Mol. Cells *32*, 349-358, October 31, 2011
DOI/10.1007/s10059-011-0088-5

Molecules and Cells

Novel GFP Expression Using a Short N-Terminal Polypeptide through the Defined Twin-Arginine Translocation (Tat) Pathway

Sang Jun Lee*, Yun Hee Han, Young Ok Kim, Bo Hye Nam, and Hee Jeong Kong

Escherichia coli **is frequently used as a convenient host organism for soluble recombinant protein expression. However, additional strategies are needed for proteins with complex folding characteristics. Here, we suggested that the acidic, neutral, and alkaline isoelectric point (pI) range curves correspond to the channels of the *E. coli* type-II cytoplasmic membrane translocation (periplasmic translocation) pathways of twin-arginine translocation (Tat), Yid, and general secretory pathway (Sec), respectively, for unfolded and folded target proteins by examining the characteristic pI values of the N-termini of the signal sequences or the leader sequences, matching with the known diameter of the translocation channels, and analyzing the N-terminal pI value of the signal sequences of the Tat substrates. To confirm these proposed translocation pathways, we investigated the soluble expression of the folded green fluorescent protein (GFP) with short N-terminal polypeptides exhibiting pI and hydrophilicity separately or collectively. This, in turn, revealed the existence of an anchor function with a specific directionality based on the N-terminal pI value (termed as N-terminal pI-specific directionality) and distinguished the presence of the *E. coli* type-II cytoplasmic membrane translocation pathways of Tat, Yid, and Sec for the unfolded and folded target proteins. We concluded that the pI value and hydrophilicity of the short N-terminal polypeptide, and the total translational efficiency of the target proteins based on the ΔG_{RNA} value of the N-terminal coding regions are important factors for promoting more efficient translocation (secretion) through the largest diameter of the Tat channel. These results show that the short N-terminal polypeptide could substitute for the Tat signal sequence with improved efficiency.**

INTRODUCTION

E. coli is frequently used as a host for recombinant protein expression because of its relatively simple, quick, and inexpensive cultivation, its well-characterized genetics, and the availability of a large number of useful molecular tools (Sørensen and Mortensen, 2005). Despite these merits, eukaryotic cell systems are gaining favor for the production of more complex recombinant proteins, such as glycosylated proteins or those with complex folding characteristics. Nevertheless, many soluble and insoluble recombinant proteins have been produced successfully in *E. coli*, although additional strategies for the production of soluble recombinant proteins in *E. coli* are needed.

In *E. coli* expression systems, proteins that are expressed in soluble form are much more economically and rapidly purified than are those that are expressed in insoluble form in inclusion bodies. Soluble recombinant proteins that are secreted into the medium or into the periplasmic space are even more readily recovered. *E. coli* has two types of secretion mechanisms that can be used to secrete a number of native proteins (Mergulhão et al., 2005). The type-I mechanisms export high molecular-weight toxins and exoenzymes to the culture medium (Fernandez and de Lorenzo, 2001), whereas the type-II mechanisms utilize a two-step process for extracellular secretion in which periplasmic translocation (Koster et al., 2000) is followed by outer membrane translocation.

In an attempt to develop a general approach to address the problem of the periplasmic translocation, we previously used hydropathy profile analysis to demonstrate that the presence of an internal, positively charged transmembrane (TM)-like domain in the olive flounder protein hepcidin I (HepI) inhibits its soluble expression. Using a novel secretion enhancer, we were able to overcome the obstacle posed by the internal TM-like domain and successfully expressed HepI in soluble form, which was consistent in showing a correlation between expression of soluble rHepI and high hydrophilicity (Lee et al., 2008b).

We also investigated Mefp1, an adhesive protein of the marine mussel *Mytilus edulis* (Waite, 1983). This large protein consists primarily of repeated decapeptide units lacking a TM-like domain (Lee et al., 2008b). When we fused 7xMefp1 to a truncated OmpA signal peptide ($OmpASP_{tr}$) in an attempt to increase its expression in soluble form, we found that as the pI value of the N-terminus of the recombinant 7xMefp1 fusion protein (rMefp1) increased, from 9.90 to 10.82, periplasmic expression of soluble rMefp1 also increased (Lee et al., 2008a).

We further investigated the effect of the pI (ranging from

Biotechnology Research Division, National Fisheries Research and Development Institute, Busan 619-902, Korea
*Correspondence: sangjl@nrfdi.go.kr

Received April 26, 2011; revised July 1, 2011; accepted July 12, 2011; published online October 17, 2011

Keywords: ΔG_{RNA} value of the N-terminal coding region, *E. coli* type-II cytoplasmic membrane translocation pathways (Tat, Yid, and Sec), highly hydrophilic N-termini of the modified signal sequences, N-terminal hydrophilicity, N-terminal pI with an anchor function (N-terminal pI specific directionality)

2.73-13.35) of the N-terminal region of 7xMefp1 on soluble and insoluble expression (Lee et al., 2010). We found that the N-terminal pI value can be used as a comprehensive biological index to represent the level of soluble rMefp1 expression and as a pI value trigger for facilitated diffusion from the total expression. When soluble expression was plotted against N-terminal pI values, we identified three curves, one for each pI range studied (acidic, neutral, alkaline). We surmised that each such curve was derived from a different periplasmic translocation (secretion) pathway involving an inner membrane channel that accommodates a specific range of N-terminal pI values (acidic, neutral, or alkaline) (Lee et al., 2010). However, *E. coli* type-II cytoplasmic membrane translocation pathways for soluble expression are very complex (Mergulhão et al., 2005), leading to the conclusion that the presently known pathways are insufficient to explain the mechanism of cytoplasmic membrane translocation for soluble expression.

We previously proposed that the pI value of the short N-terminal of the signal sequence could represent the whole length as a directional signal (Lee et al., 2008a). We also suggested the use of the novel *E. coli* type-II cytoplasmic membrane translocation pathways of Tat, Yid, and Sec for the unfolded and folded target proteins (Fig. 1) that correspond to the pI-range-specific soluble expression curves (Lee et al., 2010) based on the N-terminal pI values of known signal sequences or leader sequences and the known diameter of the translocation channels. We hypothesized that the Tat pathway has a strong co-relationship with the soluble expression of folded target proteins, but not those associated with Yid and Sec. To validate the proposed translocation pathways for the folded protein, we investigated the soluble expression of GFP with the established pI-range values of N-termini with acidic, neutral, and alkaline curves (Fig. 1A) to overcome the complexity of the pI value associated with the Tat signal sequence (Supplementary Table S2). We also examined the regular size and position of the secretion enhancer sequence for hydrophilicity (Lee et al., 2008b) to substitute for the position of the twin-arginine motif required for translocation of the Tat substrates.

Our results showed that the pI value and hydrophilicity of the short N-terminal polypeptide and the total translational efficiency of target proteins are important factors for translocation through the Tat pathway due to the diameter of the Tat translocon, which allows all folded bulky proteins to pass through and be secreted into the periplasm. These results established a sound foundation for classifying and simplifying the complex *E. coli* type-II cytoplasmic membrane translocation pathways of Tat, Yid, and Sec for the secretion of unfolded and folded target proteins, and represented an innovative breakthrough in terms of understanding the soluble expression of folded proteins with a short N-terminal polypeptide as a substituted Tat signal sequence in *E. coli*.

MATERIALS AND METHODS

Bacterial strains and plasmids

The *E. coli* strains XL-1 blue (Stratagene) and TOP10 (Invitrogen) were used for cloning, and BL21 (DE3) (Novagen) for direct expression of the fusion protein. The plasmid pBluescriptIISK(+) (Stratagene) and the TA cloning vector (Promega) were used for cloning, and pET-22b(+) (Novagen) for protein expression.

Reagents and molecular techniques

Restriction endonucleases (Roche) and a His-tag purification kit (Qiagen) were used. All other chemicals were of analytical grades. All molecular techniques were conducted as described (Sambrook et al., 1989). Nucleotide sequencing using the dideoxy chain-termination method (Sanger et al., 1977) was performed using the Sequenase 2.0 kit (United States Biochemical).

