The Protocatechuate 3,4-Dioxygenase Solubility (PCDS) Tag Enhances the Expression and Solubility of Heterogenous Proteins in Escherichia coli

Lei Zou¹, Sha Li¹, Nan Li², Shi-Long Ruan², Jing Chen¹, Jing Wu¹, Dazhong Yan¹ and Hong-Jun Chao¹*

¹School of Life Science and Technology, Wuhan Polytechnic University, Wuhan, China, ²Daye Public Inspection and Test Centre, Huangshi, China

Escherichia coli has been developed as the most common host for recombinant protein expression. Unfortunately, there are still some proteins that are resistant to high levels of heterologous soluble expression in E. coli. Protein and peptide fusion tags are one of the most important methods for increasing target protein expression and seem to influence the expression efficiency and solubility as well. In this study, we identify a short 15-residue enhancing solubility peptide, the PCDS (protocatechuate 3,4-dioxygenase solubility) tag, which enhances heterologous protein expression in E. coli. This PCDS tag is a 45-bp long sequence encoding a peptide tag involved in the soluble expression of protocatechuate 3,4-dioxygenase, encoded by the pcaHG98 genes of Pseudomonas putida NCIMB 9866. The 45-bp sequence was also beneficial for pcaHG98 gene amplification. This tag was shown to be necessary for the heterologous soluble expression of PcaHG98 in E. coli. Purified His⁶-PcaHG98e04-PCDS exhibited an activity of 205.63 ± 14.23 U/mg against protocatechuate as a substrate, and this activity was not affected by a PCDS tag. This PCDS tag has been fused to the mammalian yellow fluorescent protein (YFP) to construct YFP-PCDS without its termination codons and YFPT-PCDS with. The total protein expressions of YFP-PCDS and YFPT-PCDS were significantly amplified up to 1.6-fold and 2-fold, respectively, compared to YFP alone. Accordingly, His⁶-YFP-PCDS and His⁶-YFPT-PCDS had 1.6-fold and 3-fold higher soluble protein yields, respectively, than His⁶-YFP expressed under the same conditions. His⁶-YFP, His⁶-YFP-PCDS, and His⁶-YFPT-PCDS also showed consistent fluorescence emission spectra, with a peak at 530 nm over a scanning range from 400 to 700 nm. These results indicated that the use of the PCDS tag is an effective way to improve heterologous protein expression in E. coli.

Keywords: fusion tag, heterogenous protein, solubility expression, protocatechuate 3,4-dioxygenase, solubility peptide
INTRODUCTION

Since the mammalian hormone somatostatin was first used to realize heterologous expression in *Escherichia coli* (Itakura et al., 1977), the expression of recombinant protein has developed rapidly for various applications, including industrial enzyme production (Headon and Walsh, 1994), biopharmaceuticals (Jozala et al., 2016), and vaccine production (Christodoulides et al., 2001). The production of recombinant proteins has been implemented in many different prokaryotic and eukaryotic host systems, including *E. coli*, yeast, filamentous fungi, insect cells, *Arabidopsis*, and even in mammalian cell culture lines (Demain and Vaishnav, 2009). Among these systems, *E. coli* is often preferred due to its fast growth, high cell density cultures, rich complex media, and easy transformation (Rosano and Ceccarelli, 2014). In order to improve the heterologous protein expression of *E. coli* systems, a number of protein tags have been developed, including the maltose-binding protein (MBP; Kapust and Waugh, 1999), glutathione S-transferase (GST; Smith and Johnson, 1988), small ubiquitin-related modifier (SUMO; Marblestone et al., 2006), N-utilization substance protein A (NusA; Davis et al., 1999), thioredoxin (Trx; LaVallie et al., 1993), and ubiquitin (Bachmair et al., 1986; Varshavsky, 2005). In addition, there are some short peptide tags for *E. coli* expression systems wherein the amino acid sequences are generally 15 residues or less (Ki and Pack, 2020), including the Arg-tag (Sassenfeld and Brewer, 1984), FLAG-tag (Hopp et al., 1988), His-tag (Hochuli et al., 1987), c-myc-tag (Ferrando et al., 2001), S-tag (Karpinsky et al., 1994), Strep-tags (Schmidt and Skerra, 1993), and Fh8 and H tags (Costa et al., 2013). These peptide tags can improve the solubility of a target protein by regulating the process of protein transcription and translation (Ki and Pack, 2020). However, a specific fusion tag is not suitable for the protein expression of all proteins (Zhao et al., 2019), and the function of an expressed protein may be negatively affected by the specific fusion tag used for expression (Arnau et al., 2006). Therefore, the identification of additional fusion protein and peptide tags is conducive to finding more beneficial common features for protein heterogeneous expression.

The chemical 2,4-xylenol, known to be harmful to human health, is listed as a toxic pollutant by the U.S. Environmental Protection Agency due to its damage to the environment. The two compounds 2,4-xylenol and p-cresol can be catabolized together and have the same enzymes to catalyze the oxidation of *para*-methyl groups in *Pseudomonas putida* NCIMB 9866 (Chen et al., 2014). Moreover, the metabolism of 2,4-xylenol and p-cresol is through the protocatechuate (3,4-dihydroxybenzoate, PCA) pathway (Elmorsi and Hopper, 1977; Chen et al., 2014; Chao et al., 2016). Protocatechuate is further metabolized through the ortho ring cleavage pathway. *pcaHG* genes encoding protocatechuate-3,4-dioxygenase are located on the *pch* gene cluster in *P. putida* NCIMB 9866 (Chen et al., 2014). In this study, we found a 45-bp-coded peptide tag involved in the soluble expression of protocatechuate 3,4-dioxygenase encoded by the *pcaHG* genes of *P. putida* NCIMB 9866 in *E. coli*. These *pcaHG* genes were cloned and expressed in vitro, and a 45-bp sequence termed as the PCDS tag (protocatechuate 3,4-dioxygenase solubility tag) required for its soluble expression is characterized. PcaHG98 appears as an inclusion body without the PCDS tag in *E. coli*. Moreover, this tag could also significantly promote the heterogenous soluble expression of yellow fluorescent protein (YFP), which has relatively low expression levels in *E. coli*.

