A novel silk fibroin protein–based fusion system for enhancing the expression of nanobodies in *Escherichia coli*

Jianli Yu1,2 · Yang Guo1 · Yi Gu1,2 · Xiying Fan1 · Fei Li3 · Haipeng Song3 · Rui Nian1 · Wenshuai Liu1

Received: 1 December 2021 / Revised: 22 February 2022 / Accepted: 26 February 2022 / Published online: 4 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Nanobodies show a great potential in biomedical and biotechnology applications. Bacterial expression is the most widely used expression system for nanobody production. However, the yield of nanobodies is relatively low compared to that of eukaryotic systems. In this study, the repetitive amino acid sequence motifs (GAGAGS) found in silk fibroin protein (SFP) were developed as a novel fusion tag (SF-tag) to enhance the expression of nanobodies in *Escherichia coli*. SF-tags of 1 to 5 hexapeptide units were fused to the C-terminus of 4G8, a nanobody against human epididymis protein 4 (HE4). The protein yield of 4G8 variants was increased by the extension of hexapeptide units and achieved a 2.5–7.1-fold increase compared with that of untagged 4G8 (protein yield of 4G8-5C = 0.307 mg/g vs that of untagged 4G8 = 0.043 mg/g). Moreover, the fusion of SF-tags not only had no significant effect on the affinity of 4G8, but also showed a slight increase in the thermal stability of SF-tag-fused 4G8 variants. The fusion of SF-tags increased the transcription of 4G8 by 2.3–7.0-fold, indicating SF-tags enhanced the protein expression at the transcriptional level. To verify the applicability of the SF-tags for other nanobody expression, we further investigated the protein expression of two other anti-HE4 nanobodies 1G8 and 3A3 upon fusion with the SF-tags. Results indicated that the SF-tags enhanced the protein expression up to 5.2-fold and 5.7-fold for 1G8 and 3A3, respectively. For the first time, this study reported a novel and versatile fusion tag system based on the SFP for improving nanobody expression in *Escherichia coli*, which may enhance its potential for wider applications.

Key points

- A silk fibroin protein–based fusion tag (SF-tag) was developed to enhance the expression of nanobodies in *Escherichia coli*.
- The SF-tag enhanced the nanobody expression at the transcriptional level.
- The fusion of SF-tag had no significant effect on the affinity of nanobodies and could slightly increase the thermal stability of nanobodies.

Keywords  Nanobody · *Escherichia coli* · Expression-enhancing tag · Silk fibroin protein · Hexapeptide

Introduction

Nanobodies are derived from the heavy-chain antibodies (HC Abs), which are naturally present in all Camelidae. These HC Abs have an antigen recognition part composed of single variable domains, referred to as the variable domain of the heavy chain of a heavy-chain antibody (VHH). Owing to a small diameter of 2.5 nm and a height of 4 nm, these single variable domains are accordingly named as nanobodies by its original Belgian developer Ablynx®. In recent years, nanobodies have attracted extensive attention in the biotechnology and biomedical industries owing to their nanoscale size, high stability, strong affinity and specificity, water solubility, superior cryptic cleft accessibility, and sustainable
source. With the peculiar properties, nanobodies have been developed into various research tools for basic research, including affinity purification, chaperone-assisted crystallization, protein–protein interaction, and cellular bioimaging (Wang et al. 2016). The small structure of nanobodies facilitates their fast extravasation, deep tumor penetration, and rapid clearance in vivo, which highly benefits tumor-targeted molecular imaging. Several nanobody targeting probes for cancer diagnosis have entered clinic trials (Liu et al. 2021). Apart from the application in vivo tumor imaging, nanobodies have been developed to detect disease biomarkers in the blood or other biopsies. Further, nanobodies have also been developed for therapeutic use. Caplacizumab, a bivalent nanobody, was approved by the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) for the treatment of patients with thrombotic thrombocytopenic purpura (Sculley et al. 2019). Moreover, several nanobody therapeutics are under clinical investigations for a variety of human diseases including breast cancer, lung diseases, brain tumors, inflammatory diseases, and infectious diseases (Jovevska and Mulyldermans 2020). Recently, nanobodies that neutralize SARS-CoV-2 were identified by several independent research groups to assist in combating the pandemic diseases (Custodio et al. 2020; Hanke et al. 2020; Huo et al. 2020; Li et al. 2021; Schoof et al. 2020; Xiang et al. 2020). Hence, the nanobody is considered a next generation of antibody-derived tool in antigen-related recognition, and will occupy a momentous portion of the antibody reagent market.