Computational analysis of pI and hydrophilicity, and ΔG_{RNA} values

pI values and hydrophilicity and hydropathy profiles (Hopp & Woods scale) were analyzed using the computer program DNASIS™ (Hitachi, Japan, 1997). The presence of several TM-like domains in the primary structure of GFP was confirmed using the hydropathy profile analysis (Lee et al., 2008b). The ΔG_{RNA} values for unfolding the RNA secondary structures of the N-termini were calculated using the program Mfold (Zuker, 2003).

Construction of GFP clones

For GFP and its derivative clones, the *gfp* region of the pEGFP-N2 vector (Clontech) was amplified using several forward primers containing the *Nde*I cleavage site (CAT) and a reverse primer containing the *Xho*I cleavage site (CTC GAG) after deleting the stop codon (TAA) (Supplementary Table S3). The amplified DNA fragments were cloned into the *Nde*I-*Xho*I of pET-22b(+) and the resulting pET-22b(+) (*N-terminal-gfp-XhoI-His tag*) clones and a control clone pET-22b(+) (*gfp-XhoI-His tag*) were obtained (Supplementary Table S3).

To test the Tat signal sequence, the *gfp* region of pEGFP-N2 vector (Clontech) found in the TorA signal sequence (TorAss) (Méjean et al., 1994)-GFP clone was amplified using the forward primer ($TorAss_{20-39}$-aqaa-GFP_{1-7}) and the above reverse primer. The first amplified DNA was subcloned and the subcloned DNA was re-amplified using the forward primer ($TorAss_{1-27}$) and the above reverse primer. The DNA fragment obtained by secondary PCR was cloned into the *Nde*I-*Xho*I of pET-22b(+). Additionally, for the Sec signal sequence, the OmpA signal sequence (OmpAss) (Movva et al., 1980)-GFP and its derivative clone were constructed (Supplementary Table S3).

GFP protein expression

E. coli BL21 (DE3) cells were transformed with the plasmid constructs listed in Supplementary Table S3, and the transformants were cultured in LB medium overnight at 30°C in the presence of 100 μg/mL ampicillin. The culture was then diluted 1:100 in LB medium and grown until it reached an optical density of 0.3 at 600 nm. Then, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the culture was grown for another 3 h to allow expression of the recombinant protein. An aliquot was then removed from each culture and centrifuged. The wet weight of the cell pellet was measured and resuspended in Tris buffer (50 mM Tris, pH 8.0). Cells were disrupted by sonication, in which 15 pulses at 30% power output were applied in 2-s cycles to release total proteins, and then the supernatant was obtained as the soluble fraction by centrifugation (16,000 rpm, 30 min, 4°C). Approximately 50 μg of total protein and its counter volume of soluble fractions were used to measure the fluorescence with Perkin Elmer Victor3.

Western blotting

For Western blot analysis, protein fractions were separated by SDS-PAGE on 15% acrylamide gels (Laemmli, 1970). After electrophoresis, His-tagged GFP proteins were transferred to a Hybond-P membrane (GE Healthcare) and detected using (in

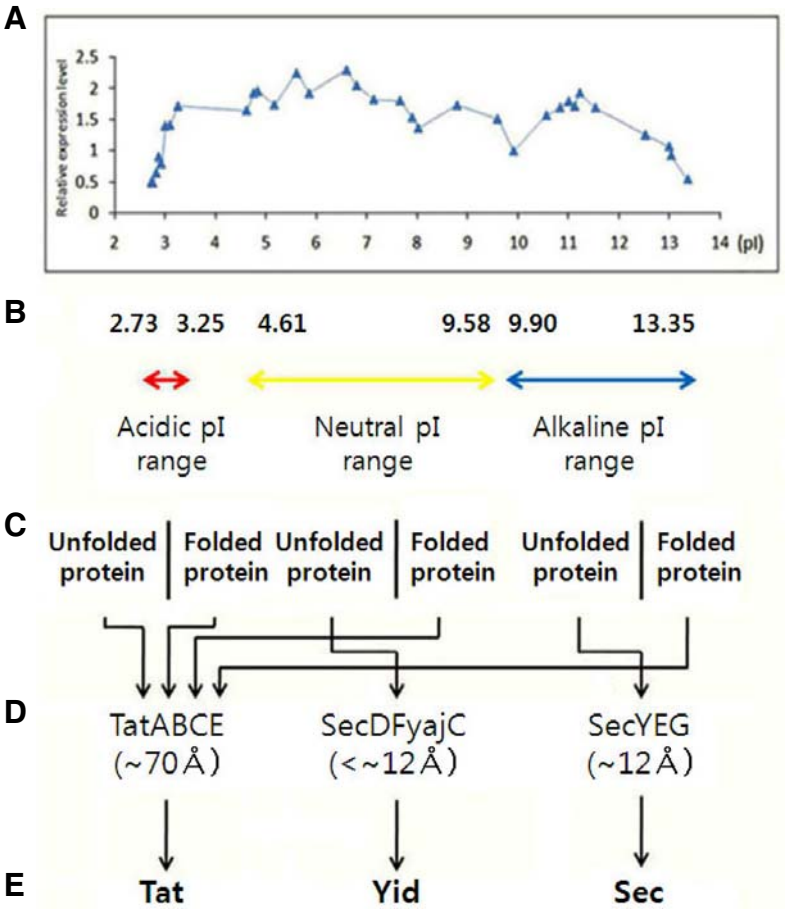

Fig. 1. Three pI range-specific (acidic, neutral, alkaline) soluble rMefp1 expression curves and their summarized defined *E. coli* type-II cytoplasmic membrane translocation pathways. (A) The curve represents the mean value of the pI range-specific soluble rMefp1 expression levels in a fused form with N-termini with pI 2.73-3.25, 4.61-9.58, and 9.90-13.35, respectively. (B) The acidic (red line), neutral (yellow line), and alkaline (blue line) pI ranges of soluble rMefp1 expression curves are delineated. (A) and (B) were adopted as referenced (Lee et al., 2010). (C) The types of unfolded and folded proteins. (D) The defined translocons and their pore sizes of Tat translocon (Sargent et al., 2002), Yid translocon (this study), and Sec translocon (van den Berg et al., 2004) (in the parentheses). (E) The defined *E. coli* type-II cytoplasmic membrane translocation pathways of Tat, Yid, and Sec are shown by their pI value ranges.

sequence) an anti-His tag (C-term) primary antibody, an alkaline phosphatase-conjugated anti-mouse secondary antibody, and a chromogenic Western blotting kit (Invitrogen), according to the manufacturers' protocols. Molecular weight markers (Benchmark, His-tagged) were used.

RESULTS AND DISCUSSION

Newly proposed *E. coli* type-II cytoplasmic membrane translocation pathways of Tat, Yid, and Sec

In our previous study, expression of a target protein lacking a TM-like domain fused to various N-terminal sequences yielded three pI range-specific curves (acidic, neutral, alkaline) for soluble expression (Lee et al., 2010). We surmised that each such curve represents a different periplasmic translocation (secretion) pathway involving an inner membrane channel that is specific for a range of N-terminal pI values (acidic, neutral, alkaline). Thus, we analyzed the pI values of the short N-terminals of the signal sequences that were representative of the whole length of the signal sequence as a directional signal (Lee et al., 2008a) to assess a clue for the corresponding translocation pathways.

In the alkaline pI range (pI 9.90-13.35; Fig. 1A), the pI values of the short N-terminals of well-known *E. coli* signal sequences were analyzed using the computer program DNASIS™ (Supplementary Table S1). The results showed that the N-domains of PhoA, OmpA, StII, PhoE, MalE, OmpC, Lpp, LTB, OmpF, LamB, and OmpT (Choi and Lee, 2004) exhibited pI values of 9.90-11.52, which is within the range of common pI values of the short N-terminals exhibiting alkaline soluble expression curves (i.e., pI 9.90-13.35; Fig. 1A). These signal sequences belong to the general secretory (Sec) pathway (Wickner et al., 1991), which is involved in the secretion of unfolded proteins across the cytoplasmic membrane.

