Materials Science and Engineering C 59 (2016) 278–285

Contents lists available at ScienceDirect

Materials Science and Engineering C

journal homepage: www.elsevier.com/locate/msec

Biosynthesis and characterization of a non-repetitive polypeptide derived from silk fibroin heavy chain

Gaoqiang Yang, Mingyang Wu, Honggen Yi, Jiannan Wang *

National Engineering Laboratory for Modern Silk, College of Textile and Clothing Engineering, Soochow University, No. 199 Ren-ai Road, Suzhou Industrial Park, Suzhou, Jiangsu Province 215123, China

ARTICLE INFO

Article history:
Received 25 June 2015
Received in revised form 26 September 2015
Accepted 8 October 2015
Available online 13 October 2015

Keywords:
Silk fibroin
Non-repetitive domain
Gene expression
Purification
Mass spectrometry
Molecular conformation
Charge property

ABSTRACT

Silk fibroin heavy chain is the major protein component of *Bombyx mori* silk fibroin and is composed of 12 repetitive and 11 non-repetitive regions, with the non-repetitive domain consisting of a hydrophilic polypeptide chain. In order to determine the biomedical function of the non-repetitive domain or potentially use it to modify hydrophobic biomaterials, high-purity isolation is necessary. Previously, we cloned and extended a gene motif (f(1)) encoding the non-repetitive domain. Here, this motif and its multimers are inserted into a glutathione S-transferase (GST)-tagged fusion-protein expression vector. Motif f(1) and multimers f(4) and f(8) were expressed in *Escherichia coli* BL21 cells following isopropyl β-D-1-thiogalactopyranoside induction, purified by GST-affinity chromatography, and single bands of purified fusion proteins GST-F(1), GST-F(4), and GST-F(8), were visualized by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Target polypeptides F(1), F(4), and F(8), were cleaved clearly from the GST-fusion tag following thrombin digestion. Mass spectrometry results indicate that the molecular weights associated with fusion proteins GST-F(1), GST-F(4), and GST-F(8) are 31.5, 43.8, and 59.0 kDa, respectively, and with the cleaved polypeptides F(1), F(4), and F(8) are 4.8, 16.8, and 32.8 kDa, respectively. The F(1), F(4), and F(8) polypeptide chains are negatively charged with isoelectric points (pI) of 3.3, 3.2, and 3.0, respectively. The molecular weight and pI values of the polypeptide chains are consistent with the predicted values and the amino acid compositions similar to predicted sequences. FTIR and CD results show the molecular conformation of F(1) was mainly random coil, and more stable α-helix structure formed in longer molecular chain.

1. Introduction

Silk fibroin heavy chain is the major protein component of silk fibroin and is composed of 12 repetitive and 11 nonrepetitive regions arranged alternately [1]. Highly-repetitive regions begin with the protein sequence GAGAGS, repeated 15 times, and end with GAAS. While the residue sequence of the non-repetitive region is relatively stable, the amino acid composition and secondary-structure characteristics vary significantly between repetitive and non-repetitive regions, resulting in different effects on overall structure and function. Given increased silk fibroin use in biomaterials research [2–5] and other promising applications, we are interested in determining the functions of silk fibroin polypeptides.

A variety of fibroin-like peptides have been characterized using chemical synthesis technology. Modeling the $(GSGAGA)_n$ consensus motif from the repetitive region of *Bombyx mori* silk fibroin suggests that it may exhibit β-turn and β-helix conformations, based on energy computations undertaken by Lazo et al. [6]. The model polypeptide $[Ala–Gly]_n$ (n = 12 and 5–9) has also been synthesized for structural analysis [7]. Silk protein-like multi-block polymers derived from the crystalline region of spider dragline silk and *B. mori* silk have been designed and are capable of spontaneous aggregation into β-sheet structures in solid state [8–10]. Fibroin-derived hexapeptides (GAGAGS, GAGAGY, GAGYGA, GAGAGA, GAGVGA, and GAGVGY) stimulate glucose transport and display beneficial effects in diabetes treatment [11]. Additionally, 3T3-L1 pre-adipocyte cell proliferation improved in culture medium when treated with the GAGAGS hexapeptide, as compared to that observed in normal cell culture medium [12].

The polypeptide chains of repetitive and non-repetitive regions contain dozens of amino acids. Given the expense and difficulty of chemically synthesizing high molecular weight peptides, as well as their inability to be regarded as silk fibroin analogs, this method is unsuitable for the systematic study of the relationships between sequences and function. Prokaryotic expression systems using *Escherichia coli* represent a protein-expression system [13,14] applicable for expressing

Abbreviations: IPTG, isopropyl-β-d-thiogalactoside; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; GST, glutathione S-transferase; f(4), four repeats of designed gene motif f(1); f(8), eight repeats of designed gene motif f(1); F(1), expression product of f(1); F(4), expression product of f(4); F(8), expression product of f(8); GST-F(1), expression product of recombinant combining with GST and F(1); GST-F(4), expression product of recombinant combining with GST and F(4); GST-F(8), expression product of recombinant combining with GST and F(8); pI, isoelectric point; *E. coli*, *Escherichia coli*.

* Corresponding author.
E-mail address: wangjn@suda.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.msec.2015.10.023
0928-4931/

high molecular weight, heterologous proteins [15]. A series of block copolymers containing repeating sequences from silk (GAGAGS) and elastin (GVGVP), were synthesized by replacing a valine residue with glutamic acid and expressed for sensitivity analysis using genetic engineering techniques [16].

We have designed and cloned gene sequences encoding the repetitive regions $(GAGAGX)_{16}$ (X = A, S, V, Y) of silk fibroin heavy chain using genetic engineering techniques and successfully expressed these polypeptide chains in *E. coli* BL21 cells. Our results indicate that the polypeptides aggregate into different molecular conformations through self-assembly [13,17]. High levels of a type of silk-elastin-like polymer, consisting of repetitive silk fibroin motifs (GAGAGS) and mammalian elastin (VPGVG), were observed in *E. coli* by Machado et al., eliciting no cytotoxicity and slightly enhancing cell viability [18]. Recombinant spider silk protein (284.9 kDa) has been produced in *E. coli*, isolated, and spun into fiber [15], while recombinant spider silk protein containing a cell-binding motif (RGD) promotes early cell adherence through formation of stress fibers and distinct focal adhesion points [19].

While peptide synthesis from the silk fibroin heavy chain repetitive region has been reported, similar efforts related to non-repetitive regions have not been undertaken. We previously designed, cloned, and extended a gene motif, f(1), encoding a non-repetitive polypeptide [20]. Here, we present an optimized method for expressing f(1), a tetramer, f(4), and an octamer, f(8), with a GST-tag in a prokaryotic system.

2. Materials and methods

2.1. Protein expression

The gene motif, f(1), and all multimers [20] were inserted into a pGEX-AgeI (derived from pGEX-KG) [13] vector to construct the expression vectors pGEX-f(n). The expression vectors pGEX-f(1), pGEX-f(4), and pGEX-f(8) were transformed into *E. coli* BL21(DE3) cells and cultured on Luria–Bertani (LB) solid medium containing 100 μg/mL ampicillin at 37 °C for 12–16 h. A single colony was used to inoculate 4 mL LB medium and cultured at 37 °C overnight, and was then amplified in 250 mL of fresh LB containing ampicillin. Different initial densities (OD_{600} = 0.3–2.1 AU, at 0.3 AU intervals) and different IPTG concentrations (0–1.0 mM, at 0.1 mM intervals) were tested to optimize the fusion proteins GST-F(1), GST-F(4), and GST-F(8) expression levels. At different time points between 0 and 8 h following IPTG induction, cells were harvested by centrifugation at 4 °C and then stored at −80 °C.