The expression of recombinant proteins is one of the most critical parts for protein applications in biotechnology and pharmaceuticals as it influences the bioactivity, cost, and yields of products. Compared to conventional antibodies, nanobodies can be expressed in Escherichia coli cells instead of mammalian cells owing to their single gene format as well as non-requirement for post-translational modifications (Liu and Huang 2018). As nanobodies contain a disulfide linkage between complementarity-determining regions (CDRs) 1 and 3, they are generally expressed in the oxidative environment of the E. coli periplasm to facilitate correct folding and disulfide bond formation (Billen et al. 2017; Salema and Fernandez 2013). Thus far, many nanobodies are successfully expressed in the periplasm of E. coli with the strains TG1 (Clontech), Shuffle® T7 (New England Biolabs), Origami (Novagen), Rosetta-gami (Novagen), and BL21 (DE3) (Novagen) (Liu and Huang 2018). However, a large amount of nanobodies is still only expressed at a very low yield due to the insufficiency of chaperones in the periplasm. Although various protein fusion tags (e.g., GST, MBP, Trx, NusA, SUMO, Fh8) have been developed to enhance protein expression and solubility (Ki and Pack 2020), large fusion tags may interfere with the protein structures or impair the affinities of nanobodies. Therefore, short peptide fusion tags to enhance nanobody expression have continued to be investigated and developed.

Silk fibroin protein (SFP) is a natural polymer-based protein from the silkworm Bombyx mori. The B. mori SFP is comprised of a heavy chain (HC) and a light chain (LC) linked by a disulfide bond (Shimura et al. 1976). P25, a 25-kDa glycoprotein, is non-covalently linked to these chains to maintain the integrity of the complex (Tanaka et al. 1999). The LC is hydrophilic and relatively elastic. The HC consists of 12 crystalline domains interspersed with amorphous regions. Each crystalline domain consists of subdomains of ~70 residues, which primarily starts with the repeated glycine-, alanine-, and serine-rich hexapeptides (GAGAGS) and terminates with a GAAS tetrapeptide (Zhou et al. 2001). SFP exhibits numerous attractive properties, such as excellent mechanical property, good biocompatibility, controllable biodegradability, aqueous processability, and ease of functionalization and patterning. Therefore, SFP has been employed in a multitude of biomaterial applications (Kundu et al. 2013). In the study by Wang et al., SFP showed antiproliferative effects on lung cancer cells both in vivo and in vitro by inducing cell apoptosis (Wang et al. 2019).

In this study, we investigated the potential of SFP hexapeptide (GAGAGS) blocks (SF-tag) as an expression enhancement tag for nanobody production in the periplasm of E. coli cells. We selected 4G8, a nanobody against human epididymis protein 4 (HE4) that was screened from an immune library, as the model nanobody (Liu et al. 2018). The SF-tag was fused to the C-terminus of 4G8. The resulting recombinant fusion proteins (termed 4G8-SF-tag fusions) were assessed for the production yields, affinities, protein secondary structures, thermal stabilities, and transcription levels compared to the untagged form. Meanwhile, the expression enhancement effect of SF-tag was also tested on two other anti-HE4 nanobodies 1G8 and 3A3. The results imply that the SF-tag is an excellent and versatile fusion tag to increase the expression levels of nanobodies in E. coli cells.

Materials and methods

Strains, plasmids, and reagents

E. coli strain TG1 (Clontech) was used as the host for phage display. E. coli strain DH5α (Invitrogen) and BL21 (DE3) (Novagen) cells were used as the host for gene cloning and nanobody expression, respectively. HEK293 cells (Thermo Fisher Scientific) were used as the host for HE4-Fc expression. A pMES4 vector containing a signal peptide pelB at the N-terminus of multiple cloning sites (MCSs) was used
for expressing recombinant nanobodies in E. coli. The pcDNA™3.1(+) vector was used for expressing recombinant HE4-Fc protein in HEK293 cells.

DNA molecules were synthesized by Tsingke (China). Lipofectamine 3000 was purchased from Invitrogen (USA). Restriction endonucleases and Gibco® FreeStyle™ 293 Expression Medium were purchased from Thermo Fisher Scientific (USA). Ni Sepharose 6 Fast Flow, HiTrap™ Protein G HP column, and CM5 sensor chip were purchased from Cytiva (USA). eStain L1 Protein Staining Kit was obtained from GenScript (China). BCA Protein Assay Kit was obtained from Tiangen Biotech (China). TriPure Isolation Reagent was purchased from Merck (Germany). 5 × All-In-One RT MasterMix Kit was purchased from ABM (Canada). SYBR Green qPCR mix and 96-well fluorescent qPCR plates were purchased from Monard (China).

Expression and purification of HE4-Fc

HE4 (NCBI Accession No. NP_006094.3) was fused with an Fc fragment of Camelus bactrianus (NCBI Accession No. ALB75438.1) at the C-terminus for immunization (Supplementary Fig. S1). A whole HE4-Fc gene was synthesized and cloned into the pcDNA™3.1(+) vector by EcoRI/HindIII restriction enzymes sites. HEK293 cells (5 × 10^7) were transfected with 100 µg of the pcDNA™3.1(+) -HE4-Fc expression vector using Lipofectamine 3000 following the manufacturer’s recommended procedure. After transfection, cells were cultured in Gibco® FreeStyle™ 293 Expression Medium at 37 °C, 100 rpm, and 5% CO₂ for 3 days. After cells were removed by centrifugation, the supernatant was filtered through 0.22-µm filters for purification of HE4-Fc.

