Molecular Biotechnology (2019) 61:451–460
https://doi.org/10.1007/s12033-019-00176-4

ORIGINAL PAPER

Optimizing Periplasmic Expression in *Escherichia coli* for the Production of Recombinant Proteins Tagged with the Small Metal-Binding Protein SmbP

Bryan D. Santos[1] · Jose Ruben Morones-Ramirez[1,2] · Isaias Balderas-Renteria[1,2] · Nestor G. Casillas-Vega[3] · David W. Galbraith[4] · Xristo Zarate[1,2]

Published online: 17 April 2019

Abstract

We have previously shown that the small metal-binding protein (SmbP) extracted from the gram-negative bacterium *Nitrosomonas europaea* can be employed as a fusion protein for the expression and purification of recombinant proteins in *Escherichia coli.* With the goal of increasing the amounts of SmbP-tagged proteins produced in the *E. coli* periplasm, we replaced the native SmbP signal peptide with three different signal sequences: two were from the proteins CusF and PelB, for transport via the Sec pathway, and one was the signal peptide from TorA, for transport via the Tat pathway. Expression of SmbP-tagged Red Fluorescent Protein (RFP) using these three alternative signal peptides individually showed a considerable increase in protein levels in the periplasm of *E. coli* as compared to its level using the SmbP signal sequence. Therefore, for routine periplasmic expression and purification of recombinant proteins in *E. coli,* we highly recommend the use of the fusion proteins PelB-SmbP or CusF-SmbP, since these signal sequences increase periplasmic production considerably as compared to the wild-type. Our work, finally, demonstrates that periplasmic expression for SmbP-tagged proteins is not limited to the Sec pathway, in that the TorA-SmbP construct can export reasonable quantities of folded proteins to the periplasm. Although the Sec route has been the most widely used, sometimes, depending on the nature of the protein of interest, for example, if it contains cofactors, it is more appropriate to consider using the Tat route over the Sec. SmbP therefore can be recommended in terms of its particular versatility when combined with signal peptides for the two different routes.

Keywords SmbP · IMAC · Protein expression and purification · Periplasm · Signal sequence · Sec pathway · PelB · RFP · CusF · Tat pathway · GFP · TorA

✉ Xristo Zarate
xristo.zaratekl@uanl.edu.mx

Bryan D. Santos
bryan.sr.qfb@gmail.com

Jose Ruben Morones-Ramirez
morones.ruben@gmail.com

Isaias Balderas-Renteria
ibalderas@hotmail.com

Nestor G. Casillas-Vega
nestor.casillas.vega@hotmail.com

David W. Galbraith
galbrait@email.arizona.edu

[1] Universidad Autonoma de Nuevo Leon, Facultad de Ciencias Quimicas, 66455 San Nicolas de los Garza, NL, Mexico

[2] Universidad Autonoma de Nuevo Leon, Centro de Investigacion en Biotecnologia y Nanotecnologia, Facultad de Ciencias Quimicas, Parque de Investigacion e Innovacion Tecnologica, 66629 Apodaca, NL, Mexico

[3] Departamento de Patologia Clinica, Universidad Autonoma de Nuevo Leon, Hospital Universitario Dr. Jose Eleuterio Gonzalez, 64460 Monterrey, NL, Mexico

[4] School of Plant Sciences and BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA

Introduction

Escherichia coli is still considered the first option as a host microorganism for the production of recombinant proteins. It is particularly attractive due to its low costs of cultivation and high production yields, and the availability of simple processes for extraction and purification of the target proteins [1, 2]. However, some difficulties may be encountered during the expression and purification of recombinant proteins such as, for example, the formation of inclusion bodies and associated complications adversely affecting the purification steps. The strategy of employing fusion proteins has assisted in dealing with some of these issues, since this strategy provides access not only to improved solubilities, but also to straightforward product purification via affinity chromatography. On the other hand, final product yields may be adversely affected due to the large size of some of these fusion protein candidates, examples including maltose-binding protein (MBP), N-utilization substance protein A (NusA), protein disulfide isomerase (PDI), as well as others. Therefore, an ideal fusion protein should be one that retains the positive attributes mentioned above but that also has a low molecular weight. Recently, we showed the small metal-binding protein (SmbP) isolated from the periplasm of *Nitrosomonas europaea* can be used as a fusion protein partner for recombinant protein production in *E. coli* [3]. In nature, SmbP acts as a metal scavenger that helps the bacterium to deal with high concentrations of metal ions. The function of SmbP involves its ability to bind a number of metal ions, including Cu(II), Ni(II), Zn(II), and even Fe(III) [4]. Therefore, SmbP is the first metal-binding protein to be used as a fusion protein partner. As for many other periplasmic proteins, SmbP contains a signal sequence, sometimes described as a signal peptide, at its N-terminus, in this case a conserved sequence directing export through the Sec pathway. If the signal sequence is removed, the truncated construct can be used for cytoplasmic protein production. Cytoplasmic expression of SmbP-tagged proteins fused to SmbP lacking the signal sequence showed good results, similar to those obtained with MBP and glutathione S-transferase (GST), with respect to increased solubility, increased protein amounts (up to 32% of total cellular protein in some cases), and improved purity after purification via metal affinity chromatography using Ni(II) ions [3]. In contrast, following expression of the fusion construct containing the SmbP signal peptide, although it was possible to obtain Red Fluorescent Protein (RFP) from the bacterial periplasm after osmotic shock, and although the cells turned visibly red, the amounts produced were much lower than those observed for cytoplasmic synthesis.

Expression in the periplasm of *E. coli* provides certain advantages over expression in the cytoplasm. First, it permits the appropriate formation of disulfide bonds which, in some proteins, are necessary for correct folding [5, 6]. Second, the periplasm contains only minor amounts of host proteins and proteases, which greatly simplifies purification. Finally, and very importantly, proteins can be purified from the periplasmic fraction following an osmotic shock, and without the need to lyse the whole cell, again simplifying purification [7]. Secretion of proteins can be directed to the Sec translocase through two mechanisms: co-translational (SRP-dependent) and post-translational pathway (SecB-dependent). Most periplasmic proteins follow the SecB-dependent pathway, in which SecB binds to the fully translated protein and maintains it in an unfolded state. SecA searches the signal peptide of the unfolded protein, and guides it to the SecYEG complex, where it is exported to the periplasm. Finally, SPase I cleaves the signal peptide releasing the protein [8, 9]. An alternative route to the periplasm is represented by the twin-arginine translocation (Tat) pathway, a protein transport system that is present in many bacteria and also in plant chloroplasts [10]. In contrast to the Sec-pathway, the Tat system exports folded proteins to the periplasm. In *E. coli*, this system comprises three integral membrane proteins, TatA, TatB, and TatC, which form the TatABC complex that works as a receptor and channel for the export of Tat substrates. As for the Sec-pathway, operation of the Tat system requires the presence of a signal peptide to direct protein export. Although Tat signal peptides share the same general structure with Sec signal peptides (n-region, h-region, and c-region), they have different characteristics: a conserved twin-arginine motif (T/SRRXFLK, where X is any polar amino acid) at the union of the n- and h-regions, and a less hydrophobic h-region. These features prevent Tat signal peptides from interacting with the Sec-pathway [9, 11].

In this work, we have explored three alternative signal sequences with the aim of improving the production of SmbP-tagged recombinant proteins in the periplasm of *E. coli.* Two sequences were for the Sec pathway, being from the *E. coli* Cu-sensitive free protein (CusF), and from pectate lyase B (PelB) of *Erwinia carotovora*. A third signal sequence, but for the Tat pathway, was taken from Trimethylamine-N-oxide reductase (TorA), also from *E. coli.* RFP was used as a reporter protein to allow easy comparison of levels of protein production from the different constructs, including the wild-type SmbP signal sequence as a control, and to permit ultrastructural localization of the protein (to the cytoplasm or the periplasm) using fluorescence microscopy. Green Fluorescent Protein (GFP) was used as a reporter to evaluate the capacity of TorA-SmbP to export folded proteins to the periplasm of *E. coli.*

Materials and Methods

Bioinformatics

For Sec and Tat signal sequence alignments, we employed the multiple sequence alignment program Clustal Omega at https://www.ebi.ac.uk/Tools/msa/clustalo/. The hydrophobicity indices of the h-regions from signal peptides were scored using the grand average of hydropathy (GRAVY) value at http://www.gravy-calculator.de/.

DNA Constructs

The construct for cytoplasmic expression, without signal sequence, of SmbP-RFP (SmbPc-RFP) and wild-type SmbP-RFP, which includes its signal sequence, were as previously reported; the SmbP sequence being directly amplified from *Nitrosomonas europaea* chromosomal DNA and cloned into pET30a (EMD Millipore) [3]. For the new constructs described in this work, the wild-type SmbP signal peptide was exchanged with those of PelB, CusF, and TorA as shown in Table 1. DNA was synthesized by GenScript, and they were provided in the pUC57 vector (only the sequences coding for the signal peptide were optimized for *E. coli* expression, the sequence for SmbP remaining unchanged from that of *N. europaea*).

Sequences of the new versions of SmbP (CusF-SmbP and PelB-SmbP) were amplified by PCR with the following primers: 5′-AGTCAGTCA**CATATG**AAAAAAGCACTG CAAGTCG-3′ (NdeI, forward), 5′-AGTCAGTCA**GGT ACC**GTGCGATTTATGTTCGGATGC-3′ (KpnI, reverse) for CusF-SmbP and 5′-ATGCATGCA**CATATG**AAATAC CTGCTGCCG-3′ (NdeI, forward), 5′-AGTCAGTCA**GGT ACC**GTGCGATTTATGTTCGGATGC-3′ for PelB-SmbP. The 50 µL reaction comprised 10 ng of template (plasmid DNA), 60 pmol of each primer, 1.5 µL of a 10 mM dNTP mix, and 2 units of high-fidelity *Vent* DNA polymerase (New England Biolabs) in 1X ThermoPol reaction buffer. Thermocycler conditions were: 95 °C for 2 min, 30 cycles of 95 °C (1 min), 56 °C (1 min), and 72 °C (1.5 min), and a final extension at 72 °C for 10 min. PCR products were desalted using the QIAquick Gel Extraction kit, and digested with NdeI and KpnI. The TorA-SmbP insert was also synthesized by GenScript and directly extracted from pUC57 by digestion with NdeI and KpnI, the fragment obtained being purified from an agarose gel using the QIAquick Gel Extraction kit. A previous pET30a-CusF-RFP construct [12] was digested with NdeI and KpnI, the CusF insert removed after gel purification, after which the new sequences (CusF-SmbP, PelB-SmbP, and TorA-SmbP) were ligated into it. All constructs were confirmed by sequencing.

Protein Expression

DNA constructs were transformed into *E. coli* BL21(DE3). A single colony was used to prepare a starter culture, by inoculation of 2 mL of Luria–Bertani (LB) broth, containing 30 µg/mL kanamycin, which was incubated at 37 °C for 16 h, with 200 rpm rotary shaking. For small-scale expression experiments, an aliquot (40 mL) of LB-kan broth was inoculated with 40 µL of starter culture and incubated at 37 °C and 200 rpm until its OD_{600} reached 0.4–0.6. Isopropyl β-D-1-thiogalactopyranoside (IPTG) from a 1 M stock solution was then added to reach a final concentration of 0.1 mM to induce expression. Incubation was continued at 25 °C for 16 h at 200 rpm. For large-scale expression experiments, cells were grown in baffled flasks until their OD_{600} reached 0.4–0.6. Expression was induced by adding IPTG to a final concentration of 0.1 mM, cells being incubated at 25 °C for a further 16 h at 200 rpm.

Fluorescence Microscopy

To analyze the periplasmic expression of all RFP constructs by fluorescence microscopy, aliquots (2 mL) of the expressed cultures were centrifuged at 10,000 rpm for 10 min and the pellets resuspended in 2 mL of PBS buffer. Cells were examined using a Leica (DM3000) fluorescence

Table 1 Signal sequences for the different SmbP constructs

Name	Signal sequence	Nucleotide sequence for DNA construct
SmbP (wild-type)	MKTTLIKVIAASVTALFLSMQVYA	ATG AAA ACA ACC CTG ATA AAA GTG ATT GCA GCC TCT GTT ACC GCA CTG TTT TTG AGC ATG CAG GTG TAT GCG
CusF	MKKALQVAMFSLFTVIGFNAQA	ATG AAA AAA GCA CTG CAA GTC GCA ATG TTC AGT CTG TTT ACC GTT ATT GGC TTT AAT GCC CAG GCT
PelB	MKYLLPTAAAGLLLLAAQPAMA	ATG AAA TAC CTG CTG CCG ACC GCT GCT GCT GGT CTG CTG CTG CTG GCT GCT CAG CCG GCT ATG GCT
TorA	MNNNDLFQASRRRFLAQLGGLT-VAGMLGPSLLTPRRATA	ATG AAC AAC AAC GAC CTG TTC CAG GCG AGC CGT CGT CGT TTC CTG GCG CAA CTG GGT GGC CTG ACC GTG GCG GGT ATG CTG GGT CCG AGC CTG CTG ACC CCG CGT CGT GCG ACC GCG

microscope at an excitation wavelength of 545 nm and an emission wavelength of 585 nm.

