Research Article

Influence of N-terminal His-tags on the production of recombinant proteins in the cytoplasm of *Bacillus subtilis*

Ngan Thi Phuong Le\(^a\), Trang Thi Phuong Phan\(^a,b\), Hanh Thi Thu Phan\(^a,b\), Tuom Thi Tinh Truong\(^a,b,d\), Wolfgang Schumann\(^a,b\), Hoang Duc Nguyen\(^a,b,*\)

\(^a\) Center for Bioscience and Biotechnology, University of Science, Ho Chi Minh City, Vietnam  
\(^b\) Vietnam National University, Ho Chi Minh City, Vietnam  
\(^c\) Laboratory of Molecular Biotechnology, University of Science, Ho Chi Minh City, Vietnam  
\(^d\) Cancer research laboratory, University of Science, Ho Chi Minh City, Vietnam

**Keywords:** His-tag  
Fusion tag  
Pgrac212 promoter  
*Bacillus subtilis*  
Low expression gene

**Abstract**

The influence of fusion tags to produce recombinant proteins in the cytoplasm of *Bacillus subtilis* is not well-studied as in *E. coli*. This study aimed to investigate the influence of His-tags with different codons on the protein production levels of the high expression gene (gfp+) and low expression gene (egfp) in the cytoplasm of *B. subtilis* cells. We used three different N-terminal His-tags, M-6xHis, MRGS-8xHis and MEA-8xHis, to investigate their effects on the production levels of GFP variants under the control of the Pgrac212 in *B. subtilis*. The fusions of His-tags with GFP + caused a reduction compared to the construct without His-tag. When three His-tags fused with egfp, the EGFp production levels were significantly increased up to 3.5-, 12-, and 15-fold. This study suggested that His-tag at the N-terminus could enhance the protein production for the low expression gene and reduce that of the high expression gene in *B. subtilis*.

1. Introduction

*B. subtilis* offers many advantages for use in the production of recombinant proteins. Unlike the Gram-negative *E. coli*, *B. subtilis* is generally recognized as safe (GRAS) because it is non-pathogenic and nontoxic [1,2]. In addition, *B. subtilis* is well known for its high capacity to secrete proteins [3]. Many modifications of the host strain improved the secretion of recombinant protein production [4]. The expression systems also developed in *B. subtilis* [5]. Most studies on the intracellular production of recombinant proteins have focused on *B. subtilis* [6] using promoter elements optimized for strong expression [7]. Vectors based on Pgrac100 were developed that showed high protein production levels in *B. subtilis* and decreased basal expression in *E. coli* [8]. Modifications of IPTG-inducible promoters have resulted in inducer-free expression plasmids [9], while some reporter genes have been successfully expressed in the cytoplasm of *B. subtilis* like BgaB [7–9], GFP [7,9], and GUS [10]. An optimal combination of strong promoters, transcription terminators, and various translation/secretion signals can achieve a high protein expression level [11]. In addition, the fusion tag also significantly affects the production levels of intracellular proteins.

Fusion tags are short peptides or proteins attached to the N- or C-terminus of the target protein to enhance production, increase solubility and facilitate its purification [12]. Fusion tags include short sequences such as polyhistidine (poly-His), polyarginine (poly-Arg), FLAG, c-myc, or Strep-tag, and proteins such as maltose-binding protein (MBP), glutathione S-transferase (GST), N-utilization substance A (NusA), thioredoxin (Trx), small ubiquitin-related modifier (SUMO) [13,14]. The larger tags are often used to increase the solubility of target proteins [15]. It is known that the fusion tag has an effect on protein production in *E. coli*. The position, sequence, and length of the fusion tag can affect protein production on several levels, including production levels, solubility, binding to the immobilized metal affinity chromatography (IMAC) ligand, tertiary structure, propensity to form crystals, and activity [16,17]. The nucleotide sequence around the translation initiation region (TIR) also substantially impacts translation efficiency in prokaryotes [13]. However, a system that optimizes both production and purification of target proteins in *B. subtilis* has not been developed yet.

A His-tag containing 2–10 histidine residues (commonly 6xHis) is the fusion tag most frequently used because of its small size (about 0.84 kDa for 6xHis). In *E. coli*, a His-tagged recombinant protein can be purified.