Purification of HE4-Fc was performed using a HiTrap™ Protein G HP column (1 mL). The column was equilibrated with 10 mM PBS (pH 7.2) at 1 mL/min, and the supernatant was injected. After a wash step with 10 mM PBS (pH 7.2), the bound protein was eluted using 0.1 M glycine–HCl (pH 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7). The UV absorbance was monitored at 280 nm. The protein peak was neutralized using 2 M Tris and dialyzed into 2.7)
levels with and without SF-tags, each sample was loaded with the same volume (20 µL) from 20 mL of periplasmic protein extracts, recovered from 200 mL culture broths. The resolved protein bands were visualized using eStain L1 Protein Staining Kit.

The recombinant proteins were purified by immobilized metal affinity chromatography using a 1-mL gravity column prepacked with Ni Sepharose 6 Fast Flow. In short, the column was pre-equilibrated with the binding buffer (20 mM Tris–HCl, 0.3 M NaCl, pH 8.0). Then, periplasmic protein extracts were loaded onto the column at a constant flow rate of 10 drops per minute. After washing with the washing buffer (20 mM imidazole, 20 mM Tris–HCl, 0.3 M NaCl, pH 8.0), the bound proteins were eluted using the elution buffer (250 mM imidazole, 20 mM Tris–HCl, 0.3 M NaCl, pH 8.0). The purified proteins were evaluated by SDS-PAGE and dialyzed into 10 mM PBS (pH 7.2). Protein concentrations were determined using BCA Protein Assay Kit. In brief, 25 µL of the standard solution (bovine serum albumin, BSA) and samples were added into a 96-well microplate, and 200 µL working solution was added to each well. After incubation at 37 ºC for 30 min, the absorbance was measured at 562 nm on a microplate reader (Spark, Switzerland). A standard curve of BSA was prepared to determine the protein concentration of each sample.

**Affinity analysis by surface plasmon resonance binding assay**

The affinity values ($K_D$) of 4G8 variants, 1G8 variants, and 3A3 variants were estimated by surface plasmon resonance (SPR) technology using the BIACORE T100 system (GE, USA). HE4-Fc (100 µg/mL) was immobilized on the CM5 sensor chip surface through amine coupling. Afterwards, the chip was incubated in 1 M ethanolamine-HCl (pH 8.5) to completely block the remaining activated NHS groups. Running buffer HBS-N (10 mM HEPES, 150 mM NaCl, pH 7.4) was pumped at a constant flow rate for setting up the assay. Nanobody variants (5 µg/mL) were sequentially injected at a flow rate of 30 µL/min for 120 s and then dissociated for 1000 s. After each injection, the chip was regenerated with 10 mM glycine–HCl (pH 1.5) for 30 s and then 0.1 M NaOH for 60 s at a flow rate of 30 µL/min. SPR signals were measured in response unit (RU) and recorded as a sensogram. The kinetic parameters, including the association rate constant ($k_{on}$), the dissociation rate constant ($k_{off}$), and the equilibrium dissociation constant ($K_D$), were calculated based on a 1:1 fitting model.

**Circular dichroism**

A Chirascan circular dichroism spectrometer (Applied Photophysics, UK) was utilized to determine the secondary structure and the melting temperature ($T_m$) of nanobody variants. The proteins were diluted to 0.25 mg/mL by 50 µM PBS (pH 7.4). Far-UV (190–250 nm) circular dichroism (CD) measurements were carried out in a quartz cuvette with a 1-mm optical path length at 1-nm resolution. Raw ellipticity data, given in millidegrees (mdeg), was converted to molar ellipticity $[\theta]$ in deg cm$^2$/dmol by the following equation (Greenfield 2006):

$$[\theta] = (\text{mdeg} \times \text{MRW}) ÷ (\text{pathlength} \times \text{[nanobody]})$$

where the mean residue weight (MRW = molecular weight of the nanobody in Da/number of backbone amino acids), pathlength = cell path length in millimeters, and [nanobody] = concentration of nanobody variants in milligrams per milliliter. Thermal unfolding experiments was followed at 200 nm with CD measurements taken every 2 ºC from 20 to 85 ºC with a temperature increase of 1 ºC/min using a Peltier junction temperature controller. The melting temperature analysis was performed in triplicate, and thermal denaturation data were fitted using the Global 3 software (version 3.1.0.1, Applied Photophysics, UK).

**Real-time RT-PCR analysis**

After inducing with 1 mM IPTG at 30 ºC for 6 h, the E. coli cells were pelleted and lysed by 1 mL TriPure Isolation Reagent. These samples were mixed with 200 µL chloroform and incubated at room temperature for 5 min, and then centrifuged at 4 ºC 12,000 rpm for 15 min. RNA was extracted from the water phase through isopropanol precipitation. After washing by 1 mL 75% ethanol, the RNA pellets were allowed to air-dry at room temperature and dissolved in 15 µL nuclease-free water. The quality and quantity of the RNAs were examined using a NanoVue Plus spectrophotometer (GE Healthcare, UK) and confirmed by electrophoresis.

eDNAs were synthesized using 5 × All-In-One RT MasterMix Kit according to the manufacturer’s procedure. The transcription levels of 4G8 were determined by real-time RT-PCR using SYBR Green qPCR Mix on a LightCycler 480 II thermal cycler system (Roche, Germany) (forward primer: TGGAGGCACCTTCAGTAACTATAAC; reverse primer: GATCCCTTCAGGGATGCTGATAG). The internal standard gene, 16S rRNA, was used as an internal control to normalize the transcription levels of the samples (forward primer: CTC TTGCCATCGGTAGTGCCC; reverse primer: CCAGTG TGGCTGTTCATCTCTCT). The real-time RT-PCR was set up as follows: 95 ºC for 30 s and 40 cycles of 95 ºC for 10 s, 60 ºC for 30 s.
Results

Effect of the repetitive SFP hexapeptide motifs on nanobody production

We immunized one alpaca with HE4-Fc. A phage display library was generated, and we performed four consecutive rounds of phage display, followed by a phage ELISA–based binding screen. We isolated three nanobodies (termed 4G8, 1G8, and 3A3) that bind specifically to HE4 (Supplementary Fig. S2).