Lysozyme/Osmotic Shock

Periplasmic fractions for all constructs were obtained using a lysozyme/osmotic shock procedure. Cells were harvested by centrifugation at 8000 × *g* for 10 min and resuspended in a hypertonic solution (200 mM Tris–HCl, 20% sucrose, 2 mM EDTA and 0.5 mg/mL lysozyme and pH 8.0) at a ratio of 10 mL/g wet cell weight. After incubation for 40 min at 4 °C, the cells were centrifuged at 8000 x g for 10 min to collect the hypertonic supernatant. The cells were resuspended in hypotonic solution (cold distilled water) at a ratio of 10 mL/g of wet cell weight. After incubation for 40 min at 4 °C, the cells were centrifuged at 8000 x g for 10 min to collect the hypotonic supernatant containing the periplasmic proteins. All fractions were analyzed by SDS-PAGE, and the purity of the target protein was determined using ImageJ software (version 1.52) from the U.S. National Institutes of Health [13].

Protein Purification

Purification for all protein products was done using the ÄKTA Prime plus fast protein liquid chromatography system (GE Healthcare). The hypotonic supernatant was loaded into a HiTrap IMAC FF 5 mL-column charged with Ni(II) previously equilibrated with binding buffer (50 mM Tris–HCl, 500 mM NaCl, pH 8.0). After loading, the column was washed with washing buffer (50 mM Tris–HCl, 500 mM NaCl, 10 mM imidazole, pH 8.0). The SmbP tagged proteins were eluted using an imidazole gradient (10–200 mM).

Quantification of RFP and GFP by Fluorescence Spectroscopy and Acquisition of the GFP Fluorescence Emission Spectrum

For quantification of RFP and GFP, calibration curves were constructed using fluorescence intensities at different concentrations ranging from 1 µg/mL to 70 µg/mL with triplicate measurements. The standard and experimental samples were each diluted using 50 mM Tris–HCl, 500 mM NaCl, pH 8.0 buffer and the fluorescence emission was obtained with a Luminescence Spectrometer LS 55 (Perkin Elmer) in triplicates as well. For GFP and RFP, the respective excitation wavelengths were set to 485 and 545 nm, and the emission wavelengths to 510 and 585 nm. The fluorescence emission spectrum for SmbP-GFP (from the TorA construct) was obtained using the same equipment, the excitation wavelength being set to 488 nm.

Tag Removal

1 mg of SmbP-RFP (in 220 µL of 50 mM Tris–HCl, 500 mM NaCl, pH 8.0) from the PelB construct was incubated with 25 units of enterokinase light chain (New England Biolabs) for 16 h at 25 °C. Profinity IMAC resin (BIO-RAD) charged with Ni(II) was then added to the reaction mixture. After incubation for 2 h at 4 °C, the resin was centrifuged at 1500 × *g* for 1 min, and the supernatant was analyzed by SDS-PAGE to confirm removal of SmbP.

Results

Figure 1a shows the alignment of several signal sequences for protein translocation in bacteria using the Clustal Omega software [14]. CusF, PelB, OmpA, and DsbA signal peptides have a shorter n-region than those of SmbP, LamB, MalE, and PhoA. They all have a positive charge due to the presence of lysine and arginine. However, the degree of net charge is variable, being +3 for MalE, +2 for CusF, SmbP, OmpA, LamB, and DsbA, and +1 for PelB and PhoA. The h-region of the signal peptides has differences in the total number of amino acids. The PelB h-region is the largest being 14 amino acids in length, whereas SmbP and OmpA have 13 amino acids, CusF, LamB, and MalE have 12 amino acids, and PhoA and DsbA only 10 amino acids. The calculated hydrophobicity indices of the signal peptide h-regions also show differences. PelB h-region has a high content of hydrophobic amino acids, with a hydrophobicity index of 2.07, but the highest value (2.62) is found for the DsbA h-region. As a signal peptide, DsbA targets entry into the SRP-dependent pathway, and previous studies have implicated the importance of this high hydrophobicity in correct targeting to this pathway [15]. Some signal peptides have helix breakers that destabilize the h-region. For example, the signal peptides from PelB, OmpA, LamB, and DsbA all have a glycine residue within the h-region, whereas SmbP, CusF, and MalE signal peptides do not. Nevertheless, the position of the glycine residue is not a fixed feature. In PelB, it is at the core of the h-region, whereas in OmpA, LamB, and DsbA the glycine is located towards the c-region.

The sequence alignments of the signal sequences of the Tat pathway (TorA, SufI, and HyaA) are illustrated in Fig. 1b. The n-region possess a T/SRRXFLK motif which is characteristic of Tat signal peptides. The h-region is less hydrophobic than the corresponding region of the Sec signal, providing a hydrophobicity score of 1.55 for TorA, 1.87 for SufI, and 1.14 for HyaA. It has been shown that increasing the hydrophobicity of the h-region of the Tat signal peptides redirects their targeting specificity to the Sec-pathway. Additional differences include that Tat-signal peptides contain more glycine residues and fewer leucine residues than Sec