---

\(^*\) Corresponding author at: Center for Bioscience and Biotechnology, University of Science, 227 Nguyen Van Cu, Ward 4, District 5, Ho Chi Minh City, Vietnam.  
E-mail addresses: nguyen_hoang.xuatban@hotmail.com, ndhoang@hcmus.edu.vn (H.D. Nguyen).  
https://doi.org/10.1016/j.btre.2022.e00754  
Received 19 February 2022; Received in revised form 28 June 2022; Accepted 14 July 2022  
Available online 19 July 2022  
2215-017X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
by IMAC under denaturing conditions and refolded if it is insoluble [18]. His-tagging offers the advantage of a simple, one-step purification using inexpensive materials. However, some studies have shown that His-tags are related to insolubility and can affect the structure and function of the target protein [19]. An earlier investigation reported that an N-terminal polyhistidine tag can influence the thermal stability of the recombinant protein [20]. Our previous work on His-tagged BgaB and GFP proteins confirmed that production levels of these recombinant proteins were significantly reduced when they were fused with N-terminal histidine residues [5,21,22]. Therefore, selecting a His-tag with an appropriate sequence for protein production is still a major challenge.

This study aimed to provide primary data to understand the influence of His-tags on the protein production levels of the high and low expression genes in B. subtilis. We evaluated the effectiveness of existing and predicted His-tag DNA sequences on the cytoplasmic production of recombinant proteins under the control of the strong promoter, Pgrca212 [23,24], in B. subtilis. Target proteins included high-expression gene sequences optimized for the expression in B. subtilis, GFP+ [25], and a low-expression gene sequence coding for the EGFP with codons optimized for mammalian cells [26]. We compared production levels of the target proteins fused with different His-tag sequences and evaluated their purity after single-step purification on a spin column.

2. Materials and methods

2.1. Bacterial strains, plasmids, oligonucleotides, and growth conditions

The plasmids and oligonucleotides used in this study are shown in Table 1 and Table 2. The E. coli strain OmniMAX from Invitrogen was used for the cloning experiments, and B. subtilis 1012 [27] obtained from MoBiTec was used for the production of proteins. Cells were grown in Luria broth (LB) at 37 °C with shaking. The antibiotics, ampicillin at 0.1 g l⁻¹ for E. coli and chloramphenicol at 0.01 g l⁻¹ for B. subtilis, were added to the culture media.

| Table 1 |
| Plasmid | Description | Source/ reference |
|---------|-------------|------------------|
| pHT100 | Pgrca100-bgaB | [7] |
| pHT100-gfp+ | Pgrca100-gfp+ | [8] |
| pHT105 | Pgrca212-egfp | This study |
| pHT106 | Pgrca212-MEA-8xHis-egfp | This study |
| pHT106 | Pgrca212-gfp+ | From lab collection |
| pHT1070 | Containing egfp gene originated from pEGFP-N1 of Clontech; using as a template to amplify egfp gene | This study |
| pHT116 | Pgrca100-MEA-8xHis-gfp | [8] |
| pHT117 | Pgrca100-MEA-8xHis-bgaB | [8] |
| pHT1222 | Pgrca212-MCS-gfp+|BamHI (high copy number); Amplifying gfp+ gene | This study |
| pHT1259 | Pgrca212-His-thrombin-MCS-Strep (high copy); used to construct pHT1262, pHT2472 | This study |
| pHT1262 | Pgrca212-M-6xHis-egfp | This study |
| pHT1266 | Pgrca212-MGRS-8xHis-MCS-Strep; used to construct pHT2466 | This study |
| pHT1611 | Pgrca212-MGRS-8xHis-bgaB | This study |
| pHT212 | Pgrca212-bgaB | [23] |
| pHT2466 | Pgrca212-MGRS-8xHis-egfp | This study |
| pHT2472 | Pgrca212-M-6xHis-gfp+ | This study |
| pHT2473 | Pgrca212-MEA-8xHis-gfp+ | This study |
| pHT2474 | Pgrca212-MGRS-8xHis-gfp+ | This study |
| pHT259 | Pgrca212-MEA-8xHis-MCS, originated from pHT1262; used to construct pHT1026, pHT2473 | This study |
| pHT261 | Pgrca212-MCS-Strep-Tag, originated from pHT212; used to construct pHT1025 | This study |