Thus, we recognized that short N-termini of pI 9.90-13.35 (Fig. 1A) have a common pI range with the N-terminal regions of signal sequences in the Sec pathway. However, these sequences were too short to be considered full-length and lacked the hydrophobic core region for SecA interaction (Wang et al., 2000) and the c-region carrying the cleavage site for the signal peptidase (von Heijne, 1998), which would minimally require the translocation pathway. The minimal Sec pathway translocation pore passes proteins through the SecYEG membrane-embedded translocon (Fekkes and Driessen, 1999), which also has the minimal stoichiometric subunits of translocase for the integral membrane trimer (Douville et al., 1995). Thus, we hypothesized that the alkaline soluble expression curve corresponded to the Sec pathway containing the Sec translocon (Fig. 1).

In the neutral pI range (pI 4.61-9.58; Fig. 1A), when the *E. coli* insertase YidC binding to Pf3 (a small phage coat protein; 1-6 aas, pI 5.70; 1-7 aas, pI 3.00; full-length, 1-44 aas, pI 6.74) was examined *in vitro* using fluorescence spectroscopy (Gerken et al., 2008), all the resulting binding curves showed a strict hyperbolic form at pH values between 5.0 and 9.0. This result provided evidence that the insertase undergoes a conformational change upon binding Pf3 in this pH range. Thus, it seems clear that YidC acts within the neutral pI range of protein translocation for soluble expression (Fig. 1A) and that YidC co-

purifies with overproduced SecDFyajC, but not with overproduced SecYEG (Nouwen and Driessen, 2002). It has also been known that the YidC pathway has a threading mechanism for unfolded proteins (DeLisa et al., 2002), YidC can interact directly with a Sec-independent membrane protein (Chen et al., 2002), and *E. coli* YidC is essential for viability (Samuelson et al., 2000). These results demonstrated the presence of the novel Yid pathway (named for its association with the YidC-related pathway), differing from the Sec pathway and possessing a SecDFyajC translocon. Thus, we hypothesized that the neutral soluble expression curve corresponded to the Yid pathway containing the Yid translocon (Fig. 1), which would be limited to the translocation of relatively small molecules (discussed further below).

We also analyzed the pI values of the known putative *E. coli* Tat signal sequences (Tullman-Ercek et al., 2007) up to the 10_{th} amino acid using the computer program DNASIS™ (Supplementary Table S2). The results showed that Tat signal sequences have acidic, neutral, and alkaline pI values in single or mixed forms at their N-termini. Although Tat signal sequences have diverse N-terminal pI values, we concluded that the conserved twin-arginine motif is essential to translocate the folded protein into the periplasm. Therefore, we presumed that the original Tat signal sequences have acidic N-termini for the corresponding channel of the Tat pathway located in the acidic pI range (pI 2.73-3.25; Fig. 1A), and thereby exclude the highly-related neutral and alkaline N-termini (pI 4.61-9.58 and pI 9.90-13.35, respectively; Fig. 1A) of the Yid and Sec pathways, respectively.

For the other Sec pathway substrates corresponding to the alkaline N-termini of the Tat signal sequences, the expression of the ribose-binding protein N-terminal sequence MNMKK (pI 10.55) is significantly reduced in the periplasm of *tat* mutants, but is restored by *trans* expression of the *tatABC* genes (Pradel et al., 2003). Additionally, the activity of *Stenotrophomonas maltophilia* L2 β-lactamase, which has the N-terminal sequence MLARRR (pI 12.80) is much lower in a Δ*tatC* mutant than in a *secY-cs* mutant (Pradel et al., 2009). These results indicated that the Tat pathway is involved in the translocation of bulky folded proteins with the alkaline N-terminal sequences for soluble expression. This seems to contradict the results described in the above paragraph, which lead us to suggest that Tat signal sequences with low pI values use the Tat pathway and that Tat signal sequences with high pI values (which by our analysis, are like Sec signal sequences) do not use the Sec pathway, but may instead also use the Tat pathway.

Additional evidence suggests that the Tat channel faces considerably greater mechanistic challenges than does the Sec channel (Gohlke et al., 2005). Studies in *E. coli* demonstrated that the Tat system consists, at least, of four integral membrane proteins: TatA, TatB, TatC, and TatE (Sargent et al., 1999). Tat channels must be able to accommodate large substrates; some folded *E. coli* Tat substrates are close to 70 Å in diameter (Sargent et al., 2002). On the other hand, the Sec translocase has only to thread unfolded chains of about 12 Å diameter (van den Berg et al., 2004).

Given this information, we hypothesized that Tat signal sequences with high pI values are not "genuine" Tat signal sequences, but are instead Sec signal sequences. This is because the large, folded Tat substrates with high-pI Sec signal sequences cannot use the small Sec channel; thus, they use the large Tat channel instead. Thus, we concluded that the acidic pI range of the signal sequences belonging to the acidic soluble expression curve corresponded to the genuine Tat pathway (Fig. 1), but that the neutral and alkaline pI values in single or mixed forms at the N-terminal of the Tat signal sequences (Supplementary Table S2) could play an alternative role for Tat signal sequences when they are linked to the twin-arginine motif located in a proper position.

Overall, we revealed that unfolded proteins possessing N-termini with leader sequences of acidic, neutral, and alkaline pI values would be secreted into the periplasm through the Tat, Yid, and Sec pathways, respectively, in a similar manner to rMefp1 (Lee et al., 2010). Here, we concluded that rMefp1 belonged to the unfolded target protein, because it was secreted to the periplasm through the three proposed translocation pathways of Tat (transports folded proteins), Yid (threading mechanism for unfolded proteins), and Sec (secretes unfolded target protein), regardless of the channel diameter (Fig. 1). Otherwise, the bulky folded protein would be translocated into the periplasm through the largest channel by the Tat pathway, despite the pI value of the N-terminus (Fig. 1) (see below). Thus, we suggest a new model for *E. coli* type-II cytoplasmic membrane translocation pathways for the soluble expression of unfolded and folded proteins through the Tat, Yid, and Sec pathways (Fig. 1), based on the characteristic pI values of the N-termini of the signal sequences or the leader sequences, the known diameter of the translocation channels, and the N-terminal pI value of the signal sequence of the Tat substrates.

Effect of N-terminal pI with an anchor function on GFP expression

To analyze the effect of the pI value of the N-terminus on the soluble expression of the folded GFP with TM-like domains in the primary structure (see "Material and Methods"), we used different pI values of N-termini from the acidic, neutral, and alkaline curves (Fig. 1A; Lee et al., 2010) to minimize the complexity associated with the pI values of the N-termini of the Tat signal sequences (Supplementary Table S2). We also used the secretion enhancer sequence of hydrophilicity for the soluble expression of the folded GFP, as in the target protein with a TM-like domain (Lee et al., 2008b), to overcome the diverse position of twin-arginine motif in the Tat signal sequence (Supplementary Table S2). We modified the truncated OmpA signal peptide of $OmpASP_{1-8}$ (Lee et al., 2008a) to M (X = aa_1) (Y = aa_2)-TAIAI ($OmpASP_{4-8}$) for the N-terminus to obtain a variety of pI values and to maintain the function of the anchor (see below) with the $OmpASP_{4-8}$ containing up to a section of the hydrophobic region (Jobling et al., 1997), which was linked to the hydrophilic poly amino acids (8xArg). We calculated the pI value of M(X)(Y) and the hydrophilicity of M(X)(Y)-TAIAI-8xArg. M(X)(Y)-TAIAI ($OmpASP_{4-8}$)-8xArg-GFP clones (Supplementary Table S3) were constructed as described in "Material and Methods".

The constructed clones were transformed into *E. coli* BL21 (DE3) and the total and soluble fractions of GFP expression were observed (Fig. 2) because it was previously confirmed that the expression ratio of the soluble/total protein was a good index of the solubility of rMefp1, even at the lowest expression level (Lee et al., 2010). The clone with acidic pI MEE (pI 3.09), and high hydrophilicity (hy) MEE-TAIAI-8xArg (hy +1.34) showed the highest soluble expression than the GFP control. The clones with neutral pI, MAA (pI 5.60), and hydrophilicity MAA-TAIAI-8xArg (hy +1.16), and that with MAH (pI 7.65) and MAH-TAIAI-8xArg (hy +1.16), showed higher and lower soluble expression levels than the GFP control, respectively. The clones with alkaline pI MKK (pI 10.55) and hydrophilicity MKK-TAIAI-8xArg (hy +1.34), and with MRR (pI 12.50) and MRR-TAIAI-8xArg (hy +1.34), showed extremely low, but still detectable, soluble expression.

