Protein expression and protein purification were performed as described previously by Wang et al. [12]. Briefly, the expression vector pGEX-(AYR)<sub>n</sub> was transformed into expression strain *E. coli* BL21 (DE3). A single colony was inoculated into 4 ml Luria-Bertani (LB medium) containing 100 µg/ml ampicillin and cultured at 37°C overnight, then amplified in 250 ml of fresh LB ampicillin medium the next day. When OD<sub>600</sub> was 0.6–0.8 AU, aqueous solution of IPTG was added into the cultures to a final concentration 1.0 mM to induce expression of the target protein. After 8 h of induction, cells were harvested by centrifugation at 4°C and stored at -80°C.

## 2.3. Protein purification

The fusion protein GST-(-RGD-)<sub>n</sub> was purified by the GST affinity purification system (Novagen). The cell pellet was suspended in 25 ml of GST bind/wash buffer and sonicated on ice (Biorupter JYD-900, Shanghai Zhixin Instrument). The lysate was centrifuged at 4°C to remove debris. The supernatant was loaded onto a 4 ml GST affinity column. Following column washes with 50 ml of GST wash buffer, the fusion protein GST-(-RGD-)<sub>n</sub> was eluted with 20 ml of GST elution buffer containing 10 mM reduced glutathione.

## 2.4. SDS-PAGE and western bolt

SDS-PAGE and western bolt analysis were performed as described by Wang et al. [12]. 5 µl of loading buffer was added to 20 µl of the whole cell lysate or purified fusion protein and boiled for 3–5 min. The mixture was loaded onto a 12% (w/v) polyacrylamide gel (Sigma), and then stained using Coomassie brilliant blue. After SDS-PAGE, the proteins were transferred onto a PVDF membrane (Millipore), and subsequently blocked with TBST containing 5% of nonfat dried milk. Next, the membrane was incubated in 1000× diluted GST antibody (Novagen) for 1 h, after washing with TBST three times, horseradish peroxidase labeled rabbit IgG antibody (ProSci incorporated) was added. The antibody was detected using DAB reagent (Wuhan Boster Biotech Co. Ltd.).

## 2.5. Cleavage of GST-(-RGD-)<sub>n</sub>

After ultrafiltration using Ultra-0.5 amicon (Millipore) with a molecular weight cutoff of 3000, 1U of thrombin (Novagen) was added per milligram of fusion protein in 1×cleavage buffer at 20°C for 16 h. After dialysis against deionized water, the reaction mixture was diluted in GST binding buffer and loaded onto GST affinity column to remove the GST tag.

## 2.6. Amino acid composition measurement

The fusion protein GST-(-RGD-)<sub>n</sub> was diluted to 0.05 mg/ml and filtered through a 0.22 µm filter to remove protein precipitates. 6 N HCl was then added to hydrolyze peptide bonds, followed by analysis of amino acid composition using Hitachi L-8800 (Hitachi) amino acid analyzer.

### 2.7. Two-dimensional gel electrophoresis

The isoelectric point (pI) of the digested protein mixture of GST-(-RGD-)<sub>12</sub> by thrombin was measured by Protean IEF Cell (Bio-Rad). 100 ml of protein mixture was resuspended in 400 ul of 2D sample buffer (8 M urea, 65 mM DTT, 4% w/v CHAPS, 0.2% w/v Bio-Lyte, 0.001% w/v bromophenol blue). Immobilized pH gradient (IPG) strip (3–10, 17 cm, Bio-Rad) was rehydrated in the sample, and isoelectric focusing (IEF) was in a Multiphor (Bio-Rad) for 60 kVh at 17°C. After focusing, the IPG strip was immediately equilibrated for 15 min in buffer (6 M urea, 2% w/v SDS, 0.375 M Tris pH 8.8, 20% v/v glycerin, and 130 mM DTT), then placed on top of the second dimensional gel (12% w/v, Sigma). Proteins were separated by SDS-PAGE and stained with Coomassie brilliant blue.

## 3. Results and discussion

### 3.1. Identification of gene extension

Plasmids pSL-AYR1–24 were digested by *Bgl*II/*Bam*HI and identified by agarose gel electrophoresis. As shown in Figure 2, two bands appeared in each lane. The molecular size of the upper band in each lane is 3216 bp, while the lower bands are 51, 102, 204, 408, 612, 816 and 1224 bps, respectively. According to the DNA molecular mass standards, the sizes of the two bands from the plasmids pSL-AYR1–24 matched the expected values.

Restriction enzymes *Bgl*II and *Bam*HI are a pair of isocaudamers, after ligation, their enzymatic sites (AGATCT and GGATCC) are destroyed and the gene sequence GGATCT encoding Gly and Ser is formed (Figure 3, box line labeled). The method is an efficient strategy for duplication of amino acid motifs connected head-to-tail.

