A directed self-assembly quasi-spider silk protein expressed in *Pichia pastoris*

Bin Liu, Tao Wang, Liyan Xiao, Guilan Zhang, Guangshen Li, Jingzhi Luo and Xiaobing Liu

**ABSTRACT**

The spider silk protein has distinctive physical, chemical, mechanical and biological properties. As a functional material, the application value of spider silk has increased in many fields. Consequently, considerable progress has been made in the expression of recombinant spider silk proteins through many host systems by gene engineering techniques. However, the mechanical properties of the silk fibre spun with the recombinant spider silk proteins are unsatisfactory because the recombinant spider silk proteins have too low of a molecular weight and do not have molecular orientation. This paper describes the construction and expression of a quasi-spider silk protein composed of spider silk protein and collagen-like peptides with the *Pichia pastoris* expression system. The quasi-spider silk protein is an ‘ABA-type’ triblock copolymer composed of triple helix-forming A blocks at both ends of the middle section. The triple helix-forming A blocks at both ends of the triblock copolymers consist of (Pro–Gly–Pro)$_n$ homopolymeric stretches, and the middle section of the molecule (B section) contains a spider silk protein that has been optimally designed. The supramolecular structure created by the three-block copolymers through directed self-assembly ensures that the artificial spider silk fibres will meet the requirement for molecular weight and definite molecular orientation, thus promoting the formation of silk fibres. The authors hope that this project will contribute to the study of materials science and biomedical engineering in regards to the huge potential of spider silk protein in these fields.

**Introduction**

Since ancient times, spider silk has captivated mankind’s interests due to its unique combination of strength and elasticity. It is one of the toughest natural fibres known to man [1,2], and its excellent mechanical properties have the potential to produce high-performance composite materials, which can be used in industry, military, aerospace and other fields [3,4]. Due to its good biocompatibility and biodegradability, as a biomedical material, the application of spider silk has been increasingly explored, e.g. in controlled drug delivery, surgical sutures, scaffolds in tissue engineering and organ transplants [2,5,6].

In contrast to the mulberry silk worm (*Bombyx mori*), which can be used for large-scale silk production, spiders are territorial and produce very little silk. As a result, the research and application of spider silk are still limited [7]. Therefore, it is particularly important to research the recombinant production of spider silk proteins. In recent years, studies focused on the sequences and structures of spider silk protein have resulted in a lot of progress in this area. However, because silk protein sequences are longer and have more repeats, the gene sequences obtained are mostly incomplete. In addition, for cDNA sequences encoding silk proteins, only partial information of the gene structure has been obtained. Moreover, there needs to be further clarification on the relationship between the structure and frequency of repeated sequences in different spider silk proteins and their properties. Also, there needs to be further study into the effect of non-repetitive C-terminus and N-terminal domains on the gene transcription, translation and post-translational processing of spider silk protein; their effect on the secondary structure and solubility of spider silk proteins and their role in the mechanical properties of spider silk fibre [7].

Meanwhile, most of the current studies primarily use MA silk and Flag silk as research objects to express the tandem recombinant spider silk protein by transformation and splicing its characterized modules [2,7]. Researchers have obtained spider silk protein genes by chemical synthesis and cloning techniques from a cDNA library of spider silk glands, and have selected different expression systems, such as plants, animals and...
microbial expression systems, to express recombinant spider silk protein [8]. Based on the foregoing, researchers have tried to spin superior performance spider silk protein fibres [8]. Plant expression systems have special advantages in terms of gene expression, modification and safety, and plants can make use of natural nutrition to manufacture their own amino acids. At present, recombinant spider silk proteins have been successfully expressed in Arabidopsis (Arabidopsis thaliana L.), potato (Solanum tuberosum L.), tobacco (Nicotiana tabacum L.) and other plants [9]. Through the prokaryotic expression system, cells can avoid the defects of early termination of transcription and translation in the processes of spider silk protein gene expression, and the expression level has been greatly improved as compared to other systems [10]. Nonetheless, it is difficult to culture cells, and large-scale productions of animal cells and plants cells expression systems are expensive. Therefore, on account of its relatively mature technology and relatively low cost for large-scale production, the microbial fermentation method has generated a lot of attention.