MATERIALS AND METHODS

Strains, Plasmids, Culture Media, Primers, and Chemicals

The strains used in this study are listed in Table 1. *Pseudomonas putida* NCIMB 9866 was grown at 30°C in minimal medium (MM; Chapman and Hopper, 1968) on different compound carbon sources. *Escherichia coli* was grown in lysogeny broth (LB) on a rotary shaker (200 rpm) at 37°C under aerobic conditions. The culture medium is usually supplemented with antibiotics at a final concentration of 50 μg/ml−1 for kanamycin sulfate (Kan) and 100 μg/ml−1 for ampicillin sodium (Amp), as necessary. The chemicals used in the experiments were purchased from Aladdin Reagents (Shanghai, China) or Sigma-Aldrich (St. Louis, MO, United States). Taq DNA polymerase and Pfu DNA polymerase were obtained from TaKaRa Bio (Dalian, China). *TransEco FastPfu* DNA Polymerase was purchased from TransGen Biotech (Beijing, China).

Gene Cloning and Expression, Protein Purification

Plasmid preparation and DNA manipulation were carried out as described previously (Green and Sambrook, 2012). The primers used in this study are listed in Table 2 and are shown in Figure 1. All the targeted genes were amplified by PCR using Taq DNA polymerase or *TransEco FastPfu* DNA Polymerase (Transgen, Beijing, China), according to its manufacturer’s recommendation. The *pcaHG98* gene was amplified with PCR with primers pcahg9803 and pcahg9804 from the extracted genomic DNA of strain NCIMB 9866. The PCR reaction volume usually uses 50 μl and consists of 5–30 ng Plasmid DNA or 100 ng Genomic DNA, 1x *TransEco FastPfu* Buffer, 250 μM dNTP, 0.4 μM of each primer, 2.5 U of *TransEco FastPfu* DNA Polymerase or Taq DNA polymerase and ddH₂O up to 50 μl. PCR amplification was performed with *TransEco FastPfu* DNA polymerase as follows: denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 58°C for 1 min, and 72°C for 1 min (1 kb/ min for Taq, 2 kb/min for FastPfu), and a final extension cycle at 72°C for 5 min. The PCR fragments were digested with *NdeI/HindIII* (Thermo Fisher Scientific, Shanghai, China) before being cloned into pET-28a (+) with T4 DNA ligase (TaKaRa, Dalian, China) to produce pET-28a-*pcaHG98*e01. The *pcaHG98* gene was amplified using primers pcahg9803 and pcahg9804 from plasmid pET-28a-*pcaHG98*e01 and cloned into *NdeI/HindIII* sites of pET-28a (+) with T4 DNA ligase to produce pET-28a-*pcaHG98*e02. The *pcaHG98* gene was amplified using primers pcahg9802 and pcahg9803 from plasmid pET-28a-*pcaHG98*e01 and cloned into *NdeI/HindIII* sites of pET-28a (+) with T4
DNA ligase to produce pET-28a-pcaHG98e03. The pcaHG98 gene was amplified using primers pcahg9801 and pcahg9804 from plasmid pET-28a-pcaHG98e01 and cloned into Ndel/HindIII sites of pET-28a (+) with T4 DNA ligase to produce pET-28a-pcaHG98e04.

The nucleotide sequence for YFP was synthesized according to the sequences of pcDNA3YFP, which was a gift from Doug Golenbock (Addgene plasmid # 13,033). The YFP gene was amplified using primers pYFP03 and pYFP04 from this synthesized YFP gene (Sangon Biotech, Shanghai, China) at 4°C for 48 h against phosphate buffer (PB) before being further preserved in glycerol at −80°C.

The expression and purification of YFP and eGFP are performed according to that of His<sub>6</sub>-PcaHG98e04-PCDS. The total protein amounts of E. coli BL21 (DE3)/pET-28a-YFP, pET-28a-eGFP, and pET-28a-eGFP-PCDS were further diaлизed away from imidazole with a Spectra/Por CE dialysis membrane with a molecular weight cut-off of 10,000 Da (Spectrum Laboratories Inc., Shanghai, China) at 4°C for 48 h against phosphate buffer (PB) before being further preserved in glycerol at −80°C.

**TABLE 1 | Strains or plasmids used in this study.**

| Strain or plasmid | Characteristic(s) * or purpose | Reference or source |
|-------------------|-------------------------------|---------------------|
| **Pseudomonas putida NCIMB** | | |
| 9,866 | 2,4-xylene and p-cresol-utilizing, wild-type strain, Amp<sup>+</sup>, Kan<sup>+</sup>, Tc<sup>+</sup> | Chapman and Hopper, 1968 |
| **Escherichia coli** | | |
| Trans1-TI Phage Resistant BL21 (DE3) | F<sup>−</sup> phi80 (lacZΔM15 ΔlacX74 hsdR (r<sub>K</sub>m<sub>K</sub> D)) ΔrepA1398 endA1 tonA<sup>+</sup> F<sup>−</sup> ompT hsdS (r<sub>K</sub>m<sub>K</sub>) gal dcm (DE3) | TransGen Biotech Novagen |
| **Plasmid** | | |
| pET-28a (+) | Expression vector, Kan<sup>+</sup> | Novagen |
| pET-28a-pcaHG98e01 | Expression vector for pcaHG, PCR amplified with primers pcahg9803 and pcahg9804 from strain NCIMB 9866 genome, inserted into the Ndel-Hind III restriction sites in pET-28a (+) | This study |
| pET-28a-pcaHG98e02 | Expression vector for pcaHG, PCR amplified with primers pcahg9801 and pcahg9802 from strain NCIMB 9866 genome, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET-28a-pcaHG98e03 | Expression vector for pcaHG, PCR amplified with primers pcahg9802 and pcahg9803 from strain NCIMB 9866 genome, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET-28a-pcaHG98e04 | Expression vector for pcaHG, PCR amplified with primers pcahg9801 and pcahg9804 from strain NCIMB 9866 genome, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET28a-YFP | Expression vector for the cDNA sequence of YFP synthesized according to the sequences from pcDNA3YFP, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET28a-YFP-PCDS | Expression vector for the cDNA sequence of YFP synthesized according to the sequences of pcDNA3YFP and PCDS tag with a stop codon, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET28a-YFP-PCDS | Expression vector for the cDNA sequence of YFP synthesized according to the sequences of pcDNA3YFP and PCDS tag without a stop codon, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pEGFP-N1 | Vector encoding a red-shifted variant of wild-type GFP that has fluorescence activity in mammalian cells | Clontech |
| pET28a-eGFP | Expression vector for the eGFP gene of pEGFP-N1, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |
| pET28a-eGFP-PCDS | Expression vector for the eGFP gene of pEGFP-N1 and PCDS tag, inserted into the Ndel-Hind III restriction sites of pET-28a (+) | This study |