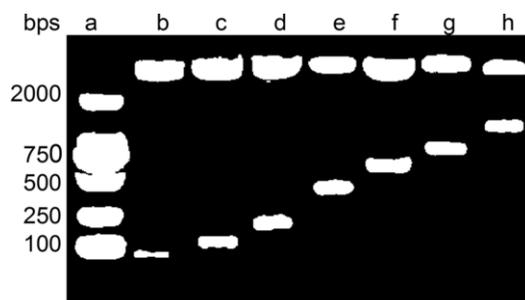


Fig. 2. Graph of DNA gel electrophoresis. a: DNA molecular mass standards; b–h from left to right: pSL-AYR1, pSL-AYR2, pSL-AYR4, pSL-AYR8, pSL-AYR12, pSL-AYR16 and pSL-AYR24.

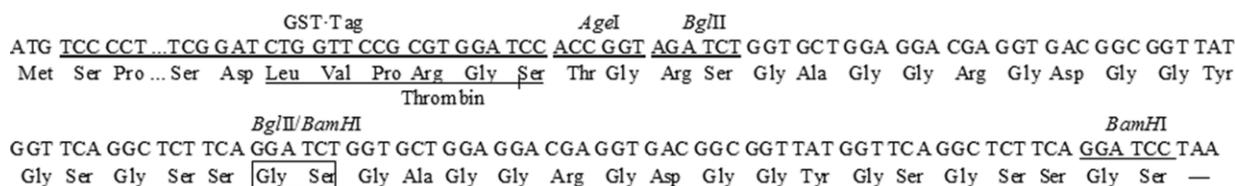


Fig. 3. Amino acid and coding sequence of dimeric GST-(-RGD-)<sub>2</sub>.

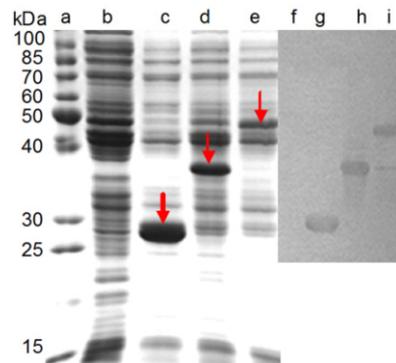


Fig. 4. SDS-PAGE and western blot analysis of total proteins in BL21. a: Protein molecular mass standards; b and f: BL21 without expression vector; c and g: BL21 with pGEX-AgeI; d and h: BL21 with pGEX-AYR12; e and i: BL21 with pGEX-AYR24.

### 3.2. Expression analysis of fusion proteins

Multimers of the  $(AYR)_n$  gene were expressed under the control of the Ptac promoter and the relevant components g10 and RBS for regulation of protein translation. A GST tag was used for initial purification purposes, while a protease (thrombin) recognition site LeuValProArgGlySer was inserted for subsequent enzymatic release of the target protein from the GST-tag by cleaving the amide linkage between Arg and Gly (Figure 3).

The expression of GST fusion proteins encoded by pGEX-AYR12 and pGEX-AYR24 in *E. coli* BL21 cells was induced by IPTG to obtain the desired products GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub>. The predicted molecular weights of GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub> are 43.1 kDa and 59.5 kDa respectively. As shown in Figure 4, approximate bands of the GST tag (about 26.3 kDa), GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub> clearly appeared in lane c, lane d and lane e (red arrowhead). The western blot result also confirmed the presence of the GST tag, GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub> in lane g, lane h and lane i, while absent in lane f representing BL21 cells not transformed with the expression vector.

Our results indicates that the fusion proteins GST-(-RGD-)<sub>24</sub> and GST-(-RGD-)<sub>12</sub> can be expressed successfully in *E. coli* BL21. However, given that the expression level of large molecular weight, foreign proteins is usually low in *E. coli* BL21, especially when the molecular weight is more than 60 kDa. Therefore, this may account for the reduced expression of GST-(-RGD-)<sub>24</sub> (59.5 kDa) compared to GST-(-RGD-)<sub>12</sub>.

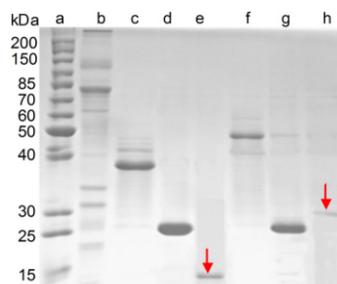


Fig. 5. SDS-PAGE of purified and cleaved proteins. a: Protein molecular mass standards; b: Thrombin; c: GST-(-RGD-)<sub>12</sub>; d and g: GST from GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub>; e: (-RGD-)<sub>12</sub>; f: GST-(-RGD-)<sub>24</sub>; h: (-RGD-)<sub>24</sub>.