With the continuous progress of technical methods for recombinant protein expression, the size of recombinant spider silk protein effectively expressed by the Escherichia coli expression system has been greatly improved. In 2016, Weichert et al. [11] expressed a spider silk protein FLAG by way of intein-based trans-splicing, and the molecular weight of this routinely produced multimer is larger than 460 kDa, which significantly exceeds the size of the native FLAG protein.

In-depth studies have been conducted on using the Pichia pastoris expression system to produce recombinant spider silk proteins [12]. This expression system can avoid the general shortcomings of the bacterial expression system and can ensure that the expression product adequately undergoes the post-translational modification process, including hydroxylation and glycosylation.

Great progress has been made in the production of recombinant spider silk protein through the expression system mentioned above; however, some problems still remain [13]. For example, the molecular weights of products expressed by the recombinant spider silk protein gene are usually not large enough as compared to natural spider silk protein. Additionally, the recombinant spider silk protein generally does not have molecular orientation, and a simulation of the fibre formation process of natural spider silk cannot be fully realized. The expression efficiency is still relatively low, and the stability of the expression system is still a problem. For these reasons, the practical applications of spider silk proteins have rarely been reported.

Another unique biological macromolecule, collagen, has comprehensive applications in biological medicine, tissue engineering, food, cosmetics and other fields due to its good biocompatibility, mechanical properties, biodegradability, low immunogenicity and other good characteristics [14]. The most typical feature of collagen is the three-helix structure consisting of three polypeptide chains, which are called α-chains. The three helical α-chains are wrapped around each other to constitute a triple helix. Another major feature of collagen is the repeated presence of –Gly–Xaa–Yaa– sequences, in which glycine is the exclusive amino acid that can be adapted to the limited space available in the centre of the collagen helix. Additionally, there is a high proportion of proline and lysine residues in the repeating –Gly–Xaa–Yaa– units [15].

The function of the collagen triple helix structure as a ‘sticky-ended’ fragment has been explored. For example, Werten et al. [16] reported two gelatins that can be used to produce customized precision gels for biomedical applications. These gelatins are nonhydroxylated triblock copolymers with terminal trimerizing blocks. Because only the end blocks can form the three-helix structure, the molecular structure of these gelatins is much more defined than that of traditional gelatin [16].

Here, we describe a quasi-spider silk protein expressed by P. pastoris, which is a fusion protein consisting of the spider silk protein and a collagen-like peptide, which is an ‘ABA-type’ triblock copolymer. Both ends (A sections) consist of (Pro–Gly–Pro)n, homopolymeric stretches, and the middle section (B section) of the molecule is an optimization-designed spider silk protein. The two A sections are ‘sticky-ended’ fragments, and their function is to directionally form the triple helix structure, which gives the whole molecule a certain molecular orientation and leads to the formation of a supramolecular structure. The supramolecule, formed by directed self-assembly, can ensure that artificial spider silk fibres will meet their required molecular weight and certain molecular orientation, thus promoting the formation of silk fibres. We hope this study will contribute to medical engineering and other related fields in regards to the huge potential of spider silk protein.

Materials and methods

Plasmids and strains

Expression vector pPIC9K, P. pastoris strain GS115 and E. coli strain Top10F’ were obtained from Shanghai Sangon (Shanghai, P.R. China). The pPIC9K plasmid contains a Saccharomyces cerevisiae α-mating factor (α-MF) secretion signal, an alcohol oxidase 1 (AOX1) promoter/terminator cassette, a HIS4 selectable marker and a Tn903 insert.
An EcoI linearized with another independent reaction system, the vector was TOP10F' by the CaCl2 method. Then, a transformant that the ligation product was immediately used to transform linearized vector to yield vector pUC57-A1S2. Thereafter, EcoRI by digesting pUC57-A1S with Figure 2. First, the monomeric gene A1S was produced The process of constructing pPIC9K-TS2T is described in Construction of pUC57-A1TS2TA2 cloning vector The multimeric gene obtained by digesting pUC57-A1TS2TA2 with EcoRI/Van91I. In another reaction system, the vector pUC57-A1S2 was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1T) was then inserted into this linearized vector to yield vector pUC57-A1TS2.

