Journal of Bioscience and Bioengineering
VOL. xx No. xx, 1–7, 2016

www.elsevier.com/locate/jbiosc

N-terminal SKIK peptide tag markedly improves expression of difficult-to-express proteins in *Escherichia coli* and *Saccharomyces cerevisiae*

Teruyo Ojima-Kato,[1,2] Satomi Nagai,[1] and Hideo Nakano[1,*]

Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan[1] *and Knowledge Hub Aichi, Aichi Science and Technology Foundation, Yakusa-cho, Toyota 470-0356, Japan*[2]

Received 22 August 2016; accepted 6 December 2016
Available online xxx

Despite advances in microbial protein expression systems, low production of proteins remains a great concern for some genes. Here we report that the insertion of a short peptide tag, consisting of Ser-Lys-Ile-Lys (SKIK), adjacent to the start codon of genes encoding difficult-to-express proteins can increase protein expression in *Escherichia coli* and *Saccharomyces cerevisiae*. Protein expression levels of a mouse monoclonal antibody (mAb), rabbit mAbs obtained from clonal B cells, and an artificially designed peptide were significantly increased simply by the addition of the SKIK tag in *E. coli* systems. In particular, a ~30-fold increase in protein production was observed for the mouse mAb, and the artificially designed peptide band became detectable in sodium dodecyl sulfate-poly acrylamide gel electrophoresis after coomassie brilliant blue staining or western blotting on adding the SKIK tag. The tag also increased the expression of tagged proteins in *S. cerevisiae* and an *E. coli* cell-free protein synthesis system. Although the mechanism of high protein expression on addition of the tag is unclear, our findings offer great benefits to biotechnology research and industry.

[**Key words:** Difficult-to-express protein; Increase of protein expression; *Escherichia coli*; *Saccharomyces cerevisiae*; N-terminal peptide tag; Cell-free protein synthesis]

Recombinant protein expression systems using microorganisms have become essential tools in industry and biotechnology research. *Escherichia coli* and *Saccharomyces cerevisiae* are the most popular hosts for the production of peptides and proteins (1,2).

Despite great achievements in microbial protein expression, many researchers often still encounter difficulties in obtaining high protein yields. In general, various parameters should be considered to improve recombinant protein expression, such as codon usage (3,4), culture media and cultivation conditions (5,6), use of peptide tags (such as maltose binding protein (7), glutathione S-transferase (8) or a SUMO tag (9,10), which can also improve the solubility of the fused proteins), coexpression of molecular chaperones (11), and the host–vector system used (12). No general rule for improving expression of difficult-to-express proteins has been found. Optimization of these parameters is laborious and time-consuming, yet it often ends with an unsatisfactory protein yield.

Numerous studies have been performed to examine the reasons for the variability in protein expression levels of different genes. The 5′-untranslated region (UTR) is known to affect both transcript and protein expression levels (13). Recent statistical studies have revealed that reduced free energy of the mRNA structure near the start codon correlates with higher protein expression level (14,15). For example, Goodman et al. (15) analyzed RNA, DNA, and protein expression levels of superfolder green fluorescent protein in *E. coli* using over 14,000 combinations of promoters, ribosome binding sites, and 11 N-terminal codons, and found that nucleotide changes in the +10 region from the start codon were most correlated with expression changes. Bivona et al. (16) reported a statistically significant enrichment in the amino acids Ala and Ser at the +2 position among highly expressed *E. coli* proteins. More recently, Boël et al. (17) reported that whole gene codon content directly modulates both translation efficiency and mRNA stability, even though the initial 16 codons are most likely to influence mRNA folding. Despite these investigations using large-scale protein expression data sets, reliable strategies for improving expression levels of difficult-to-express proteins remain elusive.

In this study, we hypothesized that the insertion of a nucleotide sequence encoding a small number of amino acids common to the N-terminus of highly expressed proteins, immediately after the start codon of genes encoding difficult-to-express proteins, would improve protein expression. Based on a comprehensive analysis of amino acid sequences from the second to the fifth residues at the N-terminus of highly expressed proteins in *E. coli* (16), we designed a short peptide tag with the sequence Ser-Lys-Ile-Lys (the SKIK tag). We then demonstrated the effect of this tag on the expression levels of several difficult-to-express proteins in *E. coli* using *in vivo* and *in vitro* expression systems. We also applied the N-terminal SKIK tag to a *S. cerevisiae* expression system. In all cases, we found that insertion of the SKIK tag was effective in increasing production of the difficult-to-express proteins.

* Corresponding author. Tel.: +81 52 789 4144; fax: +81 52 789 4145.
E-mail address: hnakano@agr.nagoya-u.ac.jp (H. Nakano).