### 3.3. SDS-PAGE of purified fusion proteins and released polypeptides

The fusion proteins GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub> were easily purified by GST affinity chromatography in high purity (Figure 5, Lane c and f). According to the protein molecular mass standards, ~40 kDa of GST-(-RGD-)<sub>12</sub> and 50~60 kDa of GST-(-RGD-)<sub>24</sub> were expressed successfully in *E. coli* BL21. Fusion protein

GST-(-RGD-)<sub>12</sub> was composed of 26.3 kDa GST and 16.8 kDa (-RGD-)<sub>12</sub>, GST-(-RGD-)<sub>24</sub> was composed of 26.3 kDa GST and 33.2 kDa (-RGD-)<sub>24</sub>. After cleavage by thrombin and dialysis, the mixed solution containing GST tag and polypeptide was purified by GST affinity column to remove GST. The flow through solution contains the released polypeptides (-RGD-)<sub>12</sub> or (-RGD-)<sub>24</sub>. Efficient thrombin cleavage can be seen in Figure 5, eluted GST (Figure 5, Lane d and g), the polypeptides (-RGD-)<sub>12</sub> (Figure 5, Lane e) and (-RGD-)<sub>24</sub> (Figure 5, Lane h) were digested by thrombin. .

### 3.4. Amino acid composition of fusion proteins

The fusion protein of GST-(-RGD-)<sub>12</sub> contained 434 amino acid residues and the GST-(-RGD-)<sub>24</sub> contained 638 amino acid residues. The experimental mole percentage of each amino acid was very close to the theoretical value except for Glu, Cys, Gly and Ser (Table 1). The fusion proteins of GST-(-RGD-)<sub>12</sub> and GST-(-RGD-)<sub>24</sub> were mainly composed of Gly, Ser, Arg, Asp, Ala and Leu, the theoretical percentages of these amino acids were 66% and 75%, respectively. These amino acids were also the main compositions in both of the measured values, and the total percentages were 62.06% and 62.62%, respectively. They basically accorded with the theoretical values. These results indirectly indicate that there was no gene mutation or mismatch during the process of protein expression.

According to amino acid composition analysis, the mole percentages of Glu and Cys were higher than theoretical values, the reason for this is that the reduced glutathione used for protein purification may not have been completely removed by ultrafiltration. Similarly, the residual reduced glutathione may also led to the lower than expected measured value for Ser.

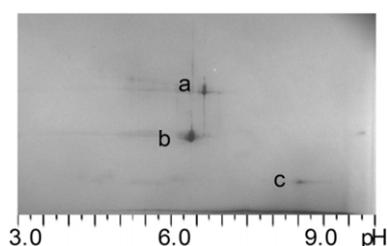
### 3.5. Two-dimensional electrophoresis analysis of GST-(-RGD-)<sub>12</sub>

The amphotericity is a key factor for analyzing the performance of a specific protein. The pI theoretical value of the fusion protein GST-(-RGD-)<sub>12</sub> was about 6.34 and (-RGD-)<sub>12</sub> was about 8.17. After digestion by thrombin, three protein points were separated on the second dimensional gel, representing GST-(-RGD-)<sub>12</sub> (Figure 6, a), GST (Figure 6, b) and (-RGD-)<sub>12</sub> (Figure 6, c). Figure 6 shows the pI value of the fusion protein GST-(-RGD-)<sub>12</sub> was about 6.5 and (-RGD-)<sub>12</sub> was about 8.5, were all close to their theoretical pI 6.34 and 8.17.

Table 1

Amino acid composition of expressed fusion proteins

Amino acid	GST-( $-RGD-$ ) <sub>12</sub> (mol %)		GST-( $-RGD-$ ) <sub>24</sub> (mol %)	
	Theoretical	Experimental	Theoretical	Experimental
Ala	5.10	6.10	5.30	6.00
Arg	5.30	4.85	5.50	4.73
Asn	0.90	0.00	0.60	0.00
Asp	6.90	8.17	6.60	7.87
Cys	0.90	2.98	0.60	3.51
Gln	1.20	0.00	0.80	0.00
Glu	3.70	7.96	2.50	8.13
Gly	28.60	25.80	36.40	27.64
His	1.40	1.43	0.90	1.29
Ile	3.00	2.87	2.00	2.45
Leu	6.50	8.00	4.40	6.72
Lys	4.80	5.02	3.30	4.46
Met	2.10	2.15	1.40	1.98
Phe	2.10	2.36	1.40	3.10
Pro	3.20	2.00	2.20	2.00
Ser	13.60	9.14	16.80	9.66
Thr	1.60	2.00	1.10	1.97
Trp	0.90	0.00	0.60	0.00
Tyr	6.00	5.11	6.00	5.31
Val	2.30	3.78	1.60	3.38

Fig. 6. Two-dimensional electrophoresis of cleaved GST-( $-RGD-$ )<sub>12</sub>. a: GST-( $-RGD-$ )<sub>12</sub>; b: GST tag; c: ( $-RGD-$ )<sub>12</sub>.