Similarly, the A1T fragment was produced by digesting pUC57-A1TA2 with EcoRI/Van91I. In another reaction system, the vector pUC57-A1S2 was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1T) was then inserted into this linearized vector to yield vector pUC57-A1TS2.

Figure 2. First, the monomeric gene A1S was produced The process of constructing pPIC9K-TS2T is described in Construction of pUC57-A1TS2TA2 cloning vector The multimeric gene obtained by digesting pUC57-A1TS2TA2 with EcoRI/Van91I. In another reaction system, the vector pUC57-A1S2 was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1T) was then inserted into this linearized vector to yield vector pUC57-A1TS2.

Similarly, the A1T fragment was produced by digesting pUC57-A1TA2 with EcoRI/Van91I. In another reaction system, the vector pUC57-A1S2 was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1T) was then inserted into this linearized vector to yield vector pUC57-A1TS2.

Finally, the A1TS2 fragment was produced by digesting pUC57-A1TS2 with EcoRI/Van91I. In another reaction system, the vector pUC57-A1TA2 was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1TS2) was then inserted into this linearized vector to yield vector pUC57-A1TS2TA2.

**Design of quasi-spider silk protein gene**

The monomeric spider silk gene (referred to hereafter as ‘S’) was designed based on spider dragline silk gene sequences (GenBank accession numbers AY555585 and AH015065). The new designed gene sequence is presented in Figure 1(A). It was properly modified and optimized according to the structural characteristics of spider silk protein and *P. pastoris* bias usage. The protein secondary structure was predicted with the Garnier method in Antherprot software. Both ends of the gene have the recognition sites of isocaudamers DraIII and Van91I, respectively, and the A1 adapter that includes EcoRI recognition sites was arranged at the 5’ end. Then, the whole gene was designated as A1S, and its length is 397 bp.

The triple helix-forming T blocks at both ends of the triblock copolymers consist of (Pro-Gly-Pro)n, homopolymeric stretches (Figure 1(B)). The length of n was tentatively chosen to be 15, in line with referenced literature [16]. The A1 and A2 adapters were arranged at both ends of the gene; the A1 adapter that includes the EcoRI recognition sites was arranged at the 5’ end, whereas the A2 adapter that includes NotI recognition sites was arranged at the 3’ end. Then, the whole gene was designated as A1TA2, and its length is 198 bp.

Both A1S and A1TA2 were synthesized with a chemical method and were then cloned in vector pUC57 to form vector pUC57-A1S and pUC57-A1TA2, respectively.

**Construction of pUC57-A1TS2TA2 cloning vector**

The process of constructing pPIC9K-TS2T is described in Figure 2. First, the monomeric gene A1S was produced by digesting pUC57-A1S with EcoRI/Van91I (Figure 2). In another independent reaction system, the vector was linearized with EcoRI/DraIII and dephosphorylated. The EcoRI/Van91I fragment (A1S) was then inserted into this linearized vector to yield vector pUC57-A1S2. Thereafter, the ligation product was immediately used to transform TOP10F’ by the CaCl2 method. Then, a transformant that contained the recombinant vector pUC57-A1S2 was selected from among colonies growing on Luria–Bertani medium plates with 0.1 mg/mL of Amp by extracting the vectors and examining their size via agarose gel electrophoresis. The transformant was designated Top10F’/pUC57-A1S2. In principle, this operation of insertional doubling can be repeated to form multimers of any desired length.

Multi-copy Mut+ pPIC9K-TS2T transformants were selected based on G418 resistance selection. The screening protocol is derived from the Multi-Copy *Pichia* Expression Kit Manual, version F (Invitrogen, Carlsbad,
A