2.2. Construction of recombinant plasmids

To investigate the influence of the different His-tags on protein production, target proteins were selected, and expression plasmids were designed with an His-tag. For pHT1611, the His-tag sequence was generated by hybridization of the complementary ON1989 ON1990 oligonucleotides. The hybridization product was cut with BamHI and cloned into plasmid pH212 at the BamHI site. In constructing pHT1025, pHT1026 and pH2466, the egfp gene was first amplified from the plasmid pHT1070 by polymerase chain reaction (PCR) using the primers ON549 and ON632. The PCR products were cut with BamHI/Smal and cloned into the plasmids, pH261, pH259, and pHT1266 at the BamHI and Smal sites. To construct pHT1262, the egfp gene was amplified using the primer pairs ON227 and ON228, with pHT1070 as a template, and the PCR products were introduced into pHT259 at the BamHI and Xbal sites. For pHT2472, pHT2473, and pHT2474, the gfp+ gene was amplified using the primer pairs ON1277 and ON742, with pHT1222 as a template. The PCR products were cut with BamHI/Smal and cloned into pH1259, pH259, and pHT1266 at the BamHI and Smal sites.

2.3. The production of recombinant proteins

The plasmids were introduced into B. subtilis 1012 by natural transformation [28]. Recombinant B. subtilis strains were streaked onto LB agar containing chloramphenicol and incubated overnight at 37 °C. A single colony was inoculated into 10 ml LB medium with chloramphenicol and shaken overnight at 37 °C. Cultures of each strain were replicated using two separate colonies. An appropriate volume of an overnight culture of each clone was transferred to 30 ml LB containing chloramphenicol to give an optical density at 600 nm (OD600) of 0.1 and incubated under shaking at 37 °C. When the OD600 of the culture reached 0.8–1, the cells were divided into three sub-cultures, and two were induced by the addition of IPTG to a final concentration of 0.1 and 1 mM, respectively. Cells were collected at 0 h just before induction and at 2 and 4 h after induction by centrifugation at 6000 g for 10 min at 4 °C. The number of B. subtilis cells sampled was equivalent to those present in 1 ml of culture with an OD600 of 2.4. Samples were prepared for target protein measurement by fluorescence and SDS-PAGE analysis.

2.4. Measurement of green fluorescent protein production in B. subtilis

Cells were lysed by incubation in 500 μl PBS containing 0.2 g l⁻¹ lysozyme at 37 °C for 30 min. Lysates were centrifuged at 12,000 g for 5 min and protein fluorescence was measured in a black microtiter plate (Nunc) using a microplate reader with an excitation wavelength of 465 (+/-8) nm and an emission wavelength of 510 (+/-8) nm for EGFP and excitation at 470 (+/-8) nm and emission at 515 (+/-8) nm for GFP+. The experiments were carried out with two different colonies, and standard errors were calculated.

2.5. SDS-PAGE and western immunoblotting

For SDS-PAGE analysis, cells were grown as described above for 4 h in the presence of different concentrations of IPTG (0, 0.1 or 1 mM), and samples taken at an OD600 of 2.4 were collected by centrifugation at 6000 g for 10 min at 4 °C. The pellets were suspended in 100 μl of the lysis PBS buffer, containing 1.25 g l⁻¹ lysozyme, and incubated at 37 °C for 5 min, after which 25 μl of 5X sample loading buffer was added to each lysate. After heating at 95 °C for 5 min, the samples were centrifuged at 12,000 g for 5 min and supernatants were loaded into an SDS-PAGE gel (12% polyacrylamide). The separated proteins were transferred from the gel to a nitrocellulose membrane, blocked with 5% skim milk in PBS-T (PBS with 0.1% Tween 20) for 1 h, then incubated with primary mouse anti-GFP serum at the dilution 1:10,000 at room temperature for 1 h. After washing with PBS-T, the nitrocellulose membrane
was incubated with HRP-conjugated anti-mouse IgG secondary antibody
for 1 hour at room temperature and washed with PBS-T. For detecting
the EGFP protein, the nitrocellulose membrane was incubated with TMB
(tetramethylbenzidine), according to the manufacturer’s procedure and
the protein bands were imaged with a scanner. The density of the target
protein was determined using AlphaEase FC 4.0 software and the rela-
tive target protein production was calculated by densitometry.

2.6. Purification of recombinant proteins using Ni-NTA spin columns

*B. subtilis* 1012 carrying different plasmids was grown in LB to mid-
log phase, and production of the recombinant proteins was induced by
the addition of 1 mM IPTG. The cells were collected by centrifugation,
resuspended in lysis buffer with 0.02 g L⁻¹ lysozyme, and disrupted by
sonication. The protocol with recommended buffers for Ni-NTA spin
column (Qiagen) was followed, with washing buffer containing 20 mM
imidazole and elution buffer containing 500 mM imidazole.