Moderately detectable levels of GFP fluorescence was pre-

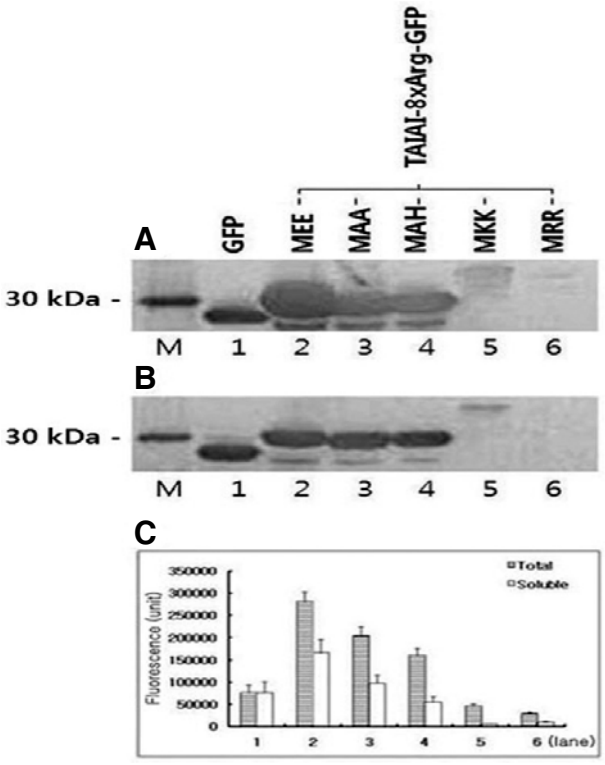

Fig. 2. Effect of pI value of the N-terminal with an anchor function on GFP expression. We constructed the M(X = aa_1)(Y = aa_2)-TAIAI (OmpASP_{4-8})-8xArg-GFP clones and calculated the pI value of the N-terminal M(X)(Y) and hydrophilicity (hy) of M(X)(Y)-TAIAI (OmpASP_{4-8})-8xArg (Supplementary Table S3). Approximately 50 μg protein of total and its counter volume of soluble fractions were used for Western blotting and fluorescence measurements as described in "Materials and Methods". The mean value of the fluorescence was obtained from three different colonies. The dark band around the size marker indicates the recombinant GFP. Western blot: total fraction (A), soluble fraction (B), fluorescence (C): total and soluble fractions. Lane M, marker; lane 1, GFP; lane 2, MEE (pI 3.09)-TAIAI-8×Arg [hy +1.34]-GFP; lane 3, MAA (pI 5.60)-TAIAI-8×Arg [hy+1.16]-GFP; lane 4, MAH (pI 7.65)-TAIAI-8×Arg [hy +1.16]-GFP; lane 5, MKK (pI 10.55)-TAIAI-8×Arg [hy +1.34]-GFP; lane 6, MRR (pI 12.50)-TAIAI-8×Arg [hy +1.34]-GFP.

sent in the total fraction, but very low levels were seen in the soluble fraction, indicating that the folded GFP with the leader peptides MKK and MRR were not secreted well into the periplasm or were blocked through the proposed Sec pathway (Fig. 1). These results were supported by the Western blot analysis. Specifically, the corresponding GFP of the total fractions showed wide and thinner bands with higher molecular weight, suggesting that the GFP bound to the membrane proteins or aggregated themselves in an insoluble form and the bands were unclear (Fig. 2A, lanes 5 and 6). However, the total fluorescence well reflected most of the cytoplasmic GFP that was converted into an insoluble form as described previously (Lee et al., 2010), so the pellet that formed after precipitation did not significantly interrupt the ability to measure low fluorescence levels in the soluble fraction (Fig. 2C, lanes 5 and 6).

When we compared the soluble GFP expression level of the clones described above, the short sequence of TAIAI (OmpASP_{4-8}) following the N-terminal leader sequence appears to act as an anchor, the pI value of which plays an important role in regulating the secretion levels of the fused GFP. These data support the presence of multiple channels from the Tat and Sec pathways at least. It is also the first evidence of a folded GFP with an anchor function and an N-terminal sequence portraying an acidic or alkaline pI value that is translocated through the Tat or Sec pathway, respectively. Therefore, we concluded that the anchor sequence assisted the N-terminal leader sequence with a specific pI value in obtaining an authentic pI-specific directionality for channel selection in the translocation of the fused GFP. The active form of the bulky folded protein with an N-terminal leader sequence having an acidic pI (Fig. 2, lane 2) should have been translocated through the approximately 70 Å diameter (Sargent et al., 2002) opening of the proposed Tat translocon (Fig. 1) and has the highest expression level of the soluble protein. Conversely, the active form of the bulky folded protein with an N-terminal leader sequence having an alkaline pI (Fig. 2, lanes 5 and 6) should target the approximate 12 Å diameter (van den Berg et al., 2004) opening of the proposed Sec translocon (Fig. 1), thus making translocation of the folded GFP difficult.

The GFP with a leader sequence having a neutral pI was expressed at a moderate level compared with the folded GFP control and at a slightly lower level compared with GFP with a leader sequence having an acidic pI (Fig. 2, lanes 3 and 4). Blocked expression was observed for the GFP with an alkaline leader peptide secreted through the Sec pathway (secretes unfolded target protein), but not for the GFP with a neutral leader peptide. Therefore, we hypothesized that the GFP with an anchor function and a neutral-pI leader peptide could exhibit a unique neutral pI-specific directionality that selects for the translocation channel found in the proposed corresponding Yid pathway (Fig. 1) in a similar manner to the acidic or alkaline pI-specific directionality, which showed the highest or extremely low GFP expression levels, respectively. However, the results showed that translocation of the folded GFP with an anchor function and a neutral N-terminus through the Yid pathway was not blocked, but rather high soluble GFP expression levels were observed (Fig. 2B, lanes 3 and 4). This suggests that the diameter of the Yid translocon is as large as that of the Tat translocon, or too small to allow translocation of the folded GFP. Nonetheless, it has been known that Yid pathway is involved in the threading mechanism for unfolded proteins only. Furthermore, it was previously unknown whether there existed another folded protein export pathway, in addition to the Tat pathway, that was large enough to allow translocation of the folded protein (Berks et al., 2000). Thus, we concluded that the Yid translocon has a much smaller diameter for transporting the folded GFP with an anchor function and a neutral N-terminal sequence compared with the Sec translocon, showing both blocked and low levels of GFP expression of those with an anchor function sequence and an alkaline N-terminus (Fig. 1).

The soluble expression kinetics of bulky folded GFPs with neutral N-termini revealed that MAA (pI 5.60), located close to the acidic region, had greater soluble expression than MAH (pI 7.65), which is close to the alkaline region (Fig. 1A). Based on this result, we determined that the acidic/neutral N-terminal MAA is close to the acidic region and has a higher specificity for the Tat channel than the alkaline/neutral N-terminal MAH, which did not appear to depend on the neutral-pI channel selectivity; MAH (pI 7.65) is closer to neutral than that of MAA (pI 5.60) (Fig. 1A). Thus, we concluded that a folded GFP with a neutral N-terminus pI value (> 5.60, < 7.65) would be secreted into the periplasm through the Tat pathway rather than Yid pathway. Therefore, the bulky folded protein was able to pass