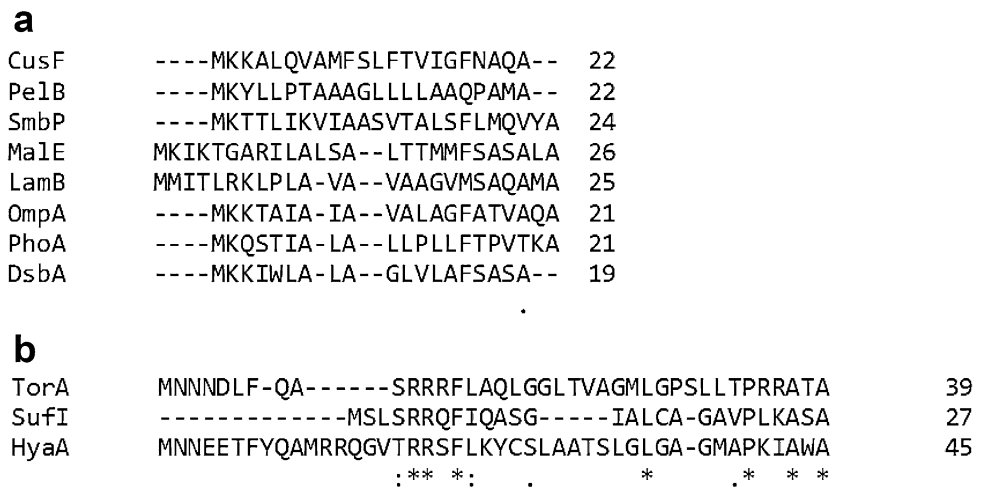

Fig. 1 Multiple sequence alignment of signal peptides with Clustal Omega. **a** Sec signal peptides: CusF - Cation efflux system protein, SmbP - Small metal binding protein, PelB - Pectate lyase B, OmpA - Outer membrane protein A, LamB - Maltose outer membrane porin, MalE - Maltodextrin-binding protein, PhoA - Alkaline phosphatase, DsbA - Disulfide interchange protein. **b** Tat signal peptides: TorA - Trimethylamine-N-oxide reductase 1, HyaA - Hydrogenase-1 small chain, SufI - Cell division protein. An asterisk (*) indicates positions which have a single conserved residue. The colon (:) indicates residues of strongly similar properties. The period (.) indicates conservation of weakly similar properties [14]

signal peptides, and the n-regions possess a higher content of arginine and negatively charged residues. In the c-region, proline is highly conserved in position -6; in -3 and -1 we find the same AXA motif as in the Sec signal peptides [16].

We examined the in vivo cellular locations of SmbP-RFP produced from all constructs using fluorescence microscopy. To examine cytoplasmic expression, RFP tagged with SmbP lacking its signal sequence (SmbPc-RFP) was employed. When observed under the microscope, we noticed an even distribution of RFP fluorescence across the whole cell (Fig. 2a). In contrast, expression from the PelB-SmbP-RFP and CusF-SmbP-RFP constructs (Fig. 2b and C) resulted in a different distribution of fluorescence, being mostly located at the poles. This was similar to the distribution seen using wild-type SmbP-RFP (Fig. 2d), although the latter showed a considerably lower intensity of fluorescence. For the TorA-SmbP-RFP construct, fluorescence was also distributed at the bacterial poles (Fig. 2e).

In the next set of experiments, we employed SDS-PAGE to examine the periplasmic protein content of *E. coli* BL21(DE3) following osmotic shock. The results for RFP tagged with wild-type SmbP, CusF-SmbP, PelB-SmbP, are shown in Fig. 3a, with their protein contents from the hypertonic and hypotonic fractions, respectively. The same analysis was performed for TorA-SmbP-RFP (Fig. 3b). Compared with the control of uninduced *E. coli,* new bands for RFP tagged with CusF-SmbP, PelB-SmbP, and TorA-SmbP appeared at the expected molecular weights in both the hypertonic and hypotonic fractions, whereas for RFP tagged with wild-type SmbP, a protein band was not apparent.

To quantify the amount of RFP produced with the different periplasmic versions of SmbP, a calibration curve was constructed; it shows linearity, with a correlation coefficient of 0.9989 (Fig. 4a). Next, large-scale preparations were made of each RFP fusion protein, the protein being purified from the osmotic shock extracts using IMAC functionalized with Ni(II) ions. After purification, the fluorescence intensities of each sample were determined (in triplicates) for quantification. For RFP tagged with PelB-SmbP, we obtained 7.9 ± 0.16 mg protein per liter, with CusF-SmbP 6.7 ± 0.09 mg/L, with TorA-SmbP 4.1 ± 0.22 mg/L, and with the wild-type 0.0067 ± 0.0003 mg/L.

To evaluate the ability of the TorA-SmbP construct to export folded protein into the periplasm, we linked this to the GFP coding sequence for expression in *E. coli*. After osmotic shock, the hypotonic fraction showed a light green color. IMAC Ni(II) chromatography was then used to purify the protein, and a fluorescence spectrum was determined from the eluted fractions (Fig. 4b). The emission maximum at 510 nm is as expected for GFP. A calibration curve was also constructed from pure GFP (insert to Fig. 4b) with a correlation coefficient of 0.9994. Thus, we calculated a production of 0.25 ± 0.003 mg of GFP tagged with TorA-SmbP per liter of cell culture.

In a final set of experiments, we employed SDS-PAGE to analyze the cleavage with enterokinase of SmbP-RFP purified from the periplasmic fraction and SmbP removal after second IMAC Ni(II) chromatography (Fig. 5). Table 2 shows the full purification summary table [17], which includes percentage yields.