3. Results

3.1. Does an N-terminal His-tag reduce the production of BgaB and
GFP+?

In the previous study, a His-tag at the N-terminus (P*_{grac100*-his-
BgaB)* drastically reduced the production of BgaB as compared to
P*_{grac100-BgaB) without a His-tag [8]. We used the P*_{grac212 promoter,
which is stronger than P*_{grac100, to evaluate the level of the fusion
protein His-BgaB, and found that the protein production from P*_{grac212-
his-BgaB (pHT1611) was equivalent to that of the P*_{grac100-his-
BgaB construct (pHT1178) (Fig. 1a). In the absence of a His-tag, the
stronger promoter P*_{grac212 in pHT212 produced a higher BgaB pro-
duction level than pHT100. However, fusion with an N-terminal His-tag
significantly reduced BgaB production by up to 60%, confirming the
negative effect of His-tag fusion at the N-terminus on production of the
BgaB protein.

Like BgaB, GFP+ can be highly overexpressed in *B. subtilis*. In the
previous study, the N-terminal His-tag also decreased the production of
the GFP+ protein as compared to the C-terminal His-tag [8]. The

### Table 2
The oligonucleotides used in this study.

| Oligonucleotide | Sequence, 5’→3’ | Used for |
|-----------------|-----------------|---------|
| ON1277          | AAAGGAGGAAGGATCCATGGCTAGCAAAGGAGAAGAACT | Amplifying gfp+ gene; constructing pHT2472, pHT2473, pHT2474 |
| ON742           | TAGGGCGGCGCGCCGGGTATTGGTAGAGCTCACTGCGGAGCTG | Amplifying gfp+ gene; constructing pHT2472, pHT2473, pHT2474; colony PCR pHT2472, pHT2473, pHT2474 |
| ON1359          | AGTGTAATCTTTAGGAGGTCT | Colony PCR pHT2472, pHT2473, pHT2474, pHT2466 |
| ON1375          | GCTTCAAGCATTGCTCCAGTAAG | Sequencing pHT2472, pHT2473, pHT2474, pHT2466 |
| ON549           | GTATCTCGGAAGCTCATGGACAGGAGGAGGAGCTG | Amplifying gfp+ gene; constructing pHT1025, pHT1026, pHT2466; colony PCR pHT1025, pHT1026 |
| ON632           | TAGGGCGGCGCGCCGGGGAGCT | Amplifying gfp+ gene; constructing pHT1025, pHT1026, pHT2466; colony PCR pHT2466 |
| ON227           | GGTCGGACCGGAGGTCTGGAGCAAAGGAGGAGGAGCTG | Amplifying gfp+ gene to construct pHT1262 |
| ON228           | CGCAGTGCAGCCTAGAGATCCGCGGCGGCTCAAG | Amplifying gfp+ gene; constructing pHT1262; colony PCR pHT1262 |
| ON653           | ACGGATAGTTCTGTGACAGCTATGG | Sequencing pHT1025 and pHT1026; colony PCR pHT1262 |
| ON314           | TGGTTCAACATTGTTCCAGGT | Sequencing pHT1025 and pHT1026; colony PCR pHT1025, pHT1026, pHT1262 |

![Fig. 1. Production of BgaB and GFP+ proteins with and without N-terminal His-tag in B. subtilis. (a) pHT100 (P*_{grac100-BgaB); pHT1178 (P*_{grac100-His-BgaB), pHT212 (P*_{grac212-BgaB) and pHT1611 (P*_{grac212-His-BgaB); (b) pHT100 gfp (P*_{grac100-gfp); pHT1169 (P*_{grac100-MEA-8xHis-gfp). Samples were harvested at 4 h after induction with IPTG at 0 mM (-) and 1 mM (+).](image-url)
production in B. subtilis of the GFP+ protein with N-terminal His-tag under the control of the Pgrac100 promoter (pHT1169) was dramatically reduced as compared to pHT100-gfp without a His-tag, and the target protein band was hard to be seen on an SDS-PAGE gel (Fig. 1b). The decrease in the GFP+ level when fused with a His-tag at the N-terminus was similar to that of the BgaB protein.

3.2. Evaluation of the effect of different His-tag sequences on the production of the GFP+ protein

The translation efficiency of prokaryotes is influenced by the nucleotide sequence around the TIR [13]. In this study, we tested three different His-tag sequences, M-6xHis, MRGS-8xHis and MEA-8xHis (Table 3) and compared their effects on GFP+ production under the control of Pgrac212. We generated three plasmids, pHT2472 (Pgrac212-M-6xHis-gfp+), pHT2473 (Pgrac212-MEA-8xHis-gfp+) and pHT2474 (Pgrac212-MRGS-8xHis-gfp+) and transformed them into competent B. subtilis cells. Quantification of His-GFP+ production was carried out by SDS-PAGE (Fig. 2a) and western blot (Fig. 2b). These results confirmed that all three His-tag sequences significantly reduced GFP+ production by about 73% as compared to the control without a His-tag. Analysis of the target protein band density compared to total cellular proteins revealed that MRGS-8xHis-gfp+ (pHT2474) had the highest production level (about 9.4% of the total cellular proteins) and was 2.5 times as high as M-6xHis-gfp+ (pHT2472).

In Fig. 2c, GFP+ fluorescence increased with increasing IPTG concentrations from 0 mM to 1 mM and increasing induction time from 0 h to 4 h. Comparing the green fluorescence intensity level of GFP+ proteins fused with different His-tags at 4 h and 1 mM IPTG, pHT2474 (MRGS-8xHis) showed the highest level and is 3.5 times as high as the amount of pHT2472 (M-6xHis). Among three His-tagged sequences (Table 3) fused to the GFP+ N-terminus, the MRGS-8xHis sequence yielded the highest GFP+ production level. Comparing fluorescence of GFP+ proteins fused with different His-tags at 4 h and 1 mM IPTG, pHT2472 (M-6xHis), pHT2473 (MEA-8xHis), pHT2474 (MRGS-8xHis) showed the green fluorescence intensity lower than pHT1066 (without His-tag) 35, 14 and 10 times, respectively.

3.3. Enhancing the expression of EGFP by fusion with different His-tags

egfp is a gfp variant that has been codon-optimized for high expression level in mammalian cells [26]. In this study, we designed plasmids containing the egfp gene fused to the three His-tag sequences as shown above (Table 3) and evaluated their effect on EGFP protein production. We generated four plasmids: pHT1025 (Pgrac212-egfp), pHT1026 (Pgrac212-M-6xHis-egfp), pHT1026 (Pgrac212-MEA-8xHis-egfp) and pHT2466 (Pgrac212-MRGS-8xHis-egfp), which were then transformed into B. subtilis and expressed. The His-EGFP protein was detected by SDS-PAGE (Fig. 3a) and Western blot (Fig. 3b). The production of the EGFP protein without His-tag after incubation with 1 mM IPTG was low, but it was seen by Western immunoblotting. All three His-tags, M-6xHis, MEA-8xHis, and MRGS-8xHis, significantly increased production of EGFP as compared to the control without His-tag (pHT1025).

In Fig. 3c, EGFP fluorescence increased with His-tag fusion and with increasing IPTG concentration from 0 mM to 1 mM and induction time from 0 h to 4 h. Comparing the fluorescence of EGFP proteins fused with different His-tags at 4 h and 1 mM IPTG, pHT1026 (M-6xHis) showed the highest green fluorescence intensity and was 15 times higher than pHT1025 without His-tag.

Each of the three His-tags increased the EGFP production to a level detectable on a stained SDS-PAGE gel. However, the production and fluorescence levels of EGFP differed with the different His-tags. The M-6xHis sequence resulted in the highest EGFP production.

3.4. Effect of adding histidine to the culture medium on the production of N-terminal His-tagged proteins

The lower production level of the BgaB and GFP+ proteins with an N-terminal poly-His-tag could be due to a deficiency in His-at the initial phase of elongation, which would reduce the speed and efficiency of translation through diminished loading of tRNAHis, limiting the transport of His-to the ribosome. The decreased supply of histidine could result in lower production of the target protein. To see if this reduction in production was due to histidine deficiency during poly-His-synthesis, we added histidine to the culture medium (Fig. 4a). The results in Fig. 4a showed that when histidine (+) was added to the culture medium and transcription induced by 1 mM IPTG, the production of His-BgaB was higher than in the absence of histidine (-). Analysis of the target protein bands using AlphaEase 4.0 software, showed that adding histidine to the culture medium increased the production level of the target protein by 32% after 4 h of IPTG induction as compared to the no-His-control. Thus, the decrease in BgaB protein with His-tag at the N-terminus was partly due to the lack of histidine during the initial phase of elongation. The experiment was repeated with the His-GFP+ and His-EGFP compared to the controls without His-tag (Fig. 4b). The fluorescence intensity of His-GFP+ and His-EGFP after 4 h of 1 mM IPTG induction was measured, and the results are shown in Fig. 4c. The gfp+ and egfp without His-tag showed no difference in expression level on SDS-PAGE with (+) or without (-) histidine addition to the culture medium. Similar to His-BgaB, His-GFP+ increased the expression level and fluorescence intensity when histidine was added. In contrast, for the low-expression gene in B. subtilis, His-tag fusion at the N-terminus increased the expression level of the EGFP protein, and the addition of histidine to the culture medium did not affect His-EGFP expression levels.