* Amp<sup>+</sup> resistant to ampicillin; Kan<sup>+</sup>, Tc<sup>+</sup>, sensitive to kanamycin and tetracycline, respectively.
**TABLE 2 | Primers used in this study.**

| Published Primer | Sequence (5’-3’) | Underlined comment |
|------------------|------------------|-------------------|
| pcahg98S01       | CGCGAGTCGAATGAGGTTGG |                  |
| pcahg98S02       | CAAGCCATTGACATAGCAACGAGTG |                  |
| pcahg9801        | GGGAAATTCCATATGTCGACGGCTCAAGAACGTCGG | NdeI |
| pcahg9802        | CCAAGCTTGAAGTCAAAAAAGACTGTCTCG | HindIII |
| pcahg9803        | GGGAAATTCCATATGTCGACGGCTCAAGAACGTCGG | NdeI |
| pcahg9804        | CCAAGCTTGAAGTCAAAAAAGACTGTCTCG | HindIII |
| pYFP01           | CCGCGAGGGCGAGCGCATATGTAAGAGCGGAG | 15 bp overlap for fusion |
| pYFP02           | TGCGGTGGCTGCGGGCATTTGAGGTCGGG | 15 bp overlap for fusion |
| pYFP03           | CCGCGAGGGCGAGCGCATATGTAAGAGCGGAG | 15 bp overlap for fusion and partial PCDS tag |
| pYFP04           | TGCGGTGGCTGCGGGCATTTGAGGTCGGG | 15 bp overlap for fusion and partial PCDS tag |
| pegf03           | CCGCGAGGGCGAGCGCATATGTAAGAGCGGAG | 15 bp overlap for fusion |
| pegf04           | GACGGAGCTGATTTCTTATGACGCTGCGTCCCG | 15 bp overlap for fusion |
| pET28a-pcas03    | CAGCTGCTGCTGCGACAGATAGCAGAACGCGCTCAGCTGCGGGCGCCACTCG | 15 bp overlap for fusion and partial PCDS tag |
| pET28a-pcas04    | GTGCCGGAGAAGCTGCTGCCGGCGATGTGCCTCTG | 15 bp overlap for fusion and partial PCDS tag |

**FIGURE 1 | Primers, PCR reactions, plasmids, and enzyme activity information used in this study. Primers are shown with a solid arrow. PCR without product (×) and with product (√). Enzyme without activity (×) and with activity (√).**

_E. coli_ BL21 (DE3)/pET-28a-YFP-PCDS and _E. coli_ BL21 (DE3)/pET-28a-YFP-PCDS refer to ultrasound-broken cells after centrifugation collection of 100 ml IPTG induced expression. The amounts of purified His<sub>6</sub>-YFP, His<sub>6</sub>-YFP-PCDS, and His<sub>6</sub>-YFP<sub>t</sub>-PCDS were detected directly after purification. Protein concentrations were determined with the common Bradford assay (Bradford, 1976).

**Molecular Weight Determination**

Molecular weight of the purified recombinant His<sub>6</sub>-PcaHG98e04-PCDS was determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and liquid chromatograph high-resolution accurate mass spectrometry (LC-MS). LC-MS data were obtained using a Thermo Scientific™ Q Exactive™ mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). The recombinant His<sub>6</sub>-PcaHG98e04-PCDS was purified by SDS-PAGE. The His<sub>6</sub>-PcaHG98e04-PCDS sample was dissolved in 2% acetonitrile and 0.1% formic acid. The mixture was vortex mixed and centrifuged at 12,000 g for 15 min. The sample was then purified by ultra-filtration, concentrated using Amicon Ultra-10K centrifugal filter device and stored in 0.1% formic acid. MS was performed using a Thermo Scientific Q Exactive mass spectrometer operated in the data-dependent/dynamic exclusion mode. A full resolution setting of 70,000 (full width at half maximum, FWHM) was used for full scan MS. All data were processed using Thermo Scientifc™ Pinpoint™ software (Thermo Fisher Scientific, Waltham, MA, United States).