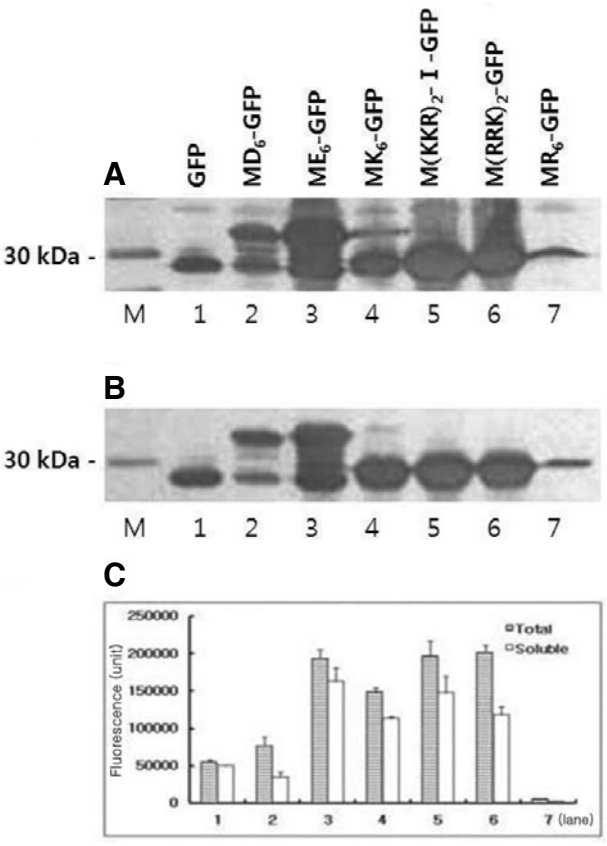

Fig. 3. Effect of the N-terminal hydrophilicity without an anchor function on GFP expression. We constructed the Met-6 x hydrophilic amino acids-GFP clones in which the 6x hydrophilic amino acids were 6xAsp, 6xGlue, 6xLys, 6xArg, 2x(Lys/Lys/Arg), and 2x (Arg/Arg/Lys). The pI values and hydrophilicities (hy) of Met-6 x hydrophilic amino acids were calculated (Supplementary Table S3). Western blotting and fluorescence measurements were conducted as in Fig. 2. The dark band around the size marker indicates the recombinant GFP. Western blot: total fraction (A), soluble fraction (B), fluorescence (C): total and soluble fractions. Lane M, marker; lane 1, GFP; lane 2, MD_6 (pI 2.56, hy +1.82)-GFP; lane 3, ME_6 (pI 2.82, hy +1.82)-GFP; lane 4, MK_6 (pI 11.21, hy +1.82)-GFP; lane 5, $M(KKR)_2$-I (pI 12.53, hy +1.82)-GFP; lane 6, $M(RRK)_2$ (pI 12.98, hy +1.82)-GFP; lane 7, MR_6 (pI 13.20, hy +1.82)-GFP.

efficiently through the Tat pathway via the translocon with the largest diameter and the passage was controlled by the pI value of the N-terminus with an anchor function sequence in the order of acidic > neutral.

Our results initially indicate that a short N-terminal polypeptide with a leader sequence composed of [N-terminal (pI value)]-space amino acids [TAIAI ($OmpASP_{4-8}$) with an anchor function sequence, but without affecting pI and hydrophilicity]-[hydrophilic secretion enhancer sequence (8xArg)], and exhibiting pI and hydrophilicity separately, is more accurate in channel selection than that of a short N-terminal polypeptide with pI and hydrophilicity occurring together without an anchor function sequence (see below). Therefore, we first localized the Tat, Yid, and Sec pathways that corresponded to acidic, neutral, and alkaline pI values of the N-termini with an anchor function using the folded target protein and deduced the diameter of Yid translocon (less than about 12 Å) (Fig. 1). The bulky folded protein was only able to successfully pass through the Tat channel via the translocon with the largest diameter, but was unable to pass through the translocon with the smallest diameter in the Yid channel, and was blocked by the intermediate-size translocon via the Sec channel, which is consistent with our suggestion for the *E. coli* type-II cytoplasmic membrane translocation pathways (Fig. 1).

Effect of N-terminal hydrophilicity on GFP expression

To investigate the effect of hydrophilicity of hydrophilic amino acids followed by Met as the N-terminus without an anchor function on GFP expression, we identified the hydrophilic leader sequences belonging to the acidic and alkaline regions of soluble expression curves (Fig. 1A; Lee et al., 2010). We then designed the Met-6x hydrophilic amino acids 6xAsp, 6xGlu, 6xLys, 6xArg (Lee et al., 2008b), 2x(Lys/Lys/Arg), and 2x(Arg/Arg/Lys), calculated the pI values and hydrophilicities of the leader sequences, and constructed the corresponding clones (Supplementary Table S3).

The constructed clones were transformed into *E. coli* BL21 (DE3) and GFP expression was observed (Fig. 3). The results showed that the acidic and highly hydrophilic leader sequence MD_6 (pI 2.56, hy 1.82) induced slightly higher total and slightly less soluble expression than that of the GFP control, whereas ME_6 (pI 2.82, hy 1.82) induced both high total and soluble expression, compared with those of the GFP control. This indicated that the acidic and high hydrophilic leader sequences MD_6 and ME_6 could induce soluble folded GFP expression through the Tat pathway.

Conversely, of the alkaline and high hydrophilic leader sequences, MK_6 (pI 11.21, hy 1.82), $M(KKR)_2$-I (pI 12.53, hy 1.82), and $M(RRK)_2$ (pI 12.98, hy 1.82) induced high total and soluble GFP expression, but MR_6 (pI 13.20, hy 1.82) induced very little total and soluble GFP expression. Additionally, for the leader sequences MK_6, $M(KKR)_2$-I and $M(RRK)_2$, the high levels of expression and fluorescence of the total protein were converted to high-level expression and fluorescence of the soluble proteins. However, in the previous section, we reported that most of the total GFPs with alkaline N-termini with an anchor function could not be converted to the soluble GFP due to blockage of the Sec pathway (Fig. 2, lanes 5 and 6). Thus, the secretion of total GFPs with the alkaline and highly hydrophilic N-termini without an anchor function should not be achieved through the relatively small pores of the Sec translocon. This indicates that cytoplasmic GFP has a prefolded, bulky active form, which could be secreted as soluble GFP through the largest diameter of the Tat channel (Fig. 3, lanes 4-7).

It was previously reported that GFP export was almost completely blocked in *tat* deletion (Δ*tatC*) mutants, whereas the active form of GFP was distributed uniformly throughout the cytoplasm, indicating that the observed export in wild-type cells occurred predominantly, if not exclusively, *via* the Tat pathway (Thomas et al., 2001). Thus, it seems clear that the GFP is first properly folded and changed to an active form in the cytoplasm, and then the active form of the bulky folded GFP should be translocated into the periplasm in a prefolded form through the Tat pathway, not via the Sec pathway. We thus concluded that the folded GFPs with the leader sequences MK_6, $M(KKR)_2$-I, and $M(RRK)_2$ could be secreted to the periplasm through the Tat pathway instead of the Sec pathway (Fig. 3, lanes 4-6). This is the first evidence that bulky folded proteins with alkaline and highly hydrophilic leader sequences are able to pass efficiently through the largest diameter Tat channel.

On the other hand, the leader sequence MR_6 induced low total GFP expression, indicating that most of the total GFP was

converted to the soluble form (Fig. 3, lane 7). This showed that the alkaline and highly hydrophilic leader sequence is involved in the high level of GFP secretion. Thus, the low GFP expression by the leader sequence MR_6 was caused by small quantitity of the total synthesized GFP, which was directly related to the solubility of the folded protein, as opposed to low secretion efficiency. Thus, we concluded that total cytoplasmic GFP expression influenced the solubility of the GFPs with alkaline and highly hydrophilic N-termini (see below).

Based on our results, we concluded that the soluble expression of folded target proteins with acidic or alkaline and highly hydrophilic N-termini without an anchor function was through the Tat pathway. The acidity and high hydrophilicity of the N-terminal of ME_6 induced one of the highest soluble expression levels observed, which is associated with a genuine Tat leader peptide as discussed above. In particular, the alkaline and highly hydrophilic nature of the N-terminal without an anchor function also enabled translocation of the folded GFP through the Tat pathway. These data suggest that high hydrophilicity of the N-terminus is more important than the pI value for determining passage through the Tat pathway rather than the Sec pathway, which is selective for alkaline pI values of the N-terminus with an anchor function. Thus, we demonstrated that highly hydrophilic N-termini without an anchor function could replace Tat signal sequences with single or mixed forms of pI values with acidic, neutral, or alkaline N-termini linked to the twin-arginine motif located in a proper position (Supplementary Table S2) and can only enable translocation of the folded target protein through the largest diameter Tat translocon (Fig. 1).