1 5’- CGGGAATTCCTCGAGAAGAGGTCACCCGGTCACTTGGTACCAGGTCAACAGGTAAG  
10 G G Y G P G Q Q G P S G P G S G A A A A  
60 GGTGTGTACGGCTCAGGTCACACAGGGTCCATCTGGTCTCCAGGTCAACAGGTAAG  
30 A G G G A G Q G G Y G L G G Q G A G R  
120 GCTGCTGGTGGGTCTCAAGGTTGTAAGGTTGGGTTGGGTCAAGGTTGCTGCTGCTGTA  
50 G G L G G Q A A A A A A S G P G G Y G P  
180 GGTGTGTCCGGTGTCCAACCTGCCTGCTGCTGCTGCTTCTGTGGTCAGGTGTTAGTGCTCA  
70 G Q G Q G P G G Y G P G Q Q G P S G P S  
240 GGTCAACAGGGTCAGGTTACGGCTCAGGTCACACAGGGTCCATCTGGTCCAGGTCT  
90 M A A A A A A A A M G P G G Y G P Q Q S  
300 ATGGTCTGCTGCTGCTGCTGCTATGGTCCAGGTGTTAGGTCAGGGTTACGGTCAGGTCAACAGGTA  
110 P G G Y G S P A  
360 CCAAGGTGTTCACGGTTTCACACCGGGTGTTAAGA - 3’

B

1 5’- CGGGAATTCCTCGAGAAGAGGTCACCCGGTCACTTGGTACCAGGTCAACAGGTAAG  
10 G G Y G P G Q Q G P S G P G S G A A A A  
60 GGTGTGTACGGCTCAGGTCACACAGGGTCCATCTGGTCTCCAGGTCAACAGGTAAG  
30 G G G G A G Q G G Y G L G G Q G A G R  
120 GCTGCTGGTGGGTCTCAAGGTTGTAAGGTTGGGTTGGGTCAAGGTTGCTGCTGCTGTA  
50 G G L G G Q A A A A A A S G P G G Y G P  
180 GGTGTGTCCGGTGTCCAACCTGCCTGCTGCTGCTGCTTCTGTGGTCAGGTGTTAGTGCTCA  
70 G Q G Q G P G G Y G P G Q Q G P S G P S  
240 GGTCAACAGGGTCAGGTTACGGCTCAGGTCACACAGGGTCCATCTGGTCCAGGTCT  
90 M A A A A A A A A M G P G G Y G P Q Q S  
300 ATGGTCTGCTGCTGCTGCTGCTATGGTCCAGGTGTTAGGTCAGGGTTACGGTCAGGTCAACAGGTA  
110 P G G Y G S P A  
360 CCAAGGTGTTCACGGTTTCACACCGGGTGTTAAGA - 3’

Van911

Van911

No1

Figure 1. Construction of the quasi-spider silk protein gene. The monomeric spider silk gene (A); the triple helix-forming T blocks (B) at both ends of the triblock copolymers consisting of (Pro–Gly–Pro)\textsubscript{15} homopolymeric stretches.

CA, USA). The copy number of pPIC9K-TS2T in the transformants may be estimated by its growth in yeast extract peptone dextrose (YPD) medium containing a gradient concentration of G418.

**Growth and induction of multi-copy transformants in shaking flasks**

Multi-copy transformants were cultured in shaking flasks for 24 h in buffered minimal glycerol medium (BMG, containing 0.4 mg/L biotin, 1.34% yeast nitrogen base without amino acids and 1% glycerol and supplemented with 100 mmol/L potassium phosphate, pH 6.0) at 28 °C. The cells were harvested and re-suspended in buffered minimal methanol medium (BMM, identical to BMG but containing 0.5% methanol instead of glycerol) to an optical density at 600 nm (OD\textsubscript{600}) of 1.0. Then, the recombinant yeasts were grown and sampled for three days at 28 °C and 220 r/min with 0.5% methanol added every day.
Fermentative production of quasi-spider silk protein in recombinant yeast