**Enzyme Activity Assays**

PCA-3,4-dioxygenase activities were detected by referring to previously described methods (Stanier and Ingraham, 1954; Gross et al., 1956). For the assay of PCA-3,4-dioxygenase activity, the reaction mixture contained 50 mM of PB (pH 7.5) and 200 μM of cell extracts or purified His<sub>6</sub>-PcaHG98e04-PCDS. All compounds except the substrate were added to the reference cuvette, and the enzyme activity assay was initiated by the addition of 30 μg of PCA. The spectra in the range of 220–400 nm were recorded every minute. The molar extinction coefficients for PCA at 290 nm and 3-carboxy-cis,cis-muconate at 270 nm were 3,890 M<sup>1</sup>cm<sup>−1</sup> (Gross et al., 1956) and 6,390 M<sup>1</sup>cm<sup>−1</sup> (Iwagami et al., 2000), respectively. One unit of enzyme activity was defined as the protein amount required for the production (or disappearance) of 1 μmol of product (or substrate) min<sup>−1</sup> at 30°C. The protein amounts of _E. coli_ BL21 (DE3)/pET-28a-pcaHG98e01 and _E. coli_ BL21 (DE3)/pET-28a-pcaHG98e04 refer to the soluble protein after the 100 ml IPTG induced cell is ultrasonically broken and centrifuged. The amount of purified...
His6-PcaHG98e04-PCDS was directly detected. All enzyme activity assays were run in triplicate in three independent experiments.

**Fluorescence-Activated Cell Sorting and Fluorescent Spectral Analysis**

The BL21 (DE3)/pET-28a(+) was used as a nonfluorescence control. The culture broth of BL21 (DE3)/pET-28a(+), BL21 (DE3)/pET-28a-YFP, BL21 (DE3)/pET-28a-YFP-PCDS, and BL21 (DE3)/pET-28a-YFPT-PCDS grown under the same conditions according to that of YFP expression. Then the cells were pelleted, washed twice with 50 mM of Tris-HCl buffer (pH 7.4), resuspended in Tris-HCl buffer, and transferred to a 2-ml round bottom centrifuge tube, respectively. The cell concentration was diluted to $5 \times 10^4$–$5 \times 10^5$ and counted. Fluorescence-activated cell sorting (FACS) analysis was carried on a Cytosens S flow cytometer (Beckman Coulter, IN, United States). The laser beam for cells was 488-nm, and a 529 ± 14-nm FL1 band pass filter used to filter the emission signal. The light signals from the cells are converted to voltage values and the minimum voltage value (400 V) was used, with a threshold of 9,200. FACS analysis was run in triplicate in three independent experiments. Data were analyzed using FlowJo v10.0.4 (Tree Star Inc., Ashland, United States) software.

Purified recombinant His$_6$-YFP, His$_6$-YFP-PCDS, and His$_6$-YFPT-PCDS were diluted to 0.1 mg/ml with 50 mM of Tris-HCl buffer (300 mM of NaCl, pH 7.4) and monitored by a Fluorescence Spectrophotometer (F96PRO, Shanghai Kingdak Scientific Instrument, Shanghai, China). The emission and excitation wavelengths of the recombinant protein were 400 and 700 nm, respectively. Fluorescent spectral analysis was run in triplicate in three independent experiments.

**Statistical Analysis**

The statistical analysis used in this study was performed by SPSS version 23.0 software (IBM SPSS Inc., New York, NY, United States). One-way analysis of variance (ANOVA) was used to calculate the values of $p$ for PCA-3,4-dioxygenase activity analyses. Paired-sample tests were used to calculate probability ($p$) values for the expression of YFP. The values of $p$ of 0.05 and 0.01 were considered as statistically significant and highly statistically significant, respectively.

**RESULTS**

**pcaHG98 Gene Cloning and Expression**

*pcaHG98*, encoded a putative protocatechuate 3,4-dioxygenase, was cloned into pET-28a (+). The *pcaHG98* gene was amplified by PCR with primers pcahg9801 and pcahg9802 (Figure 1) using strain NCIMB 9866 gDNA or bacteria lysate as a template. But the gene could not be amplified into PCR products using either Taq DNA polymerase or Pfu DNA polymerase. In order to determine this sequence on the genome of strain NCIMB 9866, two primers, pcahg98s01 and pcahg98s02 (Figure 1), were designed for sequencing analysis, and the results show that the sequence was consistent with previous reports (Chen et al., 2014). Then, two primers, pcahg9803 and pcahg9804 were redesigned (Figure 1) successfully cloned the *pcaHG98* gene into pET-28a (+) to generate pET-28a-pcaHG98e01.

The pET-28a-pcaHG98e01, which contained the amplified *pcaHG98* gene from strain NCIMB 9866 gDNA, was expressed in *E. coli* BL21(DE3). The molecular masses calculated from the nucleotide sequence of PcaG and PcaH were 22.7 and 26.3 kDa, respectively, which corresponded to the proteins observed by SDS-PAGE (Figure 2A). The PcaHG9801 protein, however, did not bind to Ni$^{2+}$-nitrilotriacetic acid agarose for chromatography. We obtained the same results three separate times. This may have been because the *pcaHG98* gene included 54 bp of upstream sequence and 45 bp of downstream sequence in pET-28a-pcaHG9801. The start and stop codons of the *pcaHG98* gene still retained with it in this recombinant. Therefore, a His-tag is not fused to the *pcaHG98* gene due to the presence of its start and stop codons.

The *pcaHG98* gene in vector pET-28a-pcaHG98e02 showed no effective expression, and most of its proteins were insoluble, of which the $\alpha$-subunit PcaG was not determined by SDS-PAGE (Figure 2A). The *pcaHG98* gene was amplified from pET-28a-pcaHG98e01 to produce pET-28a-pcaHG98e03, in which the *pcaHG98* gene is expressed in the form of inclusion body (Figure 2B). However, the *pcaHG98* gene in pET-28a-PcaHG9804 was expressed in the form of soluble proteins, which was amplified using primers pcahg9801 and pcahg9804 (Figure 2B).

**Purification and Biochemical Properties of PcaHG98e04**

The PcaHG98e04 protein was constructed using the primers pcahg9801 and pcahg9804, and the primer pcahg9801 (which removed the start codon ATG) starts with the *pcaH* gene 5'-end, with the lead pcahg9804 on the outside of the *pcaG* gene (containing 45 bases: GACATCAGCAGGGCCCGGCGG CGG). Compared with the insoluble PcaHG98e02, these 45 bases translated into polypeptides make the His$_6$-PcaHG98e04-PCDS protein soluble, the reasons for which are further explored below (Figure 2B).

His$_6$-PcaHG98e04-PCDS, purified using Ni$^{2+}$-nitrilotriacetic acid agarose chromatography, was then determined using LC–MS. The molecular weight of the protein detected was 27.6 kDa (Figure 3), which contained 22.7 kDa encoded by the *pcaG* gene and 3.3 kDa encoded by the 45-bp (the underlined DNA sequence) and other end sequences (GACATCAGCAGGGCCCGGCGG CGG). Compared with the insoluble PcaHG98e02, these 45 bases translated into polypeptides make the His$_6$-PcaHG98e04-PCDS protein soluble, the reasons for which are further explored below (Figure 2B).

**His$_6$-PcaHG98e04-PCDS Catalyzes the Oxidization of PCA**

Protocatechuate 3,4-dioxygenase catalyzes the cleavage of PCA into $\beta$-carboxy-cis, cis-muconate (Harwood and Parales, 1996). A series of DNA fragments of different lengths containing the
Zou et al. PCDS Tag Enhances Soluble Expression

Frontiers in Microbiology | www.frontiersin.org 6 November 2021 | Volume 12 | Article 779541

The complete reading frame of the pcaHG98 gene was ligated into the expression vector pET-28a (+), to make the plasmids pET-28a-pcaHG98e01, pET-28a-pcaHG98e02, pET-28a-pcaHG98e03, and pET-28a-PcaHG98e04. These recombinant plasmids were then individually expressed in *E. coli* BL21 (DE3). After induction with IPTG, *E. coli* BL21 (DE3)/pET-28a-pcaHG98e01 and *E. coli* BL21 (DE3)/pET-28a-pcaHG98e04 cell extracts were found to contain PCA 3,4-dioxygenase with specific activities of 74.7 ± 5.31 and 72.8 ± 4.41 U/mg against PCA as a substrate, respectively. No activity was detectable in *E. coli* BL21 (DE3)/pET-28a-pcaHG98e02 and *E. coli* BL21 (DE3)/pET-28a-pcaHG98e03, where the expression vector contained no soluble PcaHG98 protein. Purified His<sub>6</sub>-PcaHG98e04-PCDS exhibited 205.63 ± 14.23 U/mg activity against PCA as a substrate, which was not affected by the PCDS tag.

**Figure 4** shows the rapid transformation of PCA (λ<sub>max</sub> = 290 nm) to 3-carboxy-cis, cis-muconate (λ<sub>max</sub> = 270 nm) in the cell extract, as described by Harwood and Parales (1996).

**Biochemical Properties of Recombinant His6-YFP, His6-YFP-PCDS, and His6-YFPT-PCDS**

It has been reported that the YFP has relatively low expression in *E. coli* compared to that of codon-optimized YFP (Nguyen et al., 2019). The YFP gene was synthesized according to the sequence of the mammalian expression vector pcDNA3YFP and cloned into the *E. coli* expression vector pET28a (+) in this study. Based on our results, the YFP gene was expressed in our *E. coli* system and had a certain amount of solubility (**Figure 5B**). The 45-bp PCDS tag enhanced the soluble expression of the YFP gene in *E. coli* after fusing the PCDS tag to the YFP C-terminus (**Figure 5B**). This YFP-PCDS did not have YFP termination codons, whereas YFPt-PCDS contained YFP termination codons. The PCDS tag enhanced the expression level of YFP in *E. coli* was analyzed by fluorescence-activated cell sorting (FACS). FACS analysis showed that the fluorescence intensity of BL21 (DE3)/pET-28a-YFP-PCDS and BL21 (DE3)/pET-28a-YFPT-PCDS was higher than that of BL21 (DE3)/pET-28a-YFP, and BL21 (DE3)/pET-28a-YFPT-PCDS was the highest (**Figure 5A**). BL21 (DE3)/pET-28a-YFP showed weaker fluorescence intensity, as did the control BL21 (DE3)/pET-28a(+).
was significantly amplified up to 1.6-fold and 2-fold compared to that of YFP alone, respectively. Accordingly, His$_6$-YFP-PCDS and His$_6$-YFPt-PCDS had 1.6-fold and 3-fold higher soluble expression yield, respectively, than that of His$_6$-YFP under the same expression and purification conditions (Figure 5C). All protein amounts were determined three times in three independent experiments. His$_6$-YFP, His$_6$-YFP-PCDS, and His$_6$-YFPt-PCDS also showed consistent fluorescence emission spectra, with a peak at 530 nm across the scanning range from 400 to 700 nm (Figure 5D).

**DISCUSSION**

The β-ketoacid pathway is preceded by the preliminary conversion of a broad range of organic compounds into one of two aromatic rings, catechol or PCA (Harwood and Parales, 1996; Wells and Ragauskas, 2012). Protocatechuate 3,4-dioxygenase catalyzes the cleavage of PCA into β-carboxy-cis, cis-muconate (Harwood and Parales, 1996). pcaHG98 encodes a putative protocatechuate 3,4-dioxygenase in NCIMB 9866. This gene, unfortunately, cannot be directly amplified by PCR with primers pcahg9801 and pcahg9802 using strain NCIMB 9866 gDNA or bacteria lysate as a template using either the Taq or Pfu DNA polymerase. Sequencing analysis results show that the sequence of the pcaHG98 gene was consistent with genome sequencing data. Then, two primers, pcahg9803 and pcahg9804 were redesigned (Figure 1), and the pcaHG98 gene was successfully cloned into pET-28a (+) to generate pET-28a-pcaHG98e01. The recombinantly expressed PcaHG98e01 (Figure 2A) had protocatechuate 3,4-dioxygenase activity (Figure 4A). These results show that the 54 bp and 45 bp of the pcaHG98e01 gene (Figure 1) contain the sequences needed for its solubility expression, which were also beneficial for gene amplification. In order to determine the role of these two sequences, we constructed pET-28a-pcaHG98e03, which contained 54 bp on the N terminus, and pET-28a-pcaHG98e04, which contained 45 bp on the C terminus (Figure 1). The PcaHG98e03 protein was insoluble (Figure 2B), and the PCA 3,4-dioxygenase activity was not detectable in E. coli BL21 (DE3)/pET-28a-pcaHG98e03 (Figure 4C), while the PcaHG98e04 protein was soluble (Figure 2B), and E. coli BL21 (DE3)/pET-28a-pcaHG98e04 cell extracts were found to exhibit activity, with a specific activity of 72.8 ± 4.41 U/mg against PCA as a substrate. Therefore, the 45 bp on the C terminus of the pcaHG98e04 gene was beneficial for its soluble protein expression, which we named as the PCDS tag.

In order to analyze the solubility impact of the PCDS tag, we selected YFP, with its pI of 5.910 and its relatively low level of expression in E. coli according to the literature (Nguyen et al., 2019). Based on our results, the YFP gene was expressed in the E. coli system with a certain amount of solubility (Figure 5B). FACS analysis showed that the
fluorescence intensity of YFP-PCDS and YFPt-PCDS in *E. coli* BL21(DE3) was higher than that of YFP, and YFPt-PCDS was the highest (Figure 5A). The PCDS tag could enhance the soluble expression of the YFP gene by fusing the PCDS tag to the YFP C-terminus in *E. coli* (Figure 5B). The total protein expression of YFP-PCDS and YFPt-PCDS without and with a termination codon was significantly amplified up to 1.6-fold and 2-fold compared to that of YFP alone, respectively. Accordingly, His<sub>6</sub>-YFP-PCDS and His<sub>6</sub>-YFPt-PCDS had 1.6-fold and 3-fold higher soluble expression yield, respectively, than that of His<sub>6</sub>-YFP under the same conditions. They also showed a consistent fluorescence emission spectra, with a peak at 530 nm over a scanning range from 400 to 700 nm (Figure 5D). This result also showed that the presence of the stop codon also affected the improvement in solubility induced by the PCDS tag, which was located at a similar position as the *pcaHG98* gene. In addition, the 45 bp sequence outside the end of the *pcaHG98* gene affects its PCR cloning from the genome, but the reason is confusing and needs further analysis. A protein has its lowest solubility when the ambient pH is equal to its pI, at which point it shows a zero net charge. Therefore, optimizing the net charge to a positive or negative charge improves the solubility of a protein (Lawrence et al., 2007). Depending on the pI...
value of the target protein, the use of fusion tags to induce a rejection of static interaction between proteins can provide sufficient time for the correct folding of proteins, thus preventing protein aggregation (Zhang et al., 2004; Kato et al., 2007; Paraskevopoulou and Falcone, 2018). The pI values of PcaH and PcaG were 9.599 and 4.378, respectively. The PCDS tag having a high pI value of 9.849 may promote PcaHG soluble expression by influencing the static interaction between them. In this respect, the PCDS tag is similar to the protein fusion tag CBD. CBD also has a high pI value (Hopp et al., 1988; Marblestone et al., 2006) that improves the soluble expression of heterogeneous proteins in E. coli (Murashima et al., 2003). Most peptide fusion tags are located at the N-terminus of an expression protein, which promotes the correct folding of a protein to enable its solubility expression by affecting transcription initiation (Seo et al., 2013; Nguyen et al., 2019). In contrast to this, some studies have shown that poly-lysine or poly-arginine tags fused to the C-terminus of a target protein are more likely to enhance its solubility than a tag fused to the N-terminus (Park et al., 2003; Hage et al., 2015; Islam et al., 2015; Nautiyal and Kuroda, 2018). A PCDS tag located at the C-terminus of PcaG may promote PcaHG soluble expression by influencing the static interaction between them.

In fact, the soluble expression of recombinant proteins is controlled by many factors, including the host organism, the types of expression vector used, codon bias, culture conditions, transcription initiation, mRNA stability, and protein toxicity (Rosano and Ceccarelli, 2014). An inclusion body is a common hurdle for the heterologous expression of recombinant proteins, which is usually solved by various solubility-enhancing tags, including protein fusion tags and peptide tags (Paraskevopoulou and Falcone, 2018). These peptide tags have been developed and applied to enhance heterologous protein expression, including FLAG-tags (Einhammer and Jungbauer, 2001), Arg-tags (Sassenfeld and Brewer, 1984; Terpe, 2003), Fh8 and H tags (Costa et al., 2013), and NT11-tags (Table 3; Terpe, 2003; Nguyen et al., 2019; Ki and Pack, 2020). The advantage of these peptide tags for heterologous expression is that their amino acid sequences usually only have 15 residues or less and therefore do not affect the structure or activity of heterologous proteins (Kato et al., 2007; Paraskevopoulou and Falcone, 2018; Paraskevopoulou and Falcone, 2018).

### Table 3 | Peptide tags to improve heterologous protein expression in *Escherichia coli*

| Tag | Sequence | Number of amino acids | pI | Net charge at pH 7 | Size (Da) | GRAVY | Tagged position (terminus) | References |
|-----|----------|-----------------------|----|-------------------|----------|-------|---------------------------|------------|
| NT11 | VSEPDIYNYEK | 11 | 4.31 | –2 | 1,380 | –1.991 | N- | Nguyen et al., 2019 |
| SAP (S1) | (AEAEAKAK)2 | 16 | 7.11 | 0 | 1,614 | –0.950 | N- | Zhao et al., 2019 |
| S1v10 | (ANANARAR)2 | 16 | 12.96 | +4 | 1,667 | –1.100 | N- | Zhao et al., 2019 |
| S1hv1 | (IEIEIKIK)2 | 16 | 7.08 | 0 | 1,951 | 0.400 | N- | Zhao et al., 2019 |
| S1hv3 | (VEVEVKVK)2 | 16 | 7.08 | 0 | 1,839 | 0.250 | N- | Zhao et al., 2019 |
| S1V1 | (AEAEAH)2 | 16 | 5.06 | –4 | 1,651 | –0.775 | N- | Zhao et al., 2016 |
| DDTs | GRGRSRA | 7 | 12.79 | +3 | 758 | –1.900 | N- | Hirose et al., 2011 |
| KRRAAA | 7 | 12.49 | +3 | 742 | –0.814 | N- | Hirose et al., 2011 |
| S Tag | KETAAKFERQHMDSS | 15 | 7.54 | 0 | 1,754 | –1.393 | N- | Kim and Ranes, 1993 |
| T7B9 | EEASYSTTEEMTSEAEAEAE | 39 | 3.6 | –9 | 4,261 | –1.060 | N/V/C- | Zhang et al., 2004 |
| 6×Lys | KKKKKK | 6 | 11.33 | +7 | 914 | –3.900 | C- | Hage et al., 2015 |
| GFIL8 | GFILGFIL | 8 | 5.58 | 0 | 879 | 2.675 | C- | Wang et al., 2015 |
| LBKD | LLALLLKD | 8 | 6.71 | 0 | 940 | 1.925 | C- | Zhou et al., 2012 |
| ELK16 | LLEELKLEELK | 16 | 7.08 | 0 | 1,951 | 0.060 | C- | Wu et al., 2011 |
| A18 | EWLKAVERWLEKELF | 18 | 6.95 | 0 | 2,312 | –0.278 | C- | Xing et al., 2011 |
| SEP | GRRGGRGPRRR | 12 | 13.403 | 9 | 1,595 | –3.475 | C- | Nautiyal and Kuroda, 2018 |
| PCDS | DIAGAAAPVRNAS | 15 | 9.849 | 1 | 1,440 | 0.047 | C- | This study |
Nguyen et al., 2019; Ki and Pack, 2020). Thus, it may not be necessary to remove these peptide tags for further applications of proteins, in contrast to the case of the larger protein fusion tags. In addition, protein and peptide fusion tags are one of the most important methods to increase target protein expression, including expression efficiency and solubility. Therefore, understanding the physicochemical properties of proteins, including their pI value, net charge, and GRAY V values, can help to select and design appropriate fusion tags. Although many protein labels and peptide tags have been reported to promote protein soluble expression, there are still other proteins that cannot achieve heterologous soluble expression, which requires research and development that focused on more and more extensive technologies and fusion tags.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

Arnau, J., Lauritzen, C., Petersen, G. E., and Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr. Purif. 48, 1–13. doi: 10.1016/j.pep.2005.12.002

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science 234, 179–186. doi: 10.1126/science.3018930

Bradford, M. M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Anal. Biochem. 72, 248–254. doi: 10.1016/0003-2697(76)90573-4

Chao, H. J., Chen, Y. F., Fang, T., Xu, Y., Huang, W. E., and Zhou, N. Y. (2016). HipH catalyzes the hydroxylation of 4-hydroxyisophthalate to protocatechuic acid in 2,4-xylanol catabolism by pseudomonas putida NCIMB 9866. Appl. Environ. Microbiol. 82, 724–731. doi: 10.1128/AEM.03105-15

Chapman, P. J., and Hopper, D. J. (1968). The bacterial metabolism of 2,4-xylan. Biochem. J. 110, 491–498. doi: 10.1042/bj1100491

Chen, Y. F., Chao, H., and Zhou, N. Y. (2014). The catabolism of 2,4-xylanol and p-cresol share the enzymes for the oxidation of Para-methyl group in pseudomonas putida NCIMB 9866. Appl. Microbiol. Biotechnol. 98, 1349–1356. doi: 10.1007/s00253-013-5601-z

Christodoulides, M., Jolley, K. A., and Heckels, J. E. (2001). “Recombinant proteins in vaccine development,” in Meningococcal Vaccines: Methods and Protocols, eds. A. J. Pollard and M. C. J. Maiden (Totowa, NJ: Humana Press), 167–180.

Costa, S. J., Almeida, A., Castro, A., Domingues, L., and Besir, H. (2013). The novel FlgH and H fusion partners for soluble protein expression in Escherichia coli: a comparison with the traditional gene fusion technology. Appl. Microbiol. Biotechnol. 97, 6779–6791. doi: 10.1007/s00253-012-4559-1

Davis, G. D., Elsea, C., Newham, D. M., and Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol. Bioeng. 65, 382–388. doi: 10.1002/(SICI)1097-0290(19991120)65:4<382::AID-BIIT2>3.0.CO;2-1

Demain, A. L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. Biotechnol. Adv. 27, 297–306. doi: 10.1016/j.biotechadv.2009.01.008

Einhauser, A., and Jungbauer, A. (2001). The FLAG (TM) peptide, a versatile fusion tag for the purification of recombinant proteins. J. Biochem. Biophys. Methods 49, 455–465. doi: 10.1016/S0165-022x(01)00213-5

Einhauer, A., and Jungbauer, A. (2001). The FLAG (TM) peptide, a versatile fusion tag for the purification of recombinant proteins. J. Biochem. Biophys. Methods 49, 455–465. doi: 10.1016/S0165-022x(01)00213-5

Ferrando, A., Koncz-Kalman, Z., Farras, R., Tiburcio, A., Schell, J., and Koncz, C. (2001). Detection of in vivo protein interactions between Snf1-related kinase subunits with intron-tagged epitope-labelling in plants cells. Nucleic Acids Res. 29, 3685–3693. doi: 10.1093/nar/29.17.3685

Green, M. R., and Sambrook, J. (2012). Molecular Cloning: A Laboratory Manual, 4th Edn. New York: Cold Spring Harbor Laboratory Press.