Additionally, we hypothesized that the low hydrophilicity of the twin-arginine motif is the main component of the Tat signal sequence that induces the soluble expression of the folded protein through cytoplasmic membrane translocation, which countered the highly hydrophilic sequences, 6xAsp, 6xGlu, 6xLys, 2xLysLysArg, 2xArgArgLys, and 6xArg that were analyzed in this study. Our results show that the translocation of the folded target GFP depends on the hydrophilicity of the N-terminal, which could be distinguished by the secreted target protein rMefp1 lacking a TM-like domain in the basic unit and depends on the pI value trigger of the N-terminal (Lee et al., 2010). Thus, we initially revealed that the solubility of the tertiary structure of the folded GFP was also regulated by the hydrophilicity of the N-terminal as in the target protein with the TM-like domain (Lee et al., 2008b). However, we suggest that inhibition of the solubility of the folded GFP is rarely influenced structurally by the primary structure of the TM-like domain located internally.

Effect of the ΔG_{RNA} value of the polynucleotides encoding the N-terminal regions on GFP expression

To investigate whether a difference in translation efficiency could account for the large difference in soluble GFP expression between MK_6 and MR_6 clones (Fig. 3), which have N-termini with similar pI values and hydrophilicities (Supplementary Table S3), we examined the ΔG_{RNA} values of the polynucleotides comprising the translation initiation region of pET-22b(+) (5′-AAG AAG GAG ATA TAC AT-3′) and the sequence regions encoding MK_6-$GFP_{1\text{-}5}$ (5′-ATG AAA AAA AAA AAA AAA-ATG GTG AGC AAG GGC-3′) and MR_6-$GFP_{1\text{-}5}$ (5′-ATG CGT CGC CGT CGC CGT CGC-ATG GTG AGC AAG GGC-3′), as described in "Materials and Methods". Several ΔG_{RNA} values for an RNA molecule indicate that several secondary structures may exist, whereas lower ΔG_{RNA} values indicate more stable secondary structures.

The ΔG_{RNA} values at the position described above in the MK_6 (Lys^{AAA})$_6$ coding region were 0.60 and 1.60, and that of the MR_6 ($Arg^{CGT}Arg^{CGC}$)$_3$ coding region was -13.80. Thus, the two different N-terminal coding regions are very different with respect to ΔG_{RNA} values, with the RNA encoding the MR_6 region having a more stable secondary structure compared with the RNA encoding the MK_6 region. Therefore, we suggest that the ΔG_{RNA} values of the N-terminal coding regions may explain the total cytoplasmic GFP expression levels for MK_6 and MR_6 clones. We constructed another N-terminal clone, M(KKR)$_2$-II (Supplementary Table S3), and compared the polynucleotide regions encoding the N-termini of M(KKR)$_2$-I ($Lys^{AAA}Lys^{AAA}Arg^{CGC}$)$_2$, M(KKR)$_2$-II ($Lys^{AAG}Lys^{AAA}Arg^{CGC}$)$_2$, and M(RRK)$_2$ ($Arg^{CGT}Arg^{CGC}Lys^{AAA}$)$_2$, which all exhibit hetero-type basic hydrophilic amino acids following Met. These are variants of the MK_6 and MR_6 clones, with similar or identical pI values and hydrophilicities (Supplementary Table S3). As described above, the corresponding calculated ΔG_{RNA} values of the above clones were -1.00, -0.50, -0.30; -1.00, -0.50, -0.30; and -7.60, respectively. We then analyzed the expression of these soluble GFP fusion clones after transformation into *E. coli* BL21 (DE3) (Fig. 4).

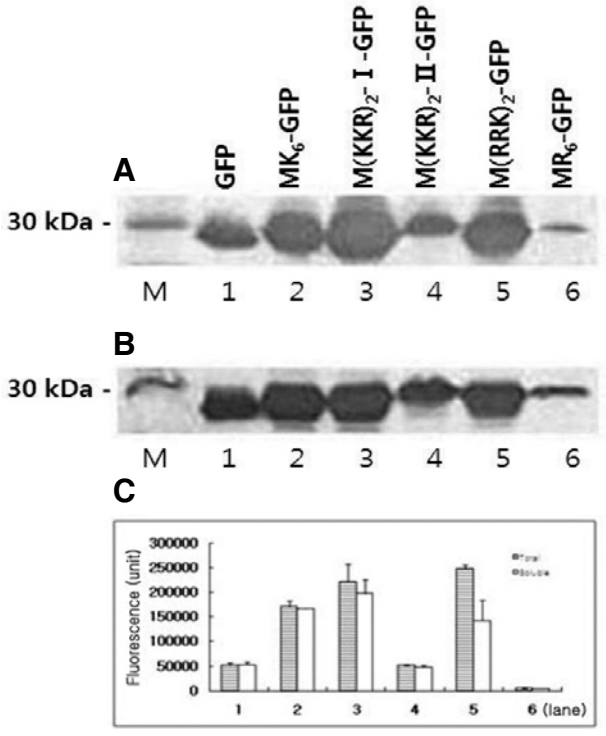

Fig. 4. Effect of the ΔG_{RNA} value of polynucleotides encoding the N-terminal regions of Met-basic hydrophilic amino acid sequences on GFP expression. The N-termini consist of homo- and hetero-type basic hydrophilic amino acids following the initial methionine. Western blotting and fluorescence measurements were conducted as in Fig. 2. The dark band around the size marker indicates the recombinant GFP. Western blot: total fraction (A), soluble fraction (B), fluorescence (C): total and soluble fractions. Lane M, marker; lane 1, GFP; lane 2, MK_6(Lys^{AAA})$_6$-GFP (ΔG_{RNA} 0.60, 1.60); lane 3, M(KKR)$_2$-I ($Lys^{AAA}Lys^{AAA}Arg^{CGC}$)$_2$-GFP)($\Delta G_{RNA}$ -1.00, -0.50, -0.30); lane 4, M(KKR)$_2$-II ($Lys^{AAG}Lys^{AAA}Arg^{CGC}$)$_2$-GFP($\Delta G_{RNA}$ -1.00, -0.50, -0.30); lane 5, M(RRK)$_2$ ($Arg^{CGT}Arg^{CGC}Lys^{AAA}$)$_2$-GFP ($\Delta G_{RNA}$ -7.60); lane 6, M(RR)$_3$ ($Arg^{CGT}Arg^{CGC}$)$_3$- GFP (ΔG_{RNA} -13.80).

The M(KKR)$_2$-I clone produced a relatively higher level of soluble GFP expression than that expected based on the