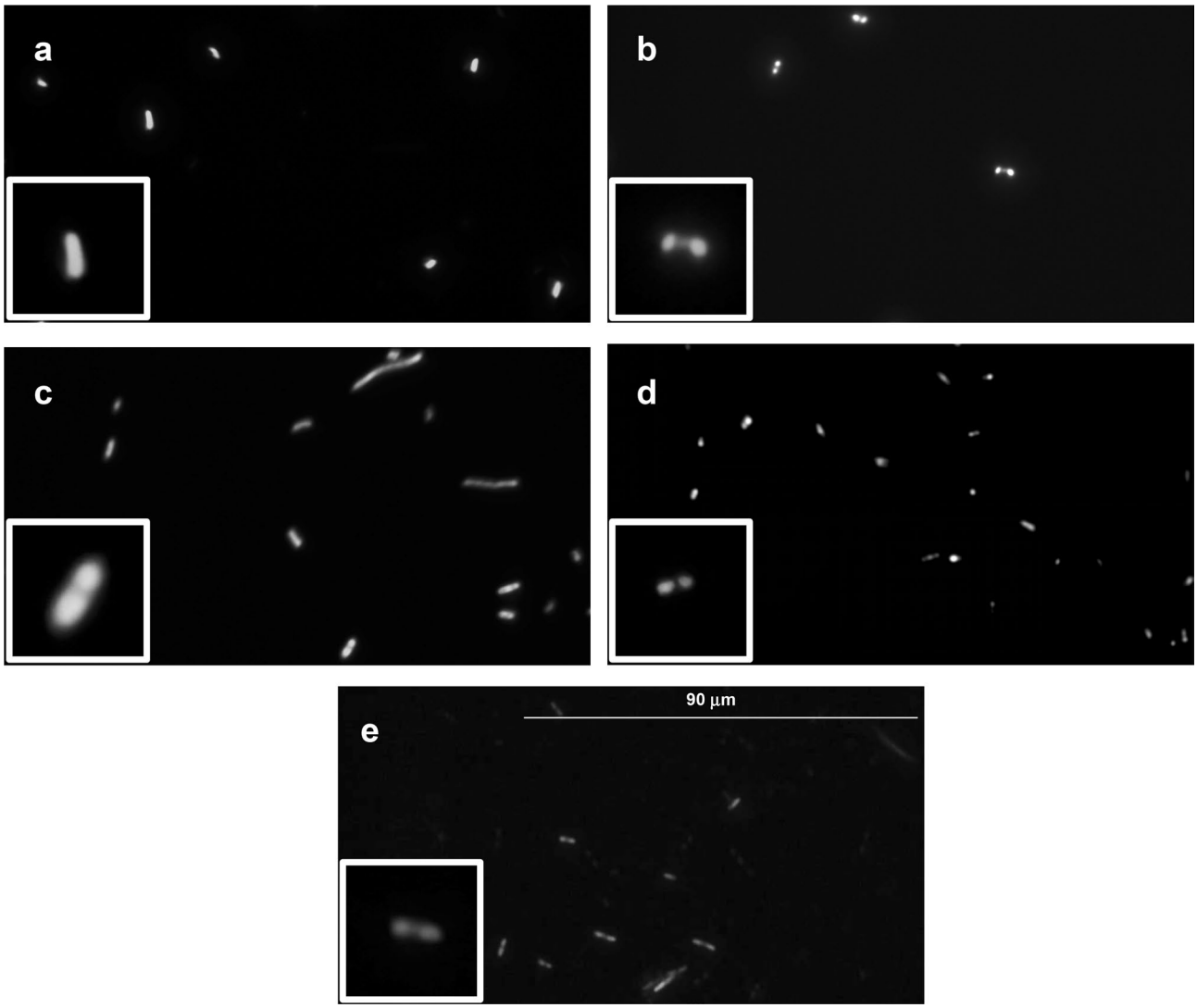

Fig. 2 Localization of RFP in *E. coli* BL21(DE3) after expression using all SmbP constructs. **a** SmbPc-RFP; **b** CusF-SmbP-RFP; **c** PelB-SmbP-RFP; **d** wild-type SmbP-RFP; **e** TorA-SmbP-RFP. All images include an inset of an individual cell

Discussion

We previously have demonstrated the use of SmbP, a protein isolated from the periplasm of the gram-negative bacterium *Nitrosomonas europaea*, as a fusion tag for both cytoplasmic and periplasmic expression of recombinant proteins [3]. However, the expression levels for periplasmic proteins were in need of improvement. To achieve this purpose, in this work, different versions of SmbP were designed by replacing its native signal sequence. The signal peptide from CusF was identified as the first option, based on results previously obtained by our research group, in which we observed that periplasmic expression was higher for CusF-RFP than for SmbP-RFP [12, 18]. CusF is an *E. coli* periplasmic metal-binding protein, with high affinities for Ag(I) and Cu(I). It is part of the CusCFBA system, whose function is to secrete these toxic ions from the cell [19]. CusF also displays some affinity for Cu(II) [20], allowing IMAC purification functionalized with these ions for CusF-tagged proteins. This underpinned our proposal that CusF is appropriate as a fusion partner for protein production both in vivo and in vitro (using cell-free expression systems) [12]. The second option, provided by the PelB signal peptide, was identified because this peptide has been applied extensively in a variety of vectors for successful periplasmic expression of a considerable number of recombinant proteins in *E. coli* [21–26]. Finally the signal peptide from TorA was chosen to test the feasibility of exporting folded, active SmbP-tagged proteins to the periplasm via the Tat pathway, TorA being the most utilized and studied signal sequence for this purpose [27].

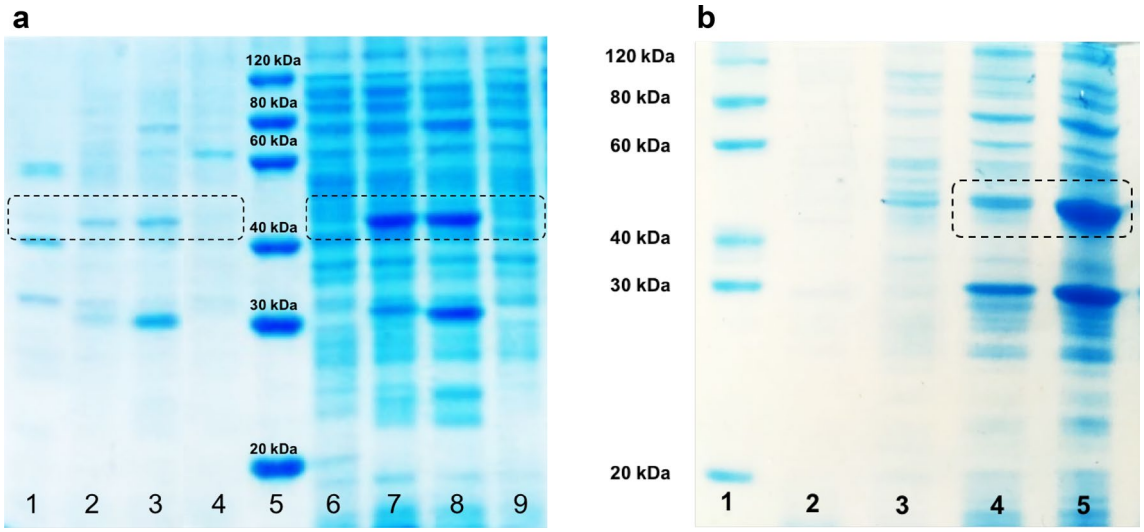

Fig. 3 12% SDS-PAGE analysis of RFP periplasmic expression. **a** Signal sequences for the Sec pathway (CusF, PelB, and wild-type). Lanes 1–4: hypertonic fractions; lane 5: protein marker; lanes 6–9: hypotonic fractions. Lanes 1 and 6: *E. coli* Bl21(DE3) uninduced controls; lanes 2 and 7: CusF-SmbP-RFP; lanes 3 and 8: PelB-SmbP-RFP; lanes 4 and 9: wild-type SmbP-RFP. Calculated molecular weight for SmbP-RFP: 37.2 kDa. **b** Periplasmic expression with TorA signal sequence. Lane 1: protein marker; lanes 2 and 4: hypertonic fraction; lanes 3 and 5: hypotonic fraction. Lanes 2 and 3: *E. coli* BL21(DE3) uninduced control; lanes 4 and 5: TorA-SmbP-RFP