Gross, S. R., Gafford, R. D., and Tamut, E. L. (1956). The metabolism of proteocathecic acid by Neurospora. J. Biol. Chem. 219, 781–796. doi: 10.1016/S0021-9258(18)65737-0

Hage, N., Renschw, J. G., Winkler, G. S., Gellert, P., Stolnik, S., and Falcone, F. H. (2013). Improved expression and purification of the helicobacter pylori adhesin BabA through the incorporation of a hexa-lysine tag. Protein Expr. Purif. 96, 25–30. doi: 10.1016/j.pep.2014.10.009

Harwood, C. S., and Parales, R. E. (1996). The beta-ketoacylase pathway and the biology of self-identity. Annu. Rev. Microbiol. 50, 553–590. doi: 10.1146/annurev.micro.50.1.553

Headon, D. R., and Walsh, G. (1994). The industrial production of enzymes. Biotechnol. Adv. 12, 635–646. doi: 10.1016/0739-9750(94)90004-3

Hirose, S., Kawamura, Y., Mori, M., Yokota, K., Noguchi, T., and Goshima, N. (2011). Development and evaluation of data-driven designed tags (DDTs) for controlling protein solubility. New Biotechnol. 28, 225–231. doi: 10.1016/j.nbt.2010.08.012

Hochuli, E., Bobeli, H., and Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. J. Chromatogr. A 411, 177–184. doi: 10.1016/S0021-9673(00)93694-9

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., et al. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. Nat. Biotechnol. 6, 1204–1210. doi: 10.1038/nbt1088-1210

Islam, M. M., Nakamura, S., Noguchi, K., Yohda, M., Kidokoro, S.-I., and Kuroda, Y. (2015). Analysis and control of protein crystallization using 4-hydroxyisophthalate hydroxylase from pseudomonas putida NCIB 9866. Eur. J. Biochem. 76, 197–208. doi: 10.1111/j.1432-1033.1977.tb11585.x

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., et al. (1977). Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science 198, 1056–1063. doi: 10.1126/science.142251

AUTHOR CONTRIBUTIONS

LZ and SL constructed the plasmids, cultured bacteria, and analyzed the data. NL and S-LR analyzed the data. H-JC designed and supervised the study. LZ, SL, JC, JW, DY, and H-JC wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDBING

This work was supported by grants from the National Natural Science Foundation of China (NSFC; grant numbers 31770119, 31400068, and 32070098). The funding organization did not influence the design of the experiment or analysis, and interpretation of data, or preparing the manuscript.

ACKNOWLEDGMENTS

The authors thank Zhen-Gang Han and Lei Lei for providing technical assistance during protein analysis.
Iwagami, S. G., Yang, K. Q., and Davies, J. (2000). Characterization of the protocatechuate acid catalytic gene from Streptomyces sp strain 2065. *Appl. Environ. Microbiol.* 66, 1499–1508. doi: 10.1128/Aem.66.4.1499-1508.2000

Jojola, A. F., Geraldes, D. C., Tundis, L. L., Feitosa, V. D. A., Breyer, C. A., Cardoso, S. L., et al. (2016). Biopharmaceuticals from microorganisms: from production to purification. *Braz. J. Microbiol.* 47, 51–63. doi: 10.1016/j.bjm.2016.10.007

Kapust, R. B., and Vaugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668–1674. doi: 10.1002/ps.8.8.1668

Karpeisky, M., SENCHenko, V. N., Dianova, M. V., and Kanevsky, V. (1994). Formation and properties of S-protein complex with S-peptide-containing fusion protein. *FEBS Lett.* 339, 209–212. doi: 10.1016/0014-5793(94)80417-6

Kato, A., Maki, K., Ebina, T., Kuwajima, K., Soda, K., and Kuroda, Y. (2007). Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers* 85, 12–18. doi: 10.1002/bip.20596

Kim, M. R., and Pack, S. P. (2020). Fusion tags to enhance heterologous protein expression. *Appl. Microbiol. Biotechnol.* 104, 2411–2425. doi: 10.1007/s00253-020-10428-4

Kipling, J. S., and Raines, R. T. (1993). Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci.* 2, 348–356. doi: 10.1002/pro.5560020307

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533. doi: 10.1007/s00253-002-1158-6

Varshavsky, A. (2005). "Ubiquitin fusion technique and related methods," in Methods Enzymol., ed. R. J. Deshaies (Academic Press), 777–799.

Wang, X., Zhou, B., Hu, W., Zhao, Q., and Lin, Z. (2015). Formation of active inclusion bodies induced by hydrophobic self-assembling peptide GFP8. *Microb. Cell Factories* 14, 88. doi: 10.1186/s12934-015-0270-0

Wells, T., and Raganaskas, A. J. (2012). Biotechnological opportunities with the beta-ketoadipate pathway. *Trends Biotechnol.* 30, 627–637. doi: 10.1016/j.tibtech.2012.09.008

Zhang, Y. B., Howitt, J., McCorckle, S., Lawrence, P., Springer, K., and Freimuth, P. (2004). Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* 36, 207–216. doi: 10.1016/j.ypep.2004.04.020

Zhao, W., Liu, L., Du, G., and Liu, S. (2018). A multifunctional tag with the ability to benefit the expression, purification, thermostability and activity of recombinant proteins. *J. Biotechnol.* 283, 1–10. doi: 10.1016/j.jbiotec.2018.07.005

Zhao, W. X., Liu, S., Du, G. C., and Zhou, J. W. (2019). An efficient expression tag library based on self-assembling amphipathic peptides. *Microb. Cell Factories* 18:91. doi: 10.1186/s12934-019-1142-9

Zhou, B. H., Xing, L., Wu, W., Zhong, X. E., and Lin, Z. L. (2012). Small surfactant-like peptides can drive soluble proteins into active aggregates. *Microb. Cell Factories* 11. doi: 10.1186/1475-2859-11-10

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zou, Li, Li, Ruan, Chen, Wu, Yan and Chao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.