To investigate the influence of the SF-tag in nanobody expression, nanobody 4G8 against HE4 was selected as the model nanobody. Previous studies showed that N-terminal fusion tags might affect the affinities of nanobodies. Thus, the SF-tag was fused at the C-terminus of 4G8. To facilitate target purification, we also added a His6 tag at the C-terminus. To examine the effect of the numbers of SFP hexapeptide units on nanobody production, 1 to 5 SFP hexapeptide units were fused to the C-terminus of 4G8 to form 4G8-1C, 4G8-2C, 4G8-3C, 4G8-4C, and 4G8-5C, respectively (Fig. 1). 4G8 without SFP hexapeptide units was used as a reference. The presence of the SF-tags on the protein molecular weight was minor owing to the small size of one SFP hexapeptide unit (0.4 kDa) (Table 1). The 4G8-SF-tag fusion genes were inserted into pMES4 expression plasmids and successfully expressed in the periplasm of E. coli BL21 (DE3) with a molecular weight identical to their expected sizes (Fig. 2A). There was no significant difference in E. coli cell growth among these recombinants, indicating that the SF-tags were suitable for protein fusion labels (Fig. 2B). As shown in SDS-PAGE analysis (Fig. 2A), the 4G8-SF-tag fusions showed a seeming yield increase compared with that of untagged-4G8. To assess whether the SFP hexapeptide units could be used as an expression-enhancing tag, the 4G8-SF-tag fusions were purified using nickel immobilized metal affinity chromatography (Fig. 2C) and the yields of purified proteins were calculated. As shown in Fig. 2D, all the 4G8-SF-tag fusions showed remarkable increases in the protein yield compared with that of untagged-4G8, resulting in fold increases in the protein yield of 4G8-1C, 4G8-2C, 4G8-3C, 4G8-4C, and 4G8-5C by 2.5-fold, 4.0-fold, 5.8-fold, 5.2-fold, and 7.1-fold, respectively. Thus, the protein yields of 4G8-SF-tag fusions were increased greatly by the extension of SFP hexapeptide units, indicating that the length of SFP hexapeptide units had a significant effect on protein production.

Fig. 1 Schematic diagrams of the target proteins with or without the C-terminal SF-tags. SF-tags were fused with 4G8 via the flexible (GGGS)3 linkers (red broken lines)

Table 1 Properties of 4G8-SF-tag fusions

| Protein | pI   | Molecular weight (Da) | GRAVY | mRNA free energy (kcal/mol) | Protein yield (mg/g) | Tm (°C) | k_{on} (M^{-1} s^{-1}) | k_{off} (s^{-1}) | K_D (M) |
|---------|------|-----------------------|-------|---------------------------|---------------------|---------|----------------------|-----------------|---------|
| 4G8     | 8.67 | 13,951.7              | -0.678| -62.1                     | 0.043               | 61.8    | 5.938 × 10^4         | 1.413 × 10^{-3} | 2.379 × 10^{-8}|
| 4G8-1C  | 8.67 | 15,126.5              | -0.624| -80.4                     | 0.108               | 62.2    | 2.792 × 10^4         | 2.764 × 10^{-3} | 9.900 × 10^{-9}|
| 4G8-2C  | 8.67 | 15,527.6              | -0.588| -83.7                     | 0.172               | 63.4    | 4.437 × 10^5         | 4.333 × 10^{-3} | 9.767 × 10^{-9}|
| 4G8-3C  | 8.67 | 15,927.0              | -0.555| -84.7                     | 0.249               | 63.4    | 2.091 × 10^5         | 2.926 × 10^{-3} | 1.400 × 10^{-8}|
| 4G8-4C  | 8.67 | 16,327.3              | -0.525| -90.1                     | 0.225               | 63.7    | 3.799 × 10^5         | 3.563 × 10^{-3} | 9.381 × 10^{-9}|
| 4G8-5C  | 8.67 | 16,727.5              | -0.496| -89.8                     | 0.307               | 63.3    | 2.797 × 10^5         | 3.226 × 10^{-3} | 1.153 × 10^{-8}|

The mRNA free energy was analyzed by an mRNA structure analyzer (http://rna.urmc.rochester.edu/RNAstructureWeb). The GRAVY was analyzed by a GRAVY calculator (http://bioinformatics.org/sms2/protein_gravy.html). The association and dissociation rate constants (k_{on} and k_{off}) were measured by SPR and used to calculate the equilibrium dissociation constant (K_D).