1389-1723/$ – see front matter
http://dx.doi.org/10.1016/j.jbiosc.2016.12.004

MATERIALS AND METHODS

Recombinant proteins for expression Several model proteins exhibiting relatively low expression in *E. coli* in our preliminary experiments (using a pET system), were chosen: (i) an anti-*E. coli* O157 fragment of antigen binding (Fab) fused with a leucine zipper (previously named Zipbody), encoded by a gene derived from mouse hybridoma (accession No. LC085624 [heavy chain] and LC085625 [light chain]). The heavy chain (Hc) with a HA tag and the light chain (Lc) with a FLAG tag are located in series between the T7 promoter and terminator in a pET vector (named pET22b-mouse Zipbody in previous work (18)). (ii) A C-terminally His-tagged peptide leucine zipper B (LZB) gene, encoding 31 amino acid residues (AQLKKKLQALKKKNAQLKWKLQALKKKLAQK), in which codon usage was optimized for expression in *E. coli*. This gene was artificially synthesized (Thermo Fisher Scientific, Waltham, MA, USA). (iii) Anti-*Listeria monocytogenes* single chain Fvs (scFvs) No. 1 and No. 4 clones fused with a HA-tag, which were previously constructed from rabbit B cells (accession nos. LC030181, LC030187, LC030182, and LC030188) (19). The codon usages in the mouse and rabbit antibody genes were not modified. All of these genes were inserted into the *Nde*I site of vector pET22b in our previous works (18,19), and the DNA sequences are shown in the Supplementary materials (Fig. S1).

Insertion of tags into the rabbit scFv expression plasmid To evaluate the influence of the tag length, SKIK (5′-TCTAAAATAAAA-3′), SKI (5′-TCTAAAATA-3′), SK (5′-TCTAAA-3′), and S (5′-TCT-3′) tags were inserted at the N-terminus of the rabbit No. 1 scFv-encoding DNA immediately after the start codon. This was achieved in a pET plasmid by the QuikChange site-directed mutagenesis method using primers 1–2 (for SKIK), 3–4 (for SKI), 5–6 (for SK), and 7–8 (for S), with KOD Plus DNA polymerase (Toyobo, Osaka, Japan). Table 1 lists the primers used in this study. The SKIK2 tag, encoded with synonymous codons (5′-TCGAAGATCAAG-3′), and the T7 tag, which is derived from the leader signal of the T7 bacteriophage capsid protein (with a DNA sequence of 5′-ATGGCTAGCATGACTGGTGGACAGCAAATGGGT-3′ and peptide sequence MASMTGGQQMG) were also inserted as above using the following primers: 9–10 (to insert the SKIK2 tag into No. 1 scFv), 11–12 (to insert the T7 tag into No. 1 scFv), and 13–12 (to insert the T7 tag into No. 4 scFv). The products were introduced into *E. coli* DH5α after *Dpn*I treatment (Takara Bio, Otsu, Shiga, Japan). Transformed colonies, growing on a Luria–Bertani (LB)-agar plate containing 100 μg/mL ampicillin, were selected.

Insertion of the SKIK tag in various proteins DNA (5′-TCTAAAATAAAA-3′) encoding the SKIK tag was inserted next to the start codon (ATG) of genes in the above pET plasmids by PCR with primers containing the SKIK coding sequence. For mouse Zipbody, two DNA fragments amplified using the primer pairs 14–15 (for the portion from the Hc to the start site of the Lc), and 16–17 from pET22b-mouse Zipbody, were connected using the Gibson assembly technique (New England Biolabs, Ipswich, MA, USA) to form a plasmid (pET22b-SKIK-mouse Zipbody) in which SKIK tags were inserted directly after the start codons of both the Hc and Lc. For rabbit scFvs and the LZB peptide, the SKIK tag was inserted by QuikChange site-directed mutagenesis using primers 18–19 (for No. 4 scFv) and 20–21 (for LZB).

To construct a cold shock expression plasmid for Zipbody, pColdI (Takara Bio) was first linearized by PCR with primers 22–23, and the Zipbody gene was amplified from the pET22b construct using primers 24–25; these fragments were assembled as described above to form pCold-mouse Zipbody. The SKIK-tagged Zipbody gene amplified from the pET-based plasmid using primers 26–25 was used to form pCold-SKIK-mouse Zipbody. A translation enhancing element included in the original plasmid was eliminated to enable proper evaluation of the influence of the SKIK tag. Unless otherwise stated, KOD Plus DNA polymerase was used for PCRs.

Protein expression in *E. coli* The mAbs (mouse Zipbody and rabbit scFvs) and LZB were expressed in *E. coli* strain Shuffle T7 Express (New England Biolabs) and *E. coli* BL21(DE3) pLysS (Thermo Fisher Scientific), respectively. Expression of protein through the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) was started when cultures reached an OD600 of 0.4–0.5. Cells were then grown aerobically, with shaking at 180 rpm, at 16°C for 24 h or 37°C for 2 h. LB, Terrific Broth (TB; 12 g Bacto tryptone, 24 g Bacto yeast extract, 8 mL glycerol, 9.4 g K_2HPO_4, 2.2 g KH_2PO_4 in 1 L) or MMI medium (12.5 g Bacto tryptone, 25 g Bacto yeast extract, 8.5 g NaCl, 20 mL 1 M Tris–HCl (pH 7.2), and 4 mL glycerol in 1 L) supplemented with 100 μg/mL ampicillin, were used. When using BL21(DE3) pLysS as host, 17 μg/mL of chloramphenicol was also included in the medium.

Cell-free protein synthesis Rabbit No. 1 scFv was used as the model protein because its expression in cell-free protein synthesis (CFPS) was low in our preliminary experiment. DNA fragments containing the T7 promoter and terminator for

TABLE 1. Primers used in this study.

No.	Sequence (5′-3′)	Description
1	AGATATACATATGTCTAAAATAAAAGACCCTATGCTGACC	Insertion of SKIK tag into the rabbit No. 1 scFv
2	TCAGCATAGGGTCTTTTATTTTAGACATATGTATATCTCC	
3	AGATATACATATGTCTAAAATAGACCCTATGCTGACC	Insertion of SKI tag into the rabbit No. 1 scFv
4	TCAGCATAGGGTCTATTTTAGACATATGTATATCTCC	
5	AGATATACATATGTCTAAAGACCCTATGCTGACC	Insertion of SK tag into the rabbit No. 1 scFv
6	TCAGCATAGGGTCTTTAGACATATGTATATCTCC	
7	AGATATACATATGTCTGACCCTATGCTGACC	Insert of S into the rabbit No. 1 scFv
8	TCAGCATAGGGTCAGACATATGTATATCTCC	
9	AGATATACATATGTCGAAGATCAAGGACCCTATGCTGACC	Insertion of SKIK2 tag into the rabbit No. 1 scFv
10	TCAGCATAGGGTCCTTGATCTTCGACATATGTATATCTCC	
11	ATGGCTAGCATGACTGGTGGACAGCAAATGGGTGACCCTATGCTGACCCAGAC	Insertion of T7 tag into the rabbit No. 1 scFv (forward primer)
12	CCACCAGTCATGCTAGCCATCATATGTATATCTCCTTCTTAAAG	Insertion of T7 tag into the rabbit scFvs (reverse primer)
13	ATGGCTAGCATGACTGGTGGACAGCAAATGGGTGATGTCGTGATGACCCAGAC	Insertion of T7 tag into the rabbit No. 4 scFv (forward primer)
14	AGGAGATATACATATGTCTAAAATAAAAGAGGTCCAGCTG	Insertion of SKIK tag into the N-terminal of Hc and Lc genes of mouse Zipbody.
15	GTCATCAAAACATCTTTTATTTTAGACATATGATATCTCCTTCT	
16	TAAAATAAAAGATGTTTTGATGACCCAAAC	
17	TAGACATATGTATATCTCCTTCT	
18	AGATATACATATGTCTAAAATAAAAGATGTCGTGATGACC	Insertion of SKIK tag into the rabbit No. 4 scFv
19	TCATCACGACATCTTTTATTTTAGACATATGTATATCTCC	
20	AGATATACATATGTCTAAAATAAAAGCCCAGCTGAAAAAAAAAAC	Insertion of SKIK tag into LZB
21	TTTCAGCTGGGCTTTTATTTTAGACATATGTATATCTCC	
22	GTAATCTCTGCTTAAAAGCACAGA	Linearization of pColdI vector
23	CATGGTGTATTACCTCTTAATAATT	
24	ATTAAGAGGTAATACACCATGGAGGTCCAGCTGCAACAGTC	Amplification of mouse Zipbody encoding genes from pET22b-mouse Zipbody (forward)
25	TGCTTTTAAGCAGAGATTACTTATTATTTGTCATCGTCGTCCTTG	Amplification of mouse no-tagged and SKIK-tagged Zipbody encoding genes (reverse)
26	ATTAAGAGGTAATACACCATGTCTAAAATAAAAGAGGTCCAGCTG	Amplification of SKIK-tagged mouse Zipbody encoding genes from pET22b-SKIK-mouse Zipbody (forward)
27	ATCTCGATCCCGCGAAATTAATACG	Amplification of DNA templates for CFPS
28	TCCGGATATAGTTCCTCCTTTCAG	
29	TAAGCTTGGTACCGAAAAAAATGGATGTCGTGATGACCCAGAC	Amplification of no-tagged rabbit No. 4 scFv (forward)
30	GGCCCTCTAGGATCAGTTAAGCGTAATCTGGAACATCGTATG	Amplification of rabbit No. 4 scFv (reverse)
31	TAAGCTTGGTACCGAAAAAAATGTCTAAAATAAAAGATGTCGTGATGACC	Amplification of SKIK-tagged rabbit No. 4 scFv (forward)
32	TTTTTTCGGTACCAAGCTTAATATTCCC	Linearization of pYC2/NT vector
33	TAACTGATCCTAGAGGGCCGCATCATG	
34	TCTCCCAAGAGCATCCTTGAAG	Real time PCR
35	TAGTGGATTTTGCATACCAGG	

Sequences corresponding to tags are underlined.

Please cite this article in press as: Ojima-Kato, T., et al., N-terminal SKIK peptide tag markedly improves expression of difficult-to-express proteins in *Escherichia coli* and *Saccharomyces cerevisiae*, J. Biosci. Bioeng., (2016), http://dx.doi.org/10.1016/j.jbiosc.2016.12.004