The fermentation was carried out in accordance with Invitrogen guidelines in a 5.0 L fermentor that was equipped with oxygen supply units and with microprocessor control of temperature, dissolved oxygen, pH, agitation and nutrient feed. The fermentor, containing 2 L of culture medium composed of 4% glycerol and Fermentation Basal Salts, was sterilized at 120 °C for 20 min. After that, 2.4 mL/L salts (Invitrogen, PTM1) and strains with approximately 5%–10% initial fermentation volume cultured in buffered glycerol complex medium (BMGY) were added by aseptic techniques. Agitation was controlled at 600 r/min, pH 6.0 was controlled with 25% ammonium hydroxide, temperature was maintained at 28 °C and the dissolved oxygen level was maintained at 25%. The glycerol batch phase was maintained for about 24 h until the glycerol was completely consumed. A 50% w/v glycerol feed containing 12 mL of PTM1 trace salts per litre of glycerol feed was initiated, and the feed rate was set to 18 mL/h per litre of initial fermentation volume. Glycerol feeding was continued until a cellular yield of 180–220 g/L of wet cells is reached. The glycerol feed was terminated, and induction was initiated by starting a

Figure 2. Construction of yeast expression vector pPIC9K-TS2T.
100% methanol feed containing 12 mL of PTM1 trace salts per litre of methanol. The feed rate was set to 3.6 mL/h per litre of initial fermentation volume. Throughout the whole process, the methanol feed was automatically adjusted according to the level of dissolved oxygen.

After fermentation, the yeast cells and the supernatant were separated by centrifugation at 7155g for 8 min (Invitrogen Kit). The obtained cell pellets were disrupted by sonication to give a total protein solution, and the supernatant was collected for the detection of extracellular secreted quasi-spider silk protein.

**Gel filtration chromatography**

The sample was loaded on a 13 × 350 mm column. Sephadex G-100 (Amersham Biosciences, Piscataway, NJ, USA) was selected as the filtration medium. The elution procedure was carried out with deionized water at a flow-rate of 0.1 mL/min. Two-millilitre fractions were collected, and the absorbance at 215 nm was monitored.

**SDS-PAGE analysis of secreted production of induced fermentation**

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed on the fermentation supernatant, which was obtained by centrifuging the samples collected at the inducing phase. The gel consisted of a 5% stacking zone and a 10% separating zone. The initial voltage was set to 100–120 V, when the dye migrated into the separating zone for about 30 min, the voltage increased to 200–220 V, and electrophoresis continued until the dye was 1 cm away from the bottom of the gel. After electrophoresis, the gel was stained with Coomassie Bright Blue R-250.

**Mass spectrometry**

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed with AB SCIEX 5800 MALDI-TOF/TOF. The samples were prepared by the dried droplet method, with sinapinic acid dissolved in 30% (v/v) acetonitrile and 4% (v/v) trifluoroacetic acid as the matrix. Measurements were made in the positive, linear mode and the accelerating voltage was 25000 V. Before the relative molecular weight of the sample was tested, the sample target was calibrated by testing external calibrants. The original data generated by 5800 MALDI-TOF/TOF were exported by software 4000 Series Explorer V3.5.

**Results and discussion**

**Gene synthesis, construction and verification of expression vector pPIC9K-TS2T**

According to the process in Figure 2, the expression vector pPIC9K-TS2T was successfully constructed. Then, a single-enzyme digestion and double-enzyme digestion identification experiment were conducted to extract the Top10F'/pPIC9K-TS2T plasmid DNA.

In the agarose gel electrophoresis of the single-enzyme digestion (Figure 3), the size of the linearized plasmid pPIC9K-TS2T was 1 kb (calculated by Quantity One, version 4.5.2, Bio-Rad), which matches very well with the theoretical size of pPIC9K-TS2T (10,302 bp). Also, the recombinant expression vector pPIC9K-TS2T was digested by EcoRI/NotI and analysed by agarose gel electrophoresis (Figure 4), and the result corresponded to the theoretical values of 9263 and 1039 bp of the two DNA fragment sizes.

Furthermore, according to the DNA sequencing report from Sangon (data not shown), the recombinant quasi-spider silk protein gene was successfully inserted into pPIC9K through adapters A1 and A2, and the insertion site immediately follows the α-mating factor (α-MF).
secretion signal gene, which ensures secretion expression of heterologous protein. Moreover, the AOX1 strong promoter (PAOX1) in pPIC9K can be induced in the presence of methanol so as to achieve high expression level of foreign proteins.