pI isoelectric point, GRAVY grand average of hydropathy
Affinity and stability characterization of the 4G8-SF-tag fusions

The antibody–antigen binding characteristic is generally evaluated by affinity analysis. To investigate whether the SF-tags interfere with the affinity of 4G8, we assessed the binding of 4G8-SF-tag fusions to HE4 by SPR binding analysis. The antigen, HE4 with an Fc fusion (HE4-Fc), was expressed by transient expression in HEK293 cells and purified by protein G chromatography. The binding of 4G8-SF-tag fusions to HE4-Fc was analyzed with a series of 4G8-SF-tag fusion concentrations ranging from 0.05 to 2 µM. The binding affinities were 23.8 nM for 4G8, 9.9 nM for 4G8-1C, 9.8 nM for 4G8-2C, 14.0 nM for 4G8-3C, 9.4 nM for 4G8-4C, and 11.5 nM for 4G8-5C (Table 1). Thus, the SPR values showed no significant difference in binding affinities between 4G8 and 4G8-SF-tag fusions, indicating that SF-tags did not affect the binding affinity of 4G8.

CD spectroscopy is a tool to analyze the secondary structure of protein. We firstly examined various secondary structures (α-helix, β-sheet, and random coils) of 4G8-SF-tag fusions by far-UV CD (190 ~ 250 nm). The spectra displayed one positive peak centered at 200 nm and one negative peak centered at 220 nm, which are the characteristic peaks for the β-sheet structure (Fig. 3A). Although the overall shapes of the far-UV CD spectra from 4G8 and 4G8-SF-tag fusions were similar, the intensity of the peak at 200 nm was lower for 4G8-SF-tag fusions, and the peak intensity was lower with the extension of the SFP hexapeptide unit length, indicating that the structure of 4G8 might change upon the addition of SF-tags.
Thermal stability is the unique property of nanobodies. The melting temperature (\(T_m\)) of a protein represents the temperature at which it exists in a state of equilibrium between unfolded and folded states. Thus, temperature-induced unfolding experiments were conducted to determine the \(T_m\) of 4G8-SF-tag fusions. Figure 3B shows the ellipticity at 200 nm plotted versus temperature for 4G8 with and without SF-tags. The \(T_m\) values of 4G8, 4G8-1C, 4G8-2C, 4G8-3C, 4G8-4C, and 4G8-5C were 61.8 °C, 62.2 °C, 63.4 °C, 63.4 °C, 63.7 °C, and 63.3 °C, respectively (Table 1). In the presence of SF-tags, the \(T_m\) values showed a minor increase, indicating a slight increase in the thermal stability of the protein. The hydrophobic interaction was reported to play an important role in increasing the thermal stability of proteins (Stepanenko et al. 2008). The GRAVY value of a protein is a measure of its hydrophobicity or hydrophilicity. Positive GRAVY values indicate hydrophobicity, and negative values mean hydrophilicity. Positive GRAVY values indicate a stronger hydrophobicity of proteins which may contribute to the higher thermal stability of 4G8-SF-tag fusions (Table 1).

**SF-tags enhanced the transcription of 4G8 protein**

To investigate the effect of SF-tag on nanobody expression at the transcription level, we performed real-time RT-PCR analysis on the encoding gene of 4G8. As expected, we observed a substantial increase in the mRNA level of the 4G8-SF-tag fusions (Fig. 4). In addition, the increase in 4G8 transcription was correlated with the size of the SFP hexapeptide unit. Compared with untagged 4G8, the mRNA levels of 4G8-1C, 4G8-2C, 4G8-3C, 4G8-4C, and 4G8-5C increased by 2.3-fold, 2.6-fold, 3.9-fold, 3.4-fold, and 7.0-fold, respectively. The results indicated that 4G8 with longer SFP hexapeptide units showed a higher transcription level, which was consistent with the protein yield results.

The structural stability of mRNA plays an important role in both transcription and translation of proteins (Mao et al. 2014; Shpaer 1985; Zur and Tuller 2012). The mRNA free energy for 4G8 with or without SF-tags was calculated by an mRNA structure analyzer (http://rna.urmc.rochester.edu/RNAstructureWeb). A lower free energy indicates a higher mRNA stability. As shown in Table 1, SF-tag fusions showed a significant decrease in the mRNA free energy, lowering the free energy of −18.3 kcal/mol for 4G8-1C, −21.6 kcal/mol for 4G8-2C, −22.6 kcal/mol for 4G8-3C, −28.0 kcal/mol for 4G8-4C, and −27.7 kcal/mol for 4G8-5C. Hence, there was a negative correlation between the mRNA abundance and mRNA free energy. Further, combined with protein expression results, it seemed that the more stable the mRNA of 4G8-SF-tag fusions become, its expression value increases.