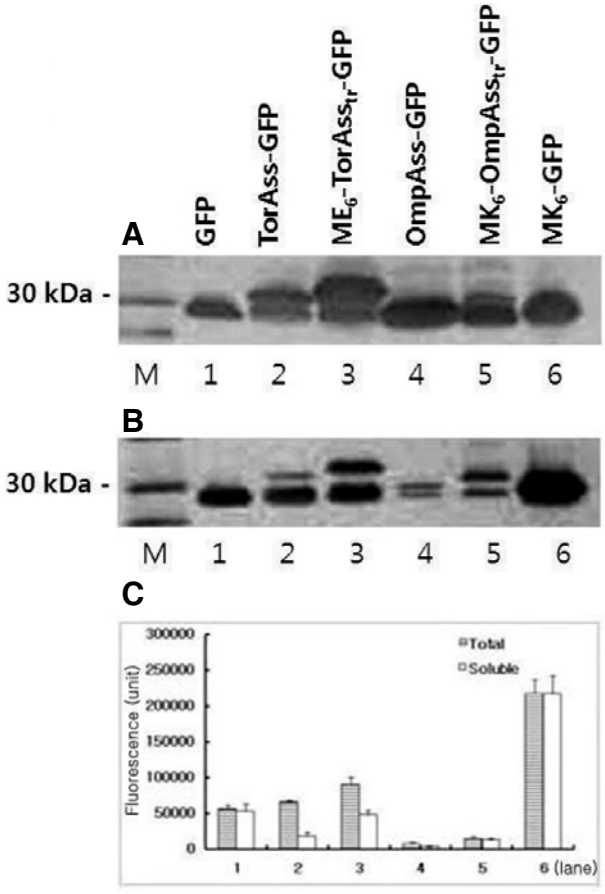

Fig. 5. Effect of high hydrophilicity in the N-terminal of the modified signal sequence on GFP expression. MNNND ($TorAss_{1-5}$) (pI 3.00, hy not tested) of the TorAss-GFP was replaced by the ME_6 (pI 2.82, hy +1.82) to construct the ME_6-$TorAss_{tr}$-GFP, and MKK ($OmpAss_{1-3}$, pI 10.55, hy not tested) of OmpAss-GFP was replaced by MK_6 (pI 11.21, hy + 1.82) to construct the corresponding MK_6-$OmpAss_{tr}$-GFP clones (Supplementary Table S3). Western blotting and fluorescence measurements were conducted as in Fig. 2. The dark band around the size marker indicates the recombinant GFP. Western blot: total fraction (A); soluble fraction (B), (C) fluorescence: total and soluble fractions. Lane M, marker; lane 1, GFP, control; lane 2, TorAss-GFP, control (upper band, TorAss-GFP; lower band, GFP); lane 3, ME_6-$TorAss_{tr}$-GFP (upper band, ME_6-$TorAss_{tr}$-GFP; lower band, GFP); lane 4, OmpAss ($OmpAss_{1-3}$, MKK)-GFP (upper band, OmpAss-GFP; lower band, GFP); lane 5, MK_6 (pI 11.35, hy +1.82)-$OmpAss_{tr}$-GFP (upper band, MK_6-Omp Ass_{tr}-GFP; lower band, GFP); lane 6, MK_6 (pI 11.35, hy +1.82)-GFP, control.

ΔG_{RNA} values. However, the clones $M(KKR)_2$-I and -II, with the same ΔG_{RNA} values, showed noticeable differences in the extent of soluble GFP expression. The $M(RRK)_2$ clone, which has a relatively low ΔG_{RNA} value, showed somewhat higher expression levels and fluorescence than expected. Thus, one N-terminal clone showed a correlation between the soluble expression level of GFP and the ΔG_{RNA} values, whereas the other showed no correlation, although all of the clones had heterotype basic hydrophilic amino acids following Met. In particular, the clones $M(KKR)_2$-I and -II showed remarkable differences in soluble GFP expression, despite having the same ΔG_{RNA}, pI, and hydrophilicity values. This remarkable difference in expression may be due to the codon wobble phenomenon (Lee et al., 2010) of the anticodon UUU for Lys between Lys^{AAA} and Lys^{AAG} (Lys^{AAA} > Lys^{AAG}) located at the 2nd codon position behind the initial Met codon. Thus, excluding this exception, we concluded that the ΔG_{RNA} value of the N-terminal coding region may be another criterion for the soluble expression of the GFP fusion protein.

Our results suggest that the hydrophilicity of the N-terminal amino acid sequence plays an important role in the secretion of GFP and confirm that the total translational level of GFP is correlated with the extent of soluble GFP expression. Therefore, when the ΔG_{RNA} value of the hydrophilic N-terminal coding region (excluding the codon wobble phenomenon) controls the total translational level of a heterologous protein, this value can also predict the extent of the secretion. Therefore, we conclude that the secretion of a heterologous protein through the Tat pathway should depend on the channel selectivity of the protein owing to the pI and hydrophilicity of its N-terminal amino acid sequence and on the total translational efficiency of the protein as determined by the ΔG_{RNA} of the N-terminal region polynucleotide sequence.

Effect of highly hydrophilic N-termini of the modified signal sequences on GFP expression

After recognizing that the acidic or alkaline nature, as well as the high hydrophilicity, of the N-terminal caused high levels of soluble GFP expression, we investigated the effects of the hydrophilicity of the N-termini of Tat and Sec signal sequences on soluble GFP expression. For Tat signal sequence, we modified the N-terminal TorA signal sequence (Méjean et al., 1994) to get a ME_6 (highly acidic and hydrophilic N-terminal; pI 2.82, hy 1.82) instead of the control $TorAss_{1-5}$ (MNNND, pI 3.00), and constructed an ME_6-truncated TorAss ($TorAss_{tr}$)-GFP clone. For the Sec signal sequence, we also constructed the control clone, OmpAss-GFP (control N-terminal, $OmpAss_{1-3}$, MKK, pI 10.55), using the OmpA signal sequence (Movva et al., 1980) and the highly alkaline and hydrophilic modified N-terminal clone MK_6-truncated OmpAss ($OmpAss_{tr}$)-GFP (MK_6, pI 11.35, hy 1.82) instead of the control N-terminal of MKK ($OmpASP_{1-3}$) (Supplementary Table S3).

The constructed clones were transformed into *E. coli* BL21 (DE3) and GFP expression was observed (Fig. 5). The total and soluble GFP expressions of the ME_6-$TorAss_{tr}$-GFP clones were slightly higher than those of the TorAss-GFP controls in the upper bands (Fig. 5AB, lanes 2 and 3). Additionally, the fluorescence levels of total and soluble ME_6-$TorAss_{tr}$-GFP expression were slightly higher than those of the TorAss-GFP controls (Fig. 5C, lanes 2 and 3). The total GFP expression of the two clones, OmpAss-GFP and MK_6-$OmpAss_{tr}$-GFP was higher than that of the TorAss-GFP control (Fig. 5A, lanes 4 and 5), whereas the soluble expression of two clones was less than that of the TorAss-GFP control (Fig. 5B, lanes 4 and 5). The fluorescence levels of total and soluble GFP expression of the two clones were also less than those of the TorAss-GFP controls, whereas the fluorescence of MK_6-$OmpAss_{tr}$-GFP was slightly higher than that of the OmpAss-GFP control (Fig. 5C, lanes 4 and 5). These results showed that the expression level of GFP with increased hydrophilicity of the modified N-terminal signal sequence was much less effective than that of the independent leader sequences of ME_6 of ME_6-GFP (Fig. 3, lane 3) and MK_6 of MK_6-GFP (Fig. 5, lane 6). This result indicated that high hydrophilicity of the N-terminal modified signal sequence did not effectively enable soluble GFP expression (see below).

Based on our results, we suggest that in the Sec signal se-

quence, the OmpA signal sequence and the modified OmpA signal sequence with high hydrophilicity of the N-terminal, the hydrophobic region bound to the SecA (Wang et al., 2000) and the cleavage site surrounded by the signal peptidase (von Heijne, 1998) are involved in reducing the folding process of GFP in the cytoplasm. In addition, relatively small amounts of the uncleaved and cleaved GFPs are secreted into the periplasm as a soluble form through the Sec pathway, as shown by the Western blot and fluorescence data (Fig. 5, lanes 4 and 5). Tat signal sequences also possess a common tripartite structure that includes a N-terminal (n-) region, a hydrophobic (h-) regions and a C-terminal (c-) region (Sargent, 2007), which is where the possible binding protein for the hydrophobic region and the apparent cleavage site of the TorA signal sequence, as well as its modified version with high hydrophilicity of the N-terminal, reduces both the GFP folding process in the cytoplasm and the translocation process of the uncleaved and cleaved target protein into the periplasm as a soluble form through the Tat pathway, as shown in Western blot and fluorescence (Fig. 5, lanes 2 and 3). Therefore, we recognized that the general expression kinetics of GFP fused with the TorA or the OmpA signal sequence with its modified versions of high acidic or alkaline and hydrophilic N-terminals, respectively, are involved in Tat- (TorA) and Sec- (OmpA) signal sequence dependent pathways (Fig. 5, lanes 2-5). Thus, we initially discovered that the modified signal sequences of N-termini that were highly acidic or alkaline and hydrophilic were less effective than the independent acidic and alkaline and highly hydrophilic leader sequences developed in this study for the soluble expression of folded proteins (Fig. 3, lanes 3-6; Fig. 4, lanes 2,3,5; Fig. 5, lane 6).

Here, we established that the acidic, neutral, and alkaline pI range curves associated with the Tat, Yid, and Sec channels of the *E. coli* type-II cytoplasmic membrane translocation pathways, respectively, for unfolded and folded target proteins by analyzing the pI value of the N-terminal of the signal sequences or the leader sequences, the known and deduced diameters of the translocation channels, and the solubility characteristics of the bulky folded target protein possessing short N-terminal polypeptides with or without an anchor function.