3.5. Purification of recombinant proteins

To determine whether His-tagged recombinant proteins produced in B. subtilis can be purified via one-step purification, we used Ni-NTA spin columns and chose B. subtilis strains containing plasmids pHT1025 (His-EGFP), pHT2473 (His-GFP+), and pHT1611 (His-BgaB). These strains were cultured for 4 h with 1 mM IPTG and cells were harvested by centrifugation at 6000 g for 10 min at 4 °C. The before-column, after-column, washes and eluted fractions were analyzed by SDS-PAGE (Fig. 5) and densitometry was done on the target protein bands in the eluted fraction using the AlphaEase 4.0 software. The purity of His-EGFP, His-GFP+, and His-BgaB recombinant proteins was greater than 88%. The result proved that proteins fused with His-tag expressed in B. subtilis could be easily purified via a single step.

4. Discussion

His-tag fusions with the target proteins were used to facilitate

Table 3

| Hist-Tag | DNA Sequence | Peptide Sequence | Length (aa) | Molecular Weight (Da) |
|----------|--------------|-----------------|-------------|----------------------|
| M-6xHis  | atgcacattcatcatcatctcttggctgtgcacacgcggatce | MBHHHHSSSLVPRGS 16 | 1811.85 |
| MRGS-8xHis | ataggggggaagtacatcatcatcatcatatccagcggt | MRSHHHHHBBBBSGS 14 | 1689.73 |
| MEA-8xHis | atgagggcatctcatcatcatatccacagcggt | MEAHHHHHHBBHSGS 13 | 1589.65 |

* Determined by using the PepDraw tool (www.pepdraw.com).
successfulness of producing recombinant proteins with high expression genes fused with His-tag in B. subtilis [29]. These examples are a few pieces of evidence for the proteins with C-terminal His-tag were at higher levels than that of proteins purified, but they could interfere with the protein production levels. Fusion tags can be used at the N-terminus or the C-terminus of recombinant proteins. Our previous reports showed that GFP and BgaB proteins with C-terminal His-tag were at higher levels than that of proteins with N-terminal His-tag in B. subtilis [8,21,22]. Another paper indicated that His-tag position at C-terminus could increase the yield and activity of CotA protein [29]. These examples are a few pieces of evidence for the successfulness of producing recombinant proteins with high expression genes fused with His-tag in B. subtilis. Surprisingly, these His-tag fusions at the N-terminus reduced the protein production levels compared with the His-tag fusion at the C-terminus. We asked what happened if His-tag was fused at the N-terminus with the low expression gene in comparison with the high expression gene. This study showed that His-tags fused at the N-terminus reduced the protein production levels of high expression genes (bgaB and gfp+) while enhancing the protein production levels of the low expression gene (egfp) in B. subtilis.

The results in this study confirmed a previous report that the DNA sequence at the N terminus affected transcription and initiation of translation [8]. The TIR sequence promotes the interaction with RNA that initiates translation [30]. Prokaryotic translation efficiency is most affected by the folding free energy of the region between nucleotides -10 and +35. Therefore, to achieve high-production protein levels, the nucleotide sequence around the TIR [31] needs to be optimized. We designed His-tags attached to the N-terminus of target proteins and then evaluated the effect of DNA encoding the His-tags with existing and predicted sequences on the production of recombinant proteins. The gfp+ has codons optimized for expression in prokaryotic cells and can be used for high production in B. subtilis. Fusing three different sequences of multiple codons for repeated histidine residues at the N-terminus of the high-production GFP protein resulted in reduced GFP+ production.