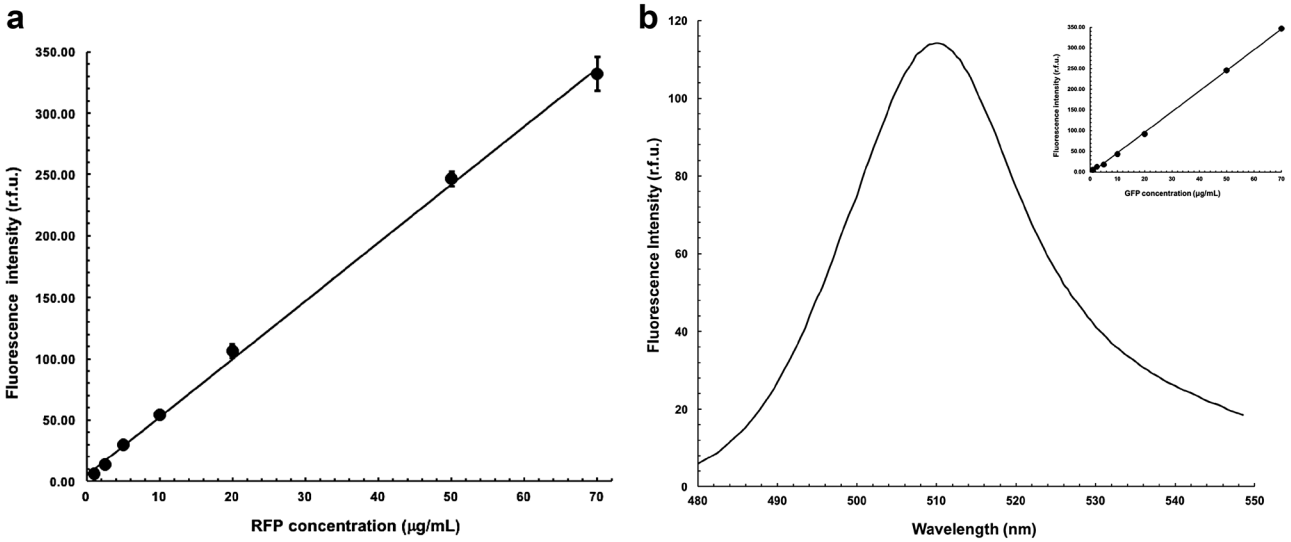

Fig. 4 **a** RFP calibration curve employed for quantification of protein produced for each SmbP construct in one liter of culture. The fluorescence intensities of specific amounts of purified RFP were determined for each standard to create the calibration curve. **b** Fluorescence spectrum of GFP purified from the periplasm of *E. coli* after expression with the TorA-SmbP-GFP construct. Excitation was at 488 nm. Inset: Calibration curve for quantification of GFP

After RFP expression using these new constructs and purification of the resultant proteins, our results indicated a decisive role for the signal peptides in terms of the amounts and locations of the fusion proteins that were produced. Fluorescence microscopy clearly differentiates between the distributions of RFP fluorescence from the cytoplasmic and periplasmic constructs. When located in the periplasm, the fluorescent proteins tended to accumulate at the poles of the cell. This characteristic distribution has been widely observed in studies of GFP secretion using the Tat pathway [10, 28], and with new versions of GFP that can be folded correctly in the periplasm after export through the Sec pathway [29]. Apparently, when GFP is expressed in the periplasm of *E. coli* and the cells are resuspended in

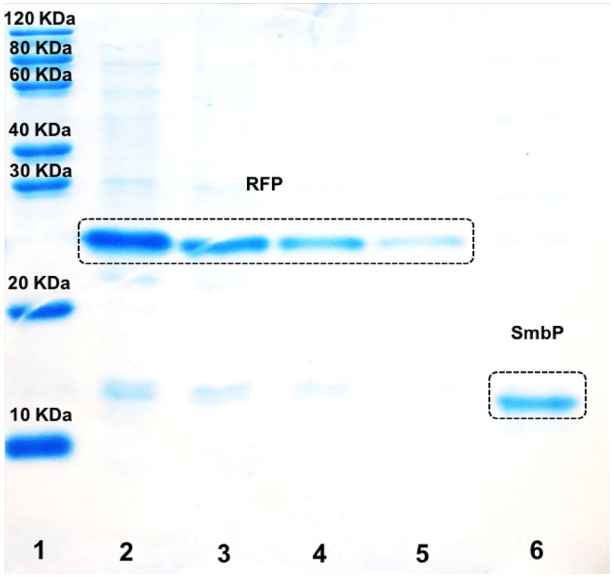

Fig. 5 15% SDS-PAGE analysis of SmbP tag removal from the PelB-SmbP-RFP construct after reaction with enterokinase. Lane 1: protein marker; lanes 2–5: RFP fractions after second purification with IMAC; Lane 6: SmbP after elution with 200 mM imidazole

PBS buffer, this results in an accumulation of the fluorescent protein at the poles.

As for the amount of RFP produced after expression with each different SmbP construct, we observed considerable differences between the wild-type and the alternative signal sequences. Most of the proteins were found in the periplasmic fraction obtained by osmotic shock, consistent with previous reports [25, 26]. After purification, we quantified the amounts of RFP based on fluorescence intensity. We generated calibration curves in a manner similar to that using GFP showing linearity with excellent correlation coefficients [30, 31]. With this information, a calibration curve using RFP as a standard provided a correlation coefficient of 0.9989 between fluorescence and RFP amount, demonstrating its linearity. Applying this to quantify the amounts of RFP produced for each construct, we found that the PelB-SmbP-RFP construct out-performed all others.