**Fusion of the SF-tags is effective for other nanobodies**

We isolated a series of nanobodies against HE4. To further explore the applicability of SF-tags on other nanobody expression, nanobodies 1G8 and 3A3 were fused with the SF-tags and expressed in E. coli cells. Without the SF-tag fusion, 1G8 showed a yield of 0.049 mg/g of wet E. coli cells, while the fusion with the SF-tags increased the yield level of 1G8 variants by 2.5-fold for 1G8-1C, 2.2-fold for 1G8-2C, 2.9-fold for 1G8-3C, 2.6-fold for 1G8-4C, and 5.2-fold for 1G8-5C (Fig. 5A, B). Nanobody 3A3 was expressed poorly in the periplasm of E. coli cells as indicated by the SDS-PAGE gel in Fig. 5C that no 3A3 protein band could be visualized. However, the expression level of 3A3 was greatly enhanced upon the fusion with SF-tags, resulting in the yield level increasing by 2.1-fold for 3A3-1C, 1.6-fold for 3A3-2C, 2.3-fold for 3A3-3C, 4.9-fold for 3A3-4C, and 5.7-fold for 3A3-5C (Fig. 5D). These results indicated that the SF-tags could enhance the expression of other nanobodies.

Further, we examined the binding affinities of 1G8 variants and 3A3 variants with antigen HE4-Fc by SPR binding analysis. As shown in Table 2, the dissociation constant (\(K_d\)) values were 0.7 nM for 1G8, 0.9 nM for 1G8-1C, 1.3 nM for 1G8-2C, 1.6 nM for 1G8-3C, 2.1 nM for 1G8-4C, 2.2 nM for 1G8-5C, 64.9 nM for 3A3, 55.8 nM for 3A3-1C, 28.7 nM for 3A3-2C, 54.0 nM for 3A3-3C, 36.5 nM for 3A3-4C, and 42.6 nM for 3A3-5C. As expected, there was no significant difference in binding affinities between the untagged nanobodies and nanobodies with SF-tag fusions, suggesting that the SF-tags did not affect the affinities of nanobodies.
Discussion

Over the decades, nanobodies have attracted an increasing amount of attention from the pharmaceutical and biotechnology industries owing to their peculiar properties. One of the main limitations in the successful commercialization of any recombinant protein is the production yield and purification level. However, a large number of nanobodies are currently mainly produced in *E. coli* cells with a low yield. Although several fusion tags have been used to functionalize nanobodies (Gotzke et al. 2019; Veggiani et al. 2020), few of them demonstrated an expressing enhancement effect on nanobodies. Previously, we found an interesting result that the expression of nanobodies was enhanced by the fusion with an SFP hexapeptide (GAGAGS). The SFP is a widely abundant natural protein derived from the silkworm cocoons, which has been extensively used as biomaterials in biological and biomedical fields (Kundu et al. 2013). To our knowledge, there is no study on the effect of SFP on the expression level of recombinant proteins. We systematically assessed the effect of SFP hexapeptide on the expression of nanobodies in the present study. Three nanobodies against HE4 (4G8, 1G8, 3A3) were selected as the model nanobodies and designed with different numbers of SFP hexapeptide units (SF-tags) at the C-terminus. To avoid interfering with the protein structure and alleviate metabolic burden on *E. coli*, short peptide tags (up to 5 SFP hexapeptide units, 30 residues) were fused to the nanobodies. All the nanobodies with SF-tag fusions showed a significant increase in protein expression level compared to the untagged nanobodies. Meanwhile, there was a trend in enhanced expression level with an increasing number of the SFP hexapeptide fusion tags. Therefore, 10-, 30-, 40-, 50-SFP hexapeptide units were fused to another model nanobody 11C12, which was isolated from immune libraries to
target carcinoembryonic antigen (CEA) (Lin et al. 2020). As expected, the yield of the 11C12 variants was greatly enhanced by the extended SF-tags as compared to the untagged 11C12 (data not shown). Hence, these results indicated that the SF-tags can be used as a novel fusion tag to enhance nanobody expression in E. coli cells.

Several mechanisms have been proposed to explain the expression-enhancing effects of fusion tags. For instance, fusion tags can enhance the protein solubility by attracting or acting as the molecular chaperones (e.g., MBP, NusA, and SUMO tags), slowing down the protein translation (e.g., NusA tag), or regulating the formation of disulfide bonds (e.g., Trx and PDI tags) (Ki and Pack 2020). However, nanobodies themselves show an excellent solubility owing to a large number of hydrophilic residues in frame region 2 (FR2). Therefore, we speculated that the enhanced expression of nanobodies with SF-tag fusions may not be due to promoting protein solubility.

Furthermore, transcription is another key factor in protein expression. To investigate whether the increased yield of SF-tag fusions is due to an increased transcription, we performed real-time RT-PCR analysis on the encoding gene of 4G8. As expected, we observed a substantial increase in the mRNA level of 4G8 in the fusion constructs. In addition, the increase in 4G8 transcription was positively correlated with the size of tags, which was consistent with the expression yield results. The mRNA secondary structure and stability have been identified as the major factors that influence several cellular processes, including transcript splicing (Patterson et al. 2002), ribosome abundance and its binding (Mao et al. 2014), and regulation of gene expression level and its accuracy (Grunberg-Manago 1999; Mao et al. 2014; Shpaer 1985). Based on our prediction of the mRNA free energy using a structure analyzer, the mRNA sequence of untagged 4G8 had a higher free energy than those of the 4G8-SF-tag fusion sequences, indicating that the mRNA structure of SF-tagged proteins is more stable than that of the untagged one. The results indicated that the stability of the mRNA secondary structure holds a positive correlation with mRNA abundance, and high abundance of mRNA favors a greater protein abundance. These findings are homologous to the findings of Victor et al. that the mRNA stability positively regulated the mRNA expression level in Saccharomyces cerevisiae (Victor et al. 2019). Greater mRNA stability tends to have a higher protein abundance, presumably by preventing the aggregation of mRNA molecules (Zur and Tuller 2012), leading to a high translation efficiency (Mao et al. 2014) or increasing the accuracy of translation (Shpaer 1985). However, some studies found that the decrease in mRNA structure stability contributed to an increase in mRNA expression, as ribosomes must unwind every structure encountered during translation (Jia and Li 2005; Kudla et al. 2009; Nguyen et al. 2019). The counterintuitive findings suggest that the stability of the mRNA structure affects the translation in a complex manner.