2.2. Protein purification

Fusion proteins were purified using the GST-affinity purification system (Novagen) as previously described [13]. Briefly, the cell pellet was suspended in 25 mL GST-bind/wash buffer and sonicated on ice (JYD-900, Shanghai Zhixin instrument Co., Ltd., China). The lysate was centrifuged at 4 °C and the supernatant was loaded onto a 4 mL GST-affinity column and washed with 50 mL GST-wash buffer. Finally, the fusion protein was eluted with 20 mL GST-elution buffer containing 10 mM reduced glutathione.

2.3. Molecular-weight determination

Molecular weight was determined qualitatively by SDS-PAGE as described previously [13]. Briefly, 5 μL loading buffer was added to 20 μL 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-Aldrich) and stained using Coomassie brilliant blue. The molecular weight was qualitatively analyzed by referring to the protein molecular weight standard.

After desalting by ZipTip C4 (Millipore), quantitative analysis of fusion-protein molecular weight was performed using a 4800 MALDI-TOF/TOF mass spectrometer (MS) (AB SCIEX, Redwood City, CA, USA).

2.4. Cleavage of fusion proteins

After ultrafiltration using Amicon Ultra-0.5 spin filters (Millipore) to remove glutathione and salt, per milligram purified fusion protein was digested with 1 U thrombin (Novagen) in lysate buffer at 20 °C for 16 h. The reaction mixture was diluted in GST-bind buffer and loaded onto a GST-affinity column to remove the GST-tag and collect target polypeptides. Samples were freeze-dried.

2.5. Amino acid composition measurement

Fusion proteins were diluted to 0.05 mg/mL and filtered through a 0.22 μm filter to remove impurities. Peptide bonds were hydrolyzed using 6 mol/L HCl and free amino acids derivatized using phenyl isothiocyanate (P1034-1ML, Sigma-Aldrich) and measured using an LC-20A high-performance liquid chromatograph (Shimadzu Crop., Japan).

2.6. Structural assay

FTIR analysis of F(1), F(4) or F(8) polypeptide after prepared in KBr pellets was performed using a FT/IR-5700 spectrometer (Nicolet, USA). The spectra were recorded with a range of 600–2000 cm^{-1} at a resolution of 4 cm^{-1}.

The ellipticities of 0.1 mg/ml purified F(1), F(4) or F(8) polypeptide solutions were obtained by a J-815 CD spectrometer (Jasco, Japan), using a 1.0 mm path-length cell at 25 °C with a accumulation time of 4 s at the rate of 100 nm min^{-1}. A blank solution was measured under the same conditions and subtracted from the spectra date.

2.7. Charge assay

The ζ-potential of each sample was measured using a ZS90 Zetasizer Nano (Malvern Instruments, Malvern, UK) in 5 mM sodium phosphate buffer with 5 mM NaCl at 25 °C. The pH of the buffer was adjusted to 2.0, 3.0, 4.0, 5.0, 6.0, and 7.0, using NaOH or HCl.

3. Results and discussion

3.1. Expression vector characteristics

A commercially available expression vector, pGEX-KG, was chosen for protein expression (Fig. 1A). Genes were expressed under control of the Ptac promoter, translation-enhancing sequence (g10), and ribosome-binding site for regulation of translation level (Fig. 1B). A GST-tag was used to purify expression products and a protease (thrombin) recognition site (LeuValProArgGlySer) was inserted to enable target-peptide release from the GST-tag by cleaving the amide linkage between Arg and Gly. The initiating codon, ATG, was placed upstream of the GST gene and the termination codon, TGA, was followed by the target gene, f(n).

The expression vectors pGEX-f(1)-pGEX-f(12) containing 1–12 gene-motif repeats were qualitatively identified by nucleic acid electrophoresis following *Bam*HI–*Hin*dIII enzyme restriction digest. Two bands were excised from each expression vector, with one representing the segment (4943 bps) of the expression vector and the other the repetitive gene-motif fragment (Fig. 2). The sizes of the repetitive fragments excised from pGEX-f(1)-pGEX-f(12) exhibited step increases in accordance with the expected values of 144, 276, 408, 540, 1068, and 1596 bps. Additionally, the sequences of the genes excised from expression vectors pGEX-f(1)-pGEX-f(12) were verified by nucleotide sequencing (Invitrogen, data not shown).