Based on the effects of short N-terminal polypeptides with an anchor function on the soluble expression of the bulky folded target protein, we concluded that acidic and neutral N-termini are related to the Tat channel, whereas alkaline N-termini are affiliated with the Sec channel as determined by the pI value range of the N-terminus. Additionally, analysis of the soluble expression of a bulky folded target protein using a highly acidic or alkaline hydrophilic N-terminals without an anchor function revealed that the hydrophilicity of the N-terminal is the major factor in determining Tat pathway secretion.

In general, our results showed that the relationship between the pI and hydrophilicity of the short N-terminal polypeptide with or without an anchor function (Figs. 2 and 3, respecitvely) is correlated with the association between the diverse pI ranges of Tat signal sequences at the N-terminus and the low hydrophilicity of the twin-arginine motif located close to or distant from the N-terminus (Supplementary Table S2). The novel design of the short N-terminal polypeptide for Tat substrates could substitute for the Tat signal sequence with better efficiencies. Therefore, we suggest that this study serves a good model for intelligent design based on a complex natural phenomenon derived through evolution.

Furthermore, we consider that this technique can be applied broadly to produce recombinant proteins and for protein transportation in many organisms given that the short N-terminal polypeptides can substitute for the whole length of the signal sequences which play common roles in protein targeting, membrane translocation, and usually exhibit interchangeable relationships between prokaryotic and eukaryotic cells (Valent et al., 1995).

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by National Fisheries Research & Development Institute (RD-11-BT-026).

REFERENCES

Berks, B.C., Sargent, F., and Palmer, T. (2000). The Tat protein export pathway. Mol. Microbiol. *35,* 260-274.

Chen, M., Samuelson, J.C., Jiang, F., Muller, M., Kuhn, A., and Dalbey, R.E. (2002). Direct interaction of YidC with the Sec-independent Pf3 coat protein during its membrane protein insertion. J. Biol. Chem. *277,* 7670-7675.

Choi, J.H., and Lee, S.Y. (2004). Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl. Microbiol. Biotechnol. *64,* 625-635.

DeLisa, M.P., Samuelson, P., Palmer, T., and Georgiou, G. (2002). Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J. Biol. Chem. *277,* 29825-29831.

Douville, K., Price, A., Eichler, J., Economou, A., and Wickner, W. (1995). SecYEG and SecA are the stoichiometeric components of preprotein translocase. J. Biol. Chem. *270,* 20106-20111.

Fekkes, P., and Driessen, A.J. (1999). Protein targeting to the bacterial cytoplasmic membrane. Microbiol. Mol. Biol. Rev. *63,* 161-173.

Fernandez, L.A., and de Lorenzo, V. (2001). Formation of disulfide bonds during secretion of proteins through the periplasmic-independent type I pathway. Mol. Microbiol. *40,* 332-346.

Gerken, U., Erhardt, D., Bär, G., Ghosh, R., and Kuhn, A. (2008). Initial binding process of the membrane insertase YidC with its substrate Pf3 coat protein is reversible. Biochemistry *47,* 6052-6058.

Gohlke, U., Pullan, L., McDevitt, C.A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H.R., and Berks, B.C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. Proc. Natl. Acad. Sci. USA *102,* 10482-10486.

Jobling, M.G., Palmer, L.M., Erbe, J.L., and Holmes, R.K. (1997). Construction and characterization of versatile cloning vectors for efficient delivery of native foreign proteins to the periplasm of *Escherichia coli*. Plasmid *38,* 158-173.

Koster, M., Bitter, W., and Tommassen, J. (2000). Protein secretion mechanisms in Gram-negative bacteria. Int. J. Med. Microbiol. *290,* 325-331.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227,* 680-685.

Lee, S.J., Han, Y.H., Nam, B.H., Kim, Y.O., and Reeves, P.R. (2008a). A novel expression system for recombinant marine mussel adhesive protein Mefp1 using a truncated OmpA signal peptide. Mol. Cells *26,* 34-40.

Lee, S.J., Park, I.S., Han, Y.H., Kim, Y.O., and Reeves, P.R. (2008b). Soluble expression of recombinant olive flounder hepcidin I with a novel secretion enhancer. Mol. Cells *26,* 140-145.

Lee, S.J., Han, Y.H., Kim, Y.O., Nam, B.H., Kong, H.J., and Kim, K.K. (2010). N-terminal pI determines the solubility of a recombinant protein lacking an internal transmembrane-like domain in *E. coli*. Mol. Cells *30,* 127-135.

Méjean, V., Iobbi-Nivol, C., Lepelletier, M., Giordano, G., Chippaux, M., and Pascal, M.C. (1994). TMAO anaerobic respiration in *Escherichia coli*: involvement of the tor operon. Mol. Microbiol. *11,* 1169-1179.

Mergulhão, F.J.M., Summers, D.K., and Monteiro, G.A. (2005). Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. *23,* 177-202.

Movva, N.R., Nakamura, K., and Inouye, M. (1980). Amino acid sequence of the signal peptide of ompA protein, a major outer

membrane protein of *Escherichia coli.* J. Biol. Chem. *255,* 27-29.
Nouwen, N., and Driessen, A.J. (2002). SecDFyajC forms a hetero-tetrameric complex with YidC. Mol. Microbiol. *44,* 1397-1405.
Pradel, N., Santini, C.-L., Ye, C.-Y., Fevat, L., Gérard, F., Alami, M., and Wu, L.-F. (2003). Influence of tat mutations on the ribose-binding protein translocation in *Escherichia coli.* BBRC. *306,* 786-791.
Pradel, N., Delmas, J., Wu, L.-F., Santini, C.-L., and Bonnet, R. (2009). Sec- and Tat-dependent translocation of β-lactamases across the *Escherichia coli* innermembrane. Antimicrob. Agents Chemother. *53,* 242-248.
Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd eds. (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).
Samuelson, J.C., Chen, M.Y., Jiang, F.L., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G.J., and Dalbey, R.E. (2000). YidC mediates membrane protein insertion in bacteria. Nature *406,* 637-641.
Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74,* 5463-5467.
Sargent, F. (2007). The twin-arginine transport system: moving folded proteins across membranes. Biochem. Soc. Trans. *35,* 835-847.
Sargent, F., Stanley, N.R., Berks, B.C., and Palmer, T. (1999). Sec-independent protein translocation in *Escherichiacoli.* A distinct and pivotal role for the TatB protein. J. Biol. Chem. *274,* 36073-36082.
Sargent, F., Berks, B.C., and Palmer, T. (2002). Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. Arch. Microbiol. *178,* 77-84.
Sørensen, H.P., and Mortensen, K.K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli.* Microbial Cell Factories *4,* 1 doi:10.1186/1475-2859-4-1.
Thomas, J.D., Daniel, R.A., Errington, J., and Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli.* Mol. Microbiol. *39,* 47-53.
Tullman-Ercek, D., DeLisa, M.P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., and Georgiou, G. (2007). Export pathway selectivity of *Escherichia coli* twin Arginine translocation signal peptides. J. Biol. Chem. *282,* 8309-8316.
Valent, Q.A., Kendall, D.A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). Early events in preparation in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. EMBO J. *22,* 5494-5505.
van den Berg, B., Clemons, W.M.Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427,* 36-44.
von Heijne, G. (1998). Protein transport: life and death of a signal peptide. Nature *396,* 111-113.
Waite, J.H. (1983). Evidence for a repeating 3,4-dihydroxyphenylalanine- and hydroxyproline-containing decapeptide in the adhesive protein of the mussel, *Mytilus edulis* L. J. Biol. Chem. *258,* 2911-2915.
Wang, L., Miller, A., and Kendall, D.A. (2000). Signal peptide determinants of SecA binding and stimulation of ATPase activity. J. Biol. Chem. *275,* 10154-10159.
Wickner, W., Driessen, A.J.M., and Hartl, F.-U. (1991). The enzymology of protein translocation across the *Escherichia coli* plasma membrane. Annu. Rev. Biochem. *60,* 101-124.
Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406-3415.