**Selection and verification of recombinant yeast**

After patching more than a thousand His^+^ transformants on YPD plates containing different concentrations of G418, 12 colonies of recombinant yeast were obtained that could grow well on YPD plates containing 4 mg/mL G418. Colony PCR was performed, and the production was analysed by agarose gel electrophoresis (Figure 5). The results from the experiment demonstrated that all of the 12 colonies were positive transformants. These successfully constructed that pPIC9K-TS2T transformants have a 1.5 kbp product containing the gene of interest of the inserted TS2T (1.0 kbp, including A1 and A2) and the flanking AOX1 sequence from GS115/pPIC9K (about 0.5 kbp). Additionally, the pPIC9K transformant has a 0.5 kbp product, which represents the AOX1 sequence of pPIC9K, whereas GS115 without any vector has a 2.2 kbp product representing the wild-type AOX1 gene.

**Analysis of large-scale expression product of multi-copy recombinant yeast**

After colony PCR, the induced expression levels of all the multi-copy recombinant yeast with multiple inserts of pPIC9K-TS2T were tested with SDS-PAGE (data not shown).
shown). Then, the recombinant yeast with the highest expression level was chosen for the large-scale expression experiment. The experimental results are shown in Figure 6.

The result from the SDS-PAGE of the fermentation supernatant and the sample purified with Sephadex G-100 Gel filtration chromatography showed that recombinant yeast can secrete the designed quasi-spider silk protein when induced by methanol. From the SDS-PAGE gel (Figure 6), it can be observed that the samples exhibit two bands: a faint blurry band (marked by the left arrow) at 34.0 kDa and a vivid band (marked by the right arrow) at 61.0 kDa. We speculate that the faint blurry band belongs to the protein of interest. The theoretical molecular weight of the protein of interest was 28.0 kDa, which is less than its apparent molecular weight. Because the quasi-spider silk protein is a fusion protein containing two sections of collagen-like peptides, we consider this a reasonable speculation because some studies show that collagenous proteins migrate in SDS-PAGE at an apparent molecular weight that is ~40% higher than the true molecular weight [17,18].

Several possible explanations for the aberrant migration behaviour of gelatins have been suggested [17]. Moreover, we speculate that the dense band is a homodimer formed by two quasi-spider silk protein monomers, and the quasi-spider silk protein is primarily presented in homodimer form. This is also due to the collagen-like peptides sections in the quasi-spider silk protein because it can form hydrogen bond connections between two or more collagens [14]. This is exactly the characteristic we expect the quasi-spider silk protein to have.

**Mass-spectrographic analysis of quasi-spider silk protein**

The purified samples were analysed by MALDI-TOF-MS to determine the exact molecular weight. The experimental results are shown in Figure 7. In Figure 7, the exact molecular weight of the quasi-spider silk protein is 29,975 Da. The theoretical design of the quasi-spider silk protein contains 339 amino acids, and its molecular weight is 28,198 Da. The actual accurate molecular weight is slightly larger than the theoretical value, which can be traced to the O-linked and N-linked glycosylation modification when proteins are secreted from *P. pastoris* [19]. However, the side chains of the mammalian cells and the *P. pastoris* yeast that extend from the sugar nucleus are not the same; thus, its glycosylation process is very complicated. Therefore, it is impossible to accurately calculate how many sites have been glycosylated and how much the glycosylation contributes to an increase in the molecular weight of each site.

At present, the research on recombinant spider silk proteins is generally based on dragline sequence motifs, such as poly-alanine, GGX and GPGXX. With chemical and genetic modifications to these silk proteins, their properties can be improved and their potential applications can be broadened. Relative studies include the introduction of disulphide bridges derived from engineered disulphides to improve the mechanical properties of recombinant spider silk [20], the integration of nanostructured carbon (e.g. carbon nanotubes, nanohorns, nanodiamonds, fullerene, etc.) to construct hybrid/composite hydrogels for medicinal use [21], the introduction of natural disorder by means of heterogeneity of bundled fibres to increase the mechanical strength of recombinant spider silk [22] and so on.

Because expressed recombinant spider silk proteins have too small of a molecular weight and they do not have molecular orientation, the goal of this study was to try to remove the defects through the development and
efficient production of block copolymers constructed by collagen-like peptides and recombinant spider silk proteins.