#### 4. Conclusion

Two ( $-RGD-$ ) contained target polypeptides ( $-RGD-$ )<sub>12</sub> and ( $-RGD-$ )<sub>24</sub> were expressed successfully in *E. coli* BL21 and released from the fusion proteins. Amino acid and two-dimensional electrophoresis analysis confirmed that the expression products matched our expected designs. Future work is required to optimize high expression and purification of the multimers ( $-RGD-$ )<sub>n</sub> in order to better study its function, as well as for its future applications in biomaterials.

#### Acknowledgement

This study is supported by National Natural Science Foundation of China (51173125), Natural Science Foundation of Jiangsu Province of China (BK2012633, BK20141210), College Natural Science Research Project of Jiangsu Province of China (12KJA43004), Science and Technology Plan

Foundation of Suzhou, China (ZXS2012002) and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

## References

- [1] S.C. Kundu, B. Kundu, S. Talukdar, S. Bano, S. Nayak, J. Kundu, B.B. Mandal, N. Bhardwaj, M. Botlagunta, B.C. Dash, C. Acharya and A.K. Ghosh, Nonmulberry silk biopolymers, *Biopolymers* **97** (2012), 455–467.
- [2] E. Ruoslahti and M.D. Pierschbacher, New perspectives in cell adhesion: RGD and integrins, *Science* **238** (1987) 491–497.
- [3] J.S. Hwang, J.S. Lee, T.W. Goo, E.Y. Yun, K.S. Lee, Y.S. Kim, B.R. Jin, S.M. Lee, K.Y. Kim, S.W. Kang and D.S. Suh, Cloning of the fibroin gene from the oak silkworm, *antheraea yamamai* and its complete sequence, *Biotechnol. Lett.* **23** (2001), 1321–1326.
- [4] H. Sezutsu and K. Yukuhiro, Dynamic rearrangement within the *antheraea pernyi* silk fibroin gene is associated with four types of repetitive units, *J. Mol. Evol.* **51** (2000), 329–338.
- [5] A. Datta, A.K. Ghosh and S.C. Kundu, Differential expression of the fibroin gene in developmental stages of silkworm, *antheraea mylitta* (saturniidae), *Comp. Biochem. Phys. B: Biochem. Mol. Biol.* **129** (2001), 197–204.
- [6] C.Z. Zhou, F. Confalonieri, M. Jacquet, R. Perasso, Z.G. Li and J. Janin, Silk fibroin: Structural implications of a remarkable amino acid sequence, *Proteins* **44** (2001), 119–122.
- [7] J.N. Wang, H.G. Yi and Y.Q. Wei, Preliminary biocompatibility evaluation of regenerated *antheraea yamamai* silk fibroin in vitro, *J. Wuhan Univ. Technol.* **26** (2011), 1044–1048.
- [8] X.Y. Luan, Y. Wang, X. Duan, Q.Y. Duan, M.Z. Li, S.Z. Lu, H.X. Zhang and X.G. Zhang, Attachment and growth of human bonemarrow derived mesenchymal stem cells on regenerated *antheraea pernyi* silkfibroin films, *Biomed. Mater.* **1** (2006), 181–187.
- [9] C. Acharya, S.K. Ghosh and S.C. Kundu, Silk fibroin protein from mulberry and non-mulberry silkworms: Cytotoxicity, biocompatibility and kinetics of L929 murine fibroblast adhesion, *J. Mater. Sci.: Mater. Med.* **19** (2008), 2827–2836.
- [10] B.B. Mandal and S.C. Kundu, Osteogenic and adipogenic differentiation of rat bone marrow cells on non-mulberry and mulberry silk gland fibroin 3D scaffolds, *Biomaterials* **30** (2009), 5019–5030.
- [11] I. Kobayashi, K. Kojima, H. Sezutsu, K. Uchino and T. Tamura, Expression of the Japanese oak silkworm *antheraea yamamai* fibroin gene in the domesticated silkworm *bombyx mori*, *Insect. Sci.* **16** (2009), 465–473.
- [12] J.N. Wang, S.Q. Yan, C.D. Lu and L. Bai, Biosynthesis and characterization of typical fibroin crystalline polypeptides of silkworm *bombyx mori*, *Mat. Sci. Eng. C* **29** (2009), 1321–1325.