Taking a closer look at the three signal sequences (SmbP, CusF, and PelB) that we used for directing protein secretion through the *E. coli* Sec pathway, it is not immediately obvious why PelB works best and why SmbP is so poor. All signal sequences for transport via the Sec pathway have a general structure that contains three sections: n-, h-, and c-. An essential feature in section n- is the presence of a positive charge from a basic amino acid, usually a lysine. Each of the three signal peptides have this amino acid at the second position. The positive charge is vital for insertion of the signal peptide into the inner membrane due to an ionic interaction with negative charges present on the membrane lipids [32]. Also, a strong initiation of translation, using the AAA codon for lysine, has been shown to have a significant influence on the ability to export proteins to the periplasm [33]. In our study, since the three signal peptides used for Sec transport each possess a lysine at the P2 position and the AAA codon, we do not believe that the differences in periplasmic accumulation are influenced by the n- section of the signal peptides. In contrast, we observe more differences in the h- regions. It has been shown that high hydrophobicity is extremely important, and if this characteristic is altered by mutation or deletion, export of the protein is inhibited [34]. The PelB sequence contains a highly hydrophobic h-section as compared to SmbP, and also has a glycine that functions as a helix breaker. It has been reported that if this glycine is replaced by a hydrophobic amino acid, the protein becomes a target of SRP for co-translational transport [35]. Furthermore, the presence of this glycine in the middle of the h- section has shown highly positive effects on the levels of production of periplasmic proteins in *E. coli* [33]. Regarding section c-, the AXA motif is widely conserved, with alanines being preferred at positions -1 and -3. Section c- is essential for the efficient cleavage of the peptide and this step dramatically influences the secretion of the proteins [32]. SmbP is the only one of the three peptides that does not contain an alanine in position -3, which might explain its performance, as shown by its lower RFP production using the wild-type SmbP construct.

We previously have reported that wild-type SmbP does not produce active GFP in the periplasm of *E. coli,* even though a protein band was observed in the SDS-PAGE, whereas the cytoplasmic construct (SmbPc-GFP) produced a considerable amount of fluorescent GFP [3]. We decided to test a signal peptide for the Tat pathway, which is known to export folded proteins to the periplasm, taking advantage of the low molecular weight of SmbP and its capacity to fold correctly in the cytoplasm. Furthermore, proteins that contain cofactors or certain posttranslational modifications have to be folded first in the cytoplasm before being exported to the periplasm through the Tat pathway [9, 36]. Interestingly,

Table 2 Purification summary for RFP tagged with PelB-SmbP from the periplasmic fraction of 1 L of cell culture

Purification step	Total protein (mg)	Amount of RFP (mg)	Yield (%)	Purity (%)
Periplasmic crude extract	267.33	40.10	100	25
IMAC	27.40	14.80	36.90	54
Cleavage and IMAC	7.92	7.21	17.97	91

the TorA-SmbP-GFP construct produced active GFP located in the periplasm, as determined from the fluorescence spectrum obtained from the periplasmic fraction after expression and purification via osmotic shock. Apparently, SmbP did not affect the capacity of the TatBCA complex to export GFP, a fluorescence spectrum, with the expected maximum emission wavelength of 510 nm, being observed for the protein produced from this construct [37]. The amount of protein produced from the TorA-SmbP-RFP construct was lower as compared with that from CusF-SmbP and PelB-SmbP, although it was much higher than that from the wild-type. As for TorA-SmbP-GFP, a considerably lower amount of protein was obtained, compared to RFP, which means that this process is target protein dependent.

Finally, SmbP tag removal can be achieved with Enterokinase, a protease that recognizes the sequence DDDDK included in all the constructs. SDS-PAGE analyses show that tag removal is easily performed with a second IMAC step as previously reported [3], obtaining pure protein with high yields, since SmbP is only 9.9 kDa in size.

Conclusions

Our results demonstrate that any of the alternative signal sequences serve to increase the production of SmbP-tagged RFP in the periplasm of *E. coli.* The signal sequence of PelB provided the best results, which we believe is due to its highly hydrophobic h- region that also contains a glycine residue that acts as a helix breaker. Close to PelB, the signal peptide from CusF also produced favorable results. Although a native sequence of *E. coli*, perhaps the position of the glycine residue does not favor optimal transport as seen for PelB. We believe that the low production of proteins in the periplasm of *E. coli* using the native signal peptide of SmbP reflects the fact that this sequence is longer and far from being optimal. Another very positive result is the demonstration of the ability to use the TorA signal sequence for the transport of folded proteins into the bacterial periplasm. This feature could contribute to producing proteins that require cofactor binding or that may otherwise be misfolded following the formation of incorrect disulfide bonds. Therefore, in addition to the periplasmic expression with the alternative peptide sequences shown here, in general, SmbP is an attractive fusion protein for the production of recombinant proteins in the cytoplasm or periplasm of *E. coli*, since it increases solubility, allows simple, one-step purification through affinity chromatography, and provides superior final yields due to its low molecular weight.

Acknowledgements We thank Mexico's Consejo Nacional de Ciencia y Tecnologia (CONACYT) for the financial support provided to BDS during his graduate studies. This work was funded by Grants UANL-PAICYT-2015 CN567-15 and CONACYT Grant CB 2012-179,774-B awarded to XZ.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest.

References

1. Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *The Protein Journal, 32*(6), 419–425.
2. Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology, 72*(2), 211–222.
3. Vargas-Cortez, T., Morones-Ramirez, J. R., Balderas-Renteria, I., & Zarate, X. (2016). Expression and purification of recombinant proteins in *Escherichia coli* tagged with a small metal-binding protein from *Nitrosomonas europaea*. *Protein Expression and Purification, 118,* 49–54.
4. Barney, B. M., LoBrutto, R., & Francisco, W. A. (2004). Characterization of a small metal binding protein from *Nitrosomonas europaea*. *Biochemistry, 43*(35), 11206–11213.
5. Dow, B. A., Tatulian, S. A., & Davidson, V. L. (2015). Use of the amicyanin signal sequence for efficient periplasmic expression in *E. coli* of a human antibody light chain variable domain. *Protein Expression and Purification, 108,* 9–12.
6. de Marco, A. (2015). Recombinant antibody production evolves into multiple options aimed at yielding reagents suitable for application-specific needs. *Microbial Cell Factories, 14*(1), 125.
7. Ramanan, R. N., Tan, J. S., Mohamed, M. S., Ling, T. C., Tey, B. T., & Ariff, A. B. (2010). Optimization of osmotic shock process variables for enhancement of the release of periplasmic interferon-a2b from *Escherichia coli* using response surface method. *Process Biochemistry, 45*(2), 196–202.
8. Mori, H., & Ito, K. (2001). The Sec protein-translocation pathway. *Trends in Microbiology, 9*(10), 494–500.
9. Natale, P., Brüser, T., & Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta - Biomembranes, 1778*(9), 1735–1756.
10. Thomas, J. D., Daniel, R. A., Errington, J., & Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*. *Molecular Microbiology, 39*(1), 47–53.
11. Patel, R., Smith, S. M., & Robinson, C. (2014). Protein transport by the bacterial Tat pathway. *Biochimica et Biophysica Acta - Molecular Cell Research, 1843*(8), 1620–1628.
12. Cantu-Bustos, J. E., Vargas-Cortez, T., Morones-Ramirez, J. R., Balderas-Renteria, I., Galbraith, D. W., McEvoy, M. M., et al. (2016). Expression and purification of recombinant proteins in *Escherichia coli* tagged with the metal-binding protein CusF. *Protein Expression and Purification, 121,* 61–65.
13. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods, 9*(7), 671–675.
14. Remmert, M., Lopez, R., Li, W., Dineen, D., Sievers, F., Gibson, T. J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*(1), 539.

15. Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D., & Beckwith, J. (2003). The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. *Journal of Bacteriology, 185*(19), 5706–5713.
16. Cristóbal, S., De Gier, J. W., Nielsen, H., & Von Heijne, G. (1999). Competition between Sec- and Tat-dependent protein translocation in *Escherichia coli*. *EMBO Journal, 18*(11), 2982–2990.
17. Burgess, R. R. (2009). Preparing a Purification Summary Table. In *Methods in Enzymology*, *Volume 463 Guide to Protein Purification, Second Edition* (pp. 29-34). Academic Press.
18. Cantu-Bustos, J. E., Cano del Villar, K. D., Vargas-Cortez, T., Morones-Ramirez, J. R., Balderas-Renteria, I., & Zarate, X. (2016). Recombinant protein production data after expression in the bacterium *Escherichia coli*. *Data in Brief, 7,* 502–508.
19. Loftin, I. R., Franke, S., Roberts, S. A., Weichsel, A., Héroux, A., Montfort, W. R., et al. (2005). A novel copper-binding fold for the periplasmic copper resistance protein CusF. *Biochemistry, 44*(31), 10533–10540.
20. Astashkin, A. V., Raitsimring, A. M., Walker, F. A., Rensing, C., & McEvoy, M. M. (2005). Characterization of the copper(II) binding site in the pink copper binding protein CusF by electron paramagnetic resonance spectroscopy. *Journal of biological inorganic chemistry: JBIC: a publication of the Society of Biological Inorganic Chemistry, 10*(3), 221–230.
21. Sockolosky, J. T., & Szoka, F. C. (2013). Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. *Protein Expression and Purification, 87*(2), 129–135.
22. Dong, Z., Zhang, J., Du, G., Chen, J., Li, H., & Lee, B. (2015). Periplasmic export of bile salt hydrolase in *Escherichia coli* by the Twin-Arginine signal peptides. *Applied Biochemistry and Biotechnology, 177*(2), 458–471.
23. Zhou, Y., Zhou, Y., Li, J., Chen, J., Yao, Y., Yu, L., et al. (2015). Efficient expression, purification and characterization of native human cystatin C in *Escherichia coli* periplasm. *Protein Expression and Purification, 111,* 18–22.
24. Samant, S., Gupta, G., Karthikeyan, S., Haq, S. F., Nair, A., Sambasivam, G., & Sukumaran, S. (2014). Effect of codon-optimized *E. coli* signal peptides on recombinant *Bacillus stearothermophilus* maltogenic amylase periplasmic localization, yield and activity. *Journal of Industrial Microbiology and Biotechnology*, *41*(9), 1435–1442.
25. Takemori, D., Yoshino, K., Eba, C., Nakano, H., & Iwasaki, Y. (2012). Extracellular production of phospholipase A 2 from *Streptomyces violaceoruber* by recombinant *Escherichia coli*. *Protein Expression and Purification, 81*(2), 145–150.
26. Mohajeri, A., Pilehvar-Soltanahmadi, Y., Pourhassan-Moghaddam, M., Abdolalizadeh, J., Karimi, P., & Zarghami, N. (2016). Cloning and expression of recombinant human endostatin in periplasm of *Escherichia coli* expression system. *Advanced Pharmaceutical Bulletin, 6*(2), 187–194.
27. Freudl, R. (2017). Beyond amino acids: Use of the *Corynebacterium glutamicum* cell factory for the secretion of heterologous proteins. *Journal of Biotechnology, 258,* 101–109.
28. Santini, C. L., Bernadac, A., Zhang, M., Chanal, A., Ize, B., Blanco, C., et al. (2001). Translocation of jellyfish Green Fluorescent Protein via the Tat system of *Escherichia coli* and change of its periplasmic localization in response to osmotic up-shock. *Journal of Biological Chemistry, 276*(11), 8159–8164.
29. Dammeyer, T., & Tinnefeld, P. (2012). Engineered fluorescent proteins illuminate the bacterial periplasm. *Computational and Structural Biotechnology Journal, 3*(4), e201210013.
30. He, X. M., Yuan, B. F., & Feng, Y. Q. (2017). Facial synthesis of nickel(II)-immobilized carboxyl cotton chelator for purification of histidine-tagged proteins. *Journal of Chromatography, B: Analytical Technologies in the Biomedical and Life Sciences, 1043,* 122–127.
31. Zárate, X., Henderson, D. C., Phillips, K. C., Lake, A. D., & Galbraith, D. W. (2010). Development of high-yield autofluorescent protein microarrays using hybrid cell-free expression with combined *Escherichia coli* S30 and wheat germ extracts. *Proteome Science, 8*(1), 32.
32. Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology, 97*(9), 3811–3826.
33. Sletta, H., Tøndervik, A., Hakvåg, S., Vee Aune, T. E., Nedal, A., Aune, R., et al. (2007). The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Applied and Environmental Microbiology, 73*(3), 906–912.
34. Suominen, I., Meyer, P., Tilgmann, C., Glumoff, T., Glumoff, V., Kapyla, J., et al. (1995). Effects of signal peptide mutations on processing of *Bacillus stearothermophilus* a-amylase in *Escherichia coli*. *Microbiology, 141*(3), 649–654.
35. Adams, H., Scotti, P. A., De Cock, H., Luirink, J., & Tommassen, J. (2002). The presence of a helix breaker in the hydrophobic core of signal sequences of secretory proteins prevents recognition by the signal-recognition particle in *Escherichia coli*. *European Journal of Biochemistry, 269*(22), 5564–5571.
36. Green, E. R., & Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. In *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition* (pp. 215–239). American Society of Microbiology.
37. Heim, R., & Tsien, R. Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology, 6*(2), 178–182.