Since the N-terminal coding sequences influences the efficiency of ribosome binding to the mRNA and ribosome extension at the initial stage of translation, it strongly affects gene expression at the level of translation [32]. At the initial phase of elongation, the synthesis of poly-histidine affects the efficiency of translation. Adding histidine to the culture medium increased production of BgaB and GFP+ with N-terminal His-tags. Therefore, the decrease in production of the high-expression genes is partly due to the lack of histidine during the initial phase of elongation. Increasing the number of histidine-coding codons affects the efficiency of translation [8]. The TIR sequence promotes the interaction with rRNA that initiates translation [30]. Prokaryotic translation efficiency is most affected by the folding free energy of the region between nucleotides -10 and +35. Therefore, to achieve high-production protein levels, the nucleotide sequence around the TIR [31] needs to be optimized. We designed His-tags attached to the N-terminus of target proteins and then evaluated the effect of DNA encoding the His-tags with existing and predicted sequences on the production of recombinant proteins. The gfp+ has codons optimized for expression in prokaryotic cells and can be used for high production in B. subtilis. Fusing three different sequences of multiple codons for repeated histidine residues at the N-terminus of the high-production GFP+ protein resulted in reduced GFP+ production.

Fig. 2. Production of GFP+ protein fused with different His-tags in B. subtilis: pHT2472 (Pgrac212-M-6xHis-gfp+); pHT2473 (Pgrac212-MEA-8xHis-gfp+) and pHT2474 (Pgrac212-MRGS-8xHis-gfp+). Samples were harvested at 4 h after induction with IPTG at 0 mM (-) and 1 mM (+). (a) SDS-PAGE; (b) Western blot. (c) GFP+ fluorescence intensity. Samples were harvested at 0 h (when the OD_600 reached 0.8–1), 2 h and 4 h after induction with IPTG at 0 mM, 0.1 mM and 1 mM.

Fig. 3. Production of EGFP protein fused with different His-tags in B. subtilis: pHT1026 (Pgrac212-MEA-8xHis-egfp+); pHT1062 (Pgrac212-His-egfp+) and pHT2466 (Pgrac212-MRGS-8xHis-egfp+). Samples were harvested at 0 h (when the OD_600 reached 0.8–1) and 2 h and 4 h after induction with IPTG at 0 mM (-) and 1 mM (+). (a) SDS-PAGE; (b) Western blot. (c) EGFP fluorescence intensity. Samples were harvested at 0 h (when the OD_600 reached 0.8–1), 2 h and 4 h after induction with IPTG at 0 mM, 0.1 mM and 1 mM.
production level of EGFP in *B. subtilis* was low. This study clearly indicated that His-tag could enhance the production of the recombinant protein for the low expression gene such as *egfp* in *B. subtilis*. In the control (pHT1025) carrying the *egfp* gene, the target protein band could not be detected by SDS-PAGE but was seen by Western immunoblotting. However, the EGFP fused with the His-tag at the N-terminus increased production of this protein significantly up to fifteen times higher than the control without His-tag. The most exciting experimental design for the study is using two *gfp* genes with different sequences, *gfp*+, optimized for bacteria with CAI 0.759 for *B. subtilis* and *egfp*, optimized for mammalian with CAI 0.637 for *B. subtilis*. The distinction in the gene optimization leads to higher GFP+ production levels in *B. subtilis* than EGFP. Based on the effect of His-tag in the N-terminus, we found that there is a difference between these two representatives. His-tags at the N-terminus reduced the protein synthesis of the highly expressed protein (GFP+) and increased the protein production of the low expression protein (EGFP) in *B. subtilis* but not in *E. coli* (data not shown).

The analysis results of codon adaptation index (CAI) *B. subtilis* by using the “CAI calculator” showed in Table 4. The CAI index are about 0.024 apart between the whole sequences of *egfp* gene and *egfp* fusion with His-tags. However, the CAI index of the first 10 codons has a big difference of about 0.199 between these genes. It is likely that the *B. subtilis* CAI index increased with His-tag sequence, and the higher the CAI for the first 10 codons, the greater the EGFP synthesis in *B. subtilis*. The codon choice could affect protein synthesis in *B. subtilis* and increase the production of the low expression protein such as EGFP.

The translation process is influenced by the secondary structure of the mRNA [34], and some studies have shown that secondary structure in the 5′ untranslated region (5′ UTR) generally reduces translation initiation efficiency and overall protein production [35]. The more negative the secondary structure energy, the greater the formation of the secondary structure. This affects the ribosomes’ binding to and moving on the mRNA during translation. We analyzed the secondary structure (SS) energy at the first ten codons of mRNA in these sequences by using the RNAfold tool, and the results are shown in Table 4. The secondary structure could affect protein synthesis in *B. subtilis* and increase the production of the low expression protein such as EGFP.