The complete artificial method was employed to obtain a quasi-spider silk protein gene, which was optimized based on the natural spider dragline silk protein gene sequence. Moreover, the monomer tandem repeating strategy made it possible to construct different lengths of recombinant spider silk protein genes with different numbers of monomers. *P. pastoris* was adopted to express this quasi-spider silk protein.

The quasi-spider silk protein is a kind of ‘ABA-type’ triblock copolymer. The triple helix structure A blocks at both ends of the triblock copolymers consist of (Pro–Gly–Pro)$_{15}$ homopolymeric stretches, and the middle section of the molecule (B section) consists of a spider silk protein that has been optimally designed. The characteristics of the three-block copolymers that form the supramolecular structure by directed self-assembly ensure that artificial spider silk fibres will meet their requirement for molecular weight and definite molecular orientation when spinning silk fibres. First, in order to obtain a large fragment or a fragment close to the natural spider silk proteins in molecular weight, the usual practice is to synthesize a silk protein gene monomer first, and then polymerize them to form dimers, trimers or multimer genes to obtain a higher molecular weight expression product. However, the excessive GC content of the spider silk protein gene limits its full expression in each expression system [23]. This quasi-spider silk protein is expected to form the trimer segment of the triple helix structure with its (Pro–Gly–Pro)$_n$ homopolymer extension situated on both ends by way of directional self-assembly. The trimer segment can lead to respective connections of both ends of each peptide chain and two other peptide chains, thereby forming a spatial network structure, which is an organized aggregate that can maintain certain integrity. It can also become a supramolecular with definite microstructure and macroscopic properties and a large amount of molecules. Therefore, the method introduced in this paper can help to produce a recombinant spider silk protein with a higher molecular weight. Moreover, the triple helix structure formed by the (Pro–Gly–Pro)$_n$ homopolymer stretches situated on both ends of the quasi-spider silk protein after orientation arrangement in a head-to-tail arrangement gives the quasi-spider silk protein a definite molecular orientation. The middle silk protein still retains the $\beta$-turn structure and the $\beta$-turn structure is folded together in series to form a spring-like structure to maintain the flexibility of the silk protein fibre [13]. At the same time, the $\beta$-sheet structure can tightly bind the
protein molecules by intermolecular force to form a crystalline region of silk fibres, which can also improve the ability of whole molecules to form a high strength fibre.

**Conclusions**

A 29,975 Da quasi-spider silk protein has been efficiently produced as an extracellular protein by recombinant yeast. It consists of triblock copolymers composed of terminal (Pro–Gly–Pro)₁₅ blocks and a spider silk protein spacer. Specifically, the end blocks can form collagen-like triple helices, which ensure that the spider silk protein can form a supramolecular structure by directed self-assembly in the silk fibres formation process, thereby meeting the molecular weight and molecular orientation requirements of artificial spider silk fibres in the process of silk fibres formation. However, this result still needs validation with further experiments. In subsequent experiments, we will focus on the secondary structure of the quasi-spider silk protein, i.e. its beta-sheet and beta-turn content, by Fourier transform infrared spectroscopy, circular dichroism, etc.

**Acknowledgments**

The authors acknowledge the assistance from Wenli Ma, director of Institute of Genetic Engineering, Southern Medical University, Guangzhou, PR China …

**Disclosure statement**

The authors declare no conflict of interest.

**Funding**

This work was supported by the National Natural Science Foundation of China [grant number 31360228].

**ORCID**

Bin Liu http://orcid.org/0000-0001-7115-913X
Tao Xiao http://orcid.org/0000-0003-4467-945X
Liyan Xiao http://orcid.org/0000-0003-4415-7192
Guilan Zhang http://orcid.org/0000-0003-3340-2623
Guangshen Li http://orcid.org/0000-0002-8023-6552
Jingzhi Luo http://orcid.org/0000-0003-2540-0441
Xiaobing Liu http://orcid.org/0000-0001-7623-0804