The SF-tags used in this study are mostly composed of hydrophobic amino acids such as glycine and alanine. Most hydrophobic residues are totally or partially buried in the protein matrix, and could create a strong hydrophobic interaction network in the protein (Stepanenko et al. 2008). Serine provides an extra hydroxyl group that is able to engage in the hydrogen bonds between the serine side chains (Mayen et al. 2015). Weak non-covalent forces such as hydrophobic interactions and hydrogen bonds have been identified to play an important role in increasing the structural and thermal stability of proteins (Kumar et al. 2000). Furthermore, the silk-like (GAGAGS) blocks have been shown to spontaneously form hydrogen-bonded β-sheet crystals, which impart thermal stability (Anderson 1998; Anderson et al. 1994). Hence, these may explain the thermostability-enhancing effect of SF-tags.

The advantage of SF-tags for nanobody expression is that the amino acid sequence is 30 residues or less, which does not impair the activity when fused to the protein of interest. Therefore, it may be not necessary to remove the fusion tags for the protein application compared with the large-sized tags. In addition, SFPs have been known to possess various biological attributes. As a biomaterial approved by FDA, SFPs show an excellent wound healing and antibacterial activity, which can be used for cell growth matrix and tissue engineering (Joseph and Raj 2012). Furthermore, the SF-derived peptides were reported to have anticancer effects. In the study by Wang et al., silk fibroin peptides with a molecular weight of ~10 kDa inhibited the growth of human lung cancer cells by inducing the cell cycle arrest at the S phase and promoting cell apoptosis (Wang et al. 2019). In a later study, the melanin synthesis in B16 melanoma cells was inhibited with the silk fibroin peptide treatment via downregulation of microphthalmia-associated transcription factor (MITF) and tyrosinase expression (Wang et al. 2021). Therefore, apart from the expression-enhancing effect found in the present study, the fusion of the SF-tags might endow nanobodies with new functions for wider applications.

In conclusion, for the first time, the repetitive amino acid sequence motifs (GAGAGS) found in silk fibroin protein were developed as a novel and versatile fusion tag (SF-tag) to enhance the expression of nanobodies in E. coli. SF-tags of 1 to 5 hexapeptide units were fused to the C-terminus of three nanobodies against human epididymis protein 4 (4G8, 1G8, and 3A3). The protein yields were increased by the extension of hexapeptide units, resulting in a 2.5–7.1-fold increase for 4G8 variants, a 2.2–5.2-fold increase for 1G8 variants, and a 1.6–5.7-fold increase for 3A3 variants compared with those of untagged forms. Moreover, the fusion of SF-tags not only had no significant effect on the affinities of the nanobodies, but also showed a slight increase in thermal
stability. The fusion of SF-tags increased the transcription of 4G8 by 2.3–7.0-fold, indicating SF-tags enhanced the protein expression at the translational level.

**Supplementary Information** The online version contains supplementary material at https://doi.org/10.1007/s00253-022-11857-7.

**Author contribution** YJL and LWS conceived the study and designed the experiments; YJL and GY conducted the experiments and analyzed the data; GUY and LF contributed to result collection and processing; LWS interpreted the data and wrote the manuscript; YJL and GY reviewed and edited the writing; and NR, SHP, and FXY reviewed the manuscript. All authors have read and approved the manuscript.

**Funding** This work was supported by Shandong Energy Institute (SEI) (SEI I20212).

**Data availability** All data generated or analyzed during this study are included in this published article.

**Declarations**

**Ethics approval** The animal study underwent approval by the Department of Science & Technology of Shandong Province (SYXK(Lu)20160004).

**Conflict of interest** The authors declare no competing interests.

**References**

Anderson JP (1998) Morphology and crystal structure of a recombinant silk-like molecule, SLPI. Biopolymers 45(4):307–321. https://doi.org/10.1002/(SICI)1097-0282(19980405)45:4<307::AID-BIPS3E3.0.CO;2-P

Anderson JP, Cappello J, Martin DC (1994) Morphology and primary crystal-structure of a silk-like protein polymer synthesized by genetically-engineered *Escherichia coli* bacteria. Biopolymers 34(8):1049–1058. https://doi.org/10.1002/bip.360340808

Billen B, Vincke C, Hansen R, Devogdlt N, Muylldermans S, Adriensen P, Geduef W (2017) Cytoplasmic versus periplasmic expression of site-specifically and bioorthogonally functionalized nanobodies using expressed protein ligation. Protein Expr Purif 133:25–34. https://doi.org/10.1016/j.pep.2017.02.009