**Table 4**

| Plasmid Characteristic | CAI of the whole fusion gene | CAI of the first 10 codons | SS (RNAfold for the first 10 codons) (J/mol) | Expression level |
|------------------------|-------------------------------|---------------------------|-------------------------------------------|-----------------|
| pHT1025 EGFP           | 0.637                         | 0.638                     | -11.5                                     | Control of EGFP |
| pHT1026 MEA-8xHis-EGFP  | 0.641                         | 0.771                     | -0.7                                      | 12-fold increase|
| pHT1262 M-6xHis-EGFP   | 0.661                         | 0.837                     | -0.6                                      | 15-fold increase|
| pHT2466 MRGS-8xHis-EGFP| 0.64                          | 0.699                     | -3.3                                      | 3.5-fold increase|
| pHT1066 GFP+           | 0.759                         | 0.81                      | -1.5                                      | Control of GFP+ |

* Determined by using the CAI calculator and 10-codon RNAfold tool.
structural energies were proportional to the production levels of EGFP proteins fused with different His-tag sequences. These results showed that His-tag affects the expression of low expression genes in B. subtilis and tends to be optimal for expression in B. subtilis.

N. T. P. Le et al.

[3] J. Neef, J. M. van Dijl, G. Buist, Recombinant protein secretion by Bacillus subtilis. Biochem. 65 (2021) 187–195, https://doi.org/10.1016/j.biochem.2021.07.005.

[4] K. Zhang, V. Guiru, S. Zhu, A. W. G. Schurmann, Novel plasmid-based fluorescent protein allows sensitive detection of gene transfer in mammalian cells, Biochem. Biophys. Res. Commun. 227 (1996) 707–711, https://doi.org/10.1016/0006-291X(96)00100-x.

[5] H. Saito, T. Shibata, T. Ando, Mapping of genes determining nonpermissiveness and host-specific restriction bacteriophages in Bacillus subtilis Marburg, Mol. Gen. Genet. MGG 170 (1979) 117–122.

[6] T. Phan, P. Huynh, T. Truong, H. Nguyen, A generic protocol for intracellular expression of recombinant proteins in Bacillus subtilis, in: N.A. Burgess-Brown (Ed.), Heterologous Gene Expression in E. Coli: Methods and Protocols, Methods in Molecular Biology, Springer, New York, NY, 2017, pp. 325–334, https://doi.org/10.1007/978-1-4939-6887-9_21.

[7] L. Li, T. Xie, Z. Liu, H. Feng, G. Wang, Activity enhancement of Gta laccase by hydrophobic engineering, histidine tag optimization and static protein, Protein Eng. Des. Sel. 31 (2018) 1–5, https://doi.org/10.1016/j.proeng.2016.12.006.

[8] G. S. Allen, A. Zavialov, R. Gursky, M. Ehrenberg, J. Frank, The Cyto-EM structure of a translation initiation complex from Escherichia coli, Cell 121 (2005) 703–712, https://doi.org/10.1016/j.cell.2005.05.018.

[9] T. K. M. Nguyen, M. R. Ki, R. G. Son, S. P. Pack, The NT11, a novel fusion tag for expression of recombinant proteins in Bacillus subtilis, in: N.A. Burgess-Brown (Ed.), Heterologous Gene Expression in E. Coli: Methods and Protocols, Methods in Molecular Biology, Springer, New York, NY, 2017, pp. 325–334, https://doi.org/10.1007/978-1-4939-6887-9_21.

[10] L. Bivona, Z. Zou, N. Stutzman, P.D. Sun, Influence of the second amino acid on recombinant protein expression, Protein Expr. Purif. 74 (2010) 248–256, https://doi.org/10.1016/j.pep.2010.06.005.

[11] C. Chiaruttini, M. Guillier, M. Lebrasseur, The role of mRNA secondary structure in bacterial translation, wRNA 11 (2020) e1579, https://doi.org/10.1586/wrna.1579.

[12] M. Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynolds, J. Moore, M.J., McFadyen, L.J., 2019. mRNA structure regulates protein expression through changes in small ribonuclease, Proc. Natl. Acad. Sci. 116, 24075–24083, https://doi.org/10.1073/pnas.1908052116.