**References**

[1] Bluem C, Nichtl A, Scheibel T. Spider silk capsules as protective reaction containers for enzymes. Adv Funct Mater. 2014;24(6):763–768.
[2] Bittencourt DMDC. Spider silks and their biotechnological applications. Short Views Insect Genomics Proteomics. 2016;4:211–227.
[3] Hinman MB, Teulé F, Perry D, et al. Modular spider silk fibers: defining new modules and optimizing fiber properties. In: Asakura T, Miller T, editors. Biotechnology of silk. Vol. 5. Netherlands: Springer; 2014. p. 137–164. (Biologically-Inspired Systems).
[4] Chiasson R, Hasan M, Nazer Q A, et al. The use of silk in nanomedicine applications. In: Howard KA, Vorup-Jensen T, Peer D, editors. Nanomedicine. New York (NY): Springer; 2016. p. 245–278.
[5] Thatikonda N. Functionalization of partial spider silk with affinity domains and its use for diverse applications [licentiate thesis]. Uppsala: Sveriges lantbruksuniv; 2016.
[6] Chung H, Kim TY, Sang YL. Recent advances in production of recombinant spider silk proteins. Curr Opin Biotechn. 2012;23(6):957–964.
[7] Xia XX, Demain AL. Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. Proc Natl Acad Sci USA. 2010;107(32):14059–14063.
[8] Weichert N, Hauptmann V, Menzel M, et al. Transglutamination allows production and characterization of native-sized ELPylated spider silk proteins from transgenic plants. Plant Biotechnol J. 2014;12(2):265–275.
[9] Peng CA, Russo J, Gravgaard C, et al. Spider silk-like proteins derived from transgenic Nicotiana tabacum. Transgenic Res. 2016;25(4):1–10.
[10] Teulé F, Miao YG, Sohn BH, et al. Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. Proc Natl Acad Sci USA. 2012;109(3):923–928.
[11] Weichert N, Hauptmann V, Helmold C, et al. Seed-specific expression of spider silk protein multimers causes long-term stability. Front Plant Sci. 2016 [cited 2016 Jan 28];7:6. DOI:10.3389/fpls.2016.00006.
[12] Jansson R, Lau CH, Ishida T, et al. Functionalized silk assembled from a recombinant spider silk fusion protein (Z-4RepCT) produced in the methylotrophic yeast Pichia pastoris. J Biotechnol. 2016;11(5):687–699.
[13] An B, Hinman MB, Holland GP, et al. Inducing β-sheets formation in synthetic spider silk fibers by aqueous post-spin stretching. Biomacromolecules. 2011;12(6):2375–2381.
[14] Gómez-Guillén MC, Giménez B, López-Caballeró ME, et al. Functional and bioactive properties of collagen and gelatin from alternative sources: a review. Food Hydrocolloids. 2011;25(8):1813–1827.
[15] Ramshaw JAM. Biomedical applications of collagens. J Biomed Mater Res B. 2015;172(1):7–11.
[16] Werten MW, Teles H, Moers AP, et al. Precision gels from collagen-inspired triblock copolymers. Biomacromolecules. 2009;10(5):1106–1113.
[17] Butkowsky RJ, Noelken ME, Hudson BG. Estimation of the size of collagenous proteins by electrophoresis and gel chromatography. Method Enzymol. 1982;82C:410–423.
[18] Werten MWT, Bosch TJVD, Wind RD, et al. High-yield secretion of recombinant gelatins by Pichia pastoris. Yeast. 1999;15(11):1087–1096.
[19] Pourcq KD, Schutter KD, Callewaert N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. Appl Microbiol Biot. 2010;87(5):1617–1631.
[20] Iglesias D, Bosi S, Melchionna M, et al. The glitter of carbon nanostructures in hybrid/composite hydrogels for medicinal use. Curr Top Med Chem. 2016;16(18):1976–1989.

[21] Grip S, Johansson J, Hedhammar M. Engineered disulfides improve mechanical properties of recombinant spider silk. Protein Sci. 2009;18(5):1012–1022.

[22] Cranford SW. Increasing silk fibre strength through heterogeneity of bundled fibrils. J R Soc Interface. 2013 [cited 2013 May 6];10(82):20130148. DOI:10.1098/rsif.2013.0148

[23] Fahnestock SR, Bedzyk LA. Production of synthetic spider dragline silk protein in Pichia pastoris. Appl Microbiol Biot. 1997;47(1):33–39.