Custudio TF, Das H, Sheward DJ, Hanke L, Pazicky S, Pieprzyk J, Sorgenfrei M, Schroer MA, Gruzinov AY, Jeffries CM, Graewert MA, Svergun DI, Dobrev N, Remans K, Seeger MA, Mclnerney GM, Murrell B, Hallberg BM, Low C (2020) Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. Nat Commun 11(1):5588. https://doi.org/10.1038/s41467-020-19204-Y

Fan XY, Zhou YH, Yu JL, Li F, Chen Q, Song HP, Feng DX, Liu WS, Nian R, Xian M (2019) Non-affinity purification of a nanobody by void-exclusion anion exchange chromatography and multi-modal weak cation exchange chromatography. Sep Purif Technol 225:88–96. https://doi.org/10.1016/j.seppur.2019.05.068

Gotze H, Kilisch M, Martinez-Carranza M, Sograte-Idrissi S, Rajavel A, Schlichteaeer T, Engels N, Jungmann R, Stenmark P, Opazo F, Frey S (2019) The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. Nat Commun 10:4403. https://doi.org/10.1038/s41467-019-12301-7

Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1(6):2876–2890. https://doi.org/10.1038/nprot.2006.202

Grunberg-Manago M (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. Annu Rev Genet 33:193–227. https://doi.org/10.1146/annurev.genet.33.1.193

Hanke L, Perez LV, Sheward DJ, Das H, Schulte T, Moliner-Morro A, Corcoran M, Achour A, Hedeestan GBK, Hallberg BM, Murrell B, Mclnerney GM (2020) An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. Nat Commun 11(1):4420. https://doi.org/10.1038/s41467-020-18174-5

Huo JD, Le Bas A, Ruza RR, Duyvesteyn HME, Mikolajek H, Malinauskas T, Tan TK, Rijal P, Dumoux M, Ward PN, Ren JS, Zhou DM, Harrison PJ, Weckener M, Clare DK, Vogirala VK, Radecke J, Moynie L, Zhao YG, Gilbert-Jaramillo J, Knight ML, Tree JA, Buttigieg KR, Coombes N, Elmore MJ, Carroll MW, Carrique L, Shah PN, James W, Townsend AR, Stuart DI, Owens RJ, Naismith JH (2020) Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. Nat Struct Mol Biol 27:846–854. https://doi.org/10.1038/s41594-020-0469-6

Jia MW, Li YD (2005) The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*. FEBS Lett 579(24):5333–5337. https://doi.org/10.1016/j.febslet.2005.08.059

Joseph B, Raj SJ (2012) Therapeutic applications and properties of silk proteins from Bombyx mori. Front Life Sci 6(3–4):55–60. https://doi.org/10.1080/21533769.2012.760491

Jovecvska I, Muyldermans S (2020) The therapeutic potential of nanobodies. BioDrugs 34(1):11–26. https://doi.org/10.1007/s41594-019-00392-z

Ki MR, Pack SP (2020) Fusion tags to enhance heterologous protein expression. Appl Microbiol Biotechnol 104(6):2411–2425. https://doi.org/10.1007/s00253-020-10402-8

Kudla G, Murray AW, Tollervy D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324(5924):255–258. https://doi.org/10.1126.science.1170160

Kumar S, Tsai CJ, Nussinov R (2000) Factors enhancing protein thermostability. Protein Eng 13(3):179–191. https://doi.org/10.1093/protein/13.3.179

Kundu B, Rajkhowa R, Kundu SC, Wang XG (2013) Silk fibroin bio-materials for tissue regenerations. Adv Drug Del Rev 65(4):457–470. https://doi.org/10.1016/j.addr.2012.09.043

Li T, Cai H, Yao H, Zhou B, Zhang N, van Vlissingen MF, Kuiken T, Han W, GeorgysKesshel CH, Gong Y, Zhao Y, Qin W, Tian XX, Peng C, Lai Y, Wang Y, Hutter CAJ, Kuo SM, Bao J, Liu C, Wang Y, Richard AS, Raoul H, Lan J, Seeger MA, Cong Y, Rockx B, Wong G, Bi Y, Lavillette D, Li D (2021) A synthetic nanobody targeting RBD protects hamsters from SARS-CoV-2 infection. Nat Commun 12(1):4635. https://doi.org/10.1038/s41467-021-24905-z

Lin JT, Yu JL, Wang H, Xu YR, Li F, Chen XH, Liang YL, Tang JS, Wu LL, Zhou ZW, Chen CL, Liu MJ, Chun X, Nian R, Song HP (2020) Development of a highly thermostable immunoassay based on a nanobody-alkaline phosphatase fusion protein for carcinoembryonic antigen detection. Anal Bioanal Chem 412(8):1723–1728. https://doi.org/10.1007/s00216-020-02456-4

Liu MM, Li L, Xian M, Nian R, Song HP, Feng D (2018) Recent advances in the selection and identification of antigen-specific nanobodies. Mol Immunol 96:37–47. https://doi.org/10.1016/j.molimm.2018.02.012
