Analysis of Factors Affecting the Periplasmic Production of Recombinant Proteins in *Escherichia coli*

MERGULHÃO, FILIPE J.*1,2* AND GABRIEL A. MONTEIRO1

1Centro de Engenharia Biológica e Química, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
2LEPAE, Faculty of Engineering of the University of Porto, Chemical Engineering Department, 4200-465 Porto, Portugal

Received: November 21, 2006 Accepted: January 4, 2007

Abstract Five fusion proteins between Z domains derived from Staphylococcal Protein A and Green Fluorescent Protein or Human Proinsulin were produced on the periplasm of *Escherichia coli*. The effects of the molecular weight and amino acid composition of the translocated peptide, culture medium composition, and growth phase of the bacterial culture were analyzed regarding the expression and periplasmic secretion of the recombinant proteins. It was found that secretion was not affected by the size of the translocated peptide (17–42 kDa) and that the highest periplasmic production values were obtained on the exponential phase of growth. Moreover, the highest periplasmic values were obtained in minimal medium, showing the relevance of the culture medium composition on secretion. *In silico* prediction analysis suggested that with respect to the five proteins used in this study, those that are prone to form α-helix structures are more translocated to the periplasm.

Keywords: *Escherichia coli*, periplasmic secretion, secondary structure, protein size, translocation efficiency

The Gram-negative bacteria *Escherichia coli* has been the favored host for the purpose of recombinant protein production [2, 23, 36]. The secretory production of recombinant proteins by *E. coli* has several advantages over intracellular production as inclusion bodies. In most cases, targeting protein production to the periplasmic space or to the culture medium facilitates downstream processing, folding, and *in vivo* stability, enabling the production of soluble and biologically active proteins at a reduced process cost. Although recent advances in this field have shed some light on the most important secretion mechanisms [4, 22], attempts to secrete recombinant proteins can still face several problems, such as the incomplete translocation across the inner membrane [3] and proteolytic degradation [14].

It has been reported [33] that an optimum translational level exists to achieve high-level secretion of heterologous proteins, otherwise secretion severely drops off. The secretion capacity of the *E. coli* transport machinery seems to be limited [18, 29], and when this capacity is overwhelmed, the excess of expressed recombinant protein is likely to accumulate in inclusion bodies [10]. It is therefore important to optimize the expression level, which can be done for instance by carefully balancing the promoter strength and gene copy number [19, 24] or by manipulation of culture conditions [10, 13]. Several factors can influence the secretion of a recombinant protein in *E. coli*. The most commonly referred are the size of the passenger polypeptide [17, 28] and the amino acid composition of the leader peptide [5, 16, 26] and of the target protein [15]. In this work, the influences of culture medium composition, growth phase of the bacterial culture, protein size, and amino acid composition were evaluated regarding expression and periplasmic secretion. The signal sequence from Staphylococcal Protein A (SpA) was used for translocation [25] of five protein fusions between Z domains derived from SpA and Green Fluorescent Protein (GFP) or Human Proinsulin. Fusions of different proteins with the synthetic Z domains used on this work have numerous applications, namely on process integration, bacterial and phage display, vaccine development, and immunodetection [34]. The use of GFP as a fusion partner is further advantageous because of the possibility of localization of the chimeras inside the cells [6, 37].

**MATERIALS AND METHODS**

Expression Vector Construction

The enhanced GFP coding sequence was amplified from plasmid pEGFP-N1 (Clontech, Palo Alto, CA, U.S.A.)
using primers that include flanking EcoRI and BamHI restriction sites (in bold):
eGFPf: 5'-G TA AGGAGCTG TTCAC-3'
GAA TTC
eGFPr: 5'-C AGGACTTG TACAGCTGTC-3'
CATGCCGAG

The PCR fragment was double-digested with EcoRI and BamHI (Promega, Madison, WI, U.S.A.) and cloned into vectors pFM7 [20], pFM17, and pFM18 [21], generating plasmids pFM20 (expressing ZZ-GFP), pFM21 (expressing Z-GFP), and pFM22 (expressing non-fused GFP). Automated DNA sequencing was performed on both strands of all plasmids pFM20 (expressing ZZ-GFP), pFM21 (expressing Z-GFP), and pFM22 (expressing non-fused GFP). Automated DNA sequencing was performed on both strands of all three vectors (StabVida, Oeiras, Portugal). The amino acid alignment of the proteins produced by these expression vectors was presented in Mergulhão et al. [21] with GFP being expressed instead of Proinsulin. Vector pFM23 (for cytoplasmic production of GFP-His being expressed instead of Proinsulin. Vector pFM is presented in Mergulhão et al. [21] with GFP being expressed instead of Proinsulin. Vector pFM23 (for cytoplasmic production of GFP-His) was constructed by digestion of pFM20 with Ndel/BamHI and cloning of the GFP gene into pET28A (Novagen, Madison, WI, U.S.A.).

Cultivation Conditions
E. coli JM109(DE3) (Promega) harboring plasmids pFM20, pFM21, or pFM22 was grown in 250-ml shake-flasks (37°C, 220 rpm) with 50 ml of LB (Sigma, St. Louis, MO, U.S.A.) supplemented with 100 µg/ml of ampicillin (Sigma). A correlation between the optical density of the culture and the number of cells was determined as previously described [18].

GFP Production and Purification
E. coli JM109(DE3) was transformed with plasmid pFM23 and grown in four 500-ml shake-flasks, each containing 125 ml of LB, as described above. The culture was induced with 0.2 mM IPTG (Calbiochem, San Diego, CA, U.S.A.), and 6 h after induction, the cells were harvested by centrifugation (15 min, 2,744 × g) and resuspended in 12 ml of Buffer I (50 mM NaH2PO4, 300 mM NaCl, pH 8). Cells were disrupted by sonication, and after centrifugation (15 min, 17,640 × g), the supernatant was incubated overnight (at 4°C with stirring) with 1.5 ml of Ni-NTA resin (Qiagen). The resin was packed into a FPLC column (Pharmacia, Uppsala, Sweden) and washed with Buffer I containing 20 mM imidazole (Calbiochem). Enhanced GFP was eluted in Buffer I containing 300 mM imidazole and dialyzed overnight at 4°C (against 10 mM NaH2PO4, pH 8). Protein was analyzed by SDS-PAGE (a single band was visible on a Coomassie-blue stained gel) and quantitated by the BCA method (Pierce, Rockford, IL, U.S.A.).

Protein Analysis
For Human Proinsulin fusions, protein analysis was performed as described in Mergulhão et al. [19], and for GFP, the methods are described below.

Cytoplasmic Fraction Analysis. A culture volume of 400 µl was used to harvest the cells by centrifugation. The pellet was resuspended in 400 µl of distilled water (dH2O) and added to a fluorescent cuvette containing 2 ml of dH2O. Fluorescence was read in a VersaFluorimeter (Bio Rad, Hercules, CA, U.S.A.) with the emission filter 510-520 nm and excitation filter 420-480 nm. A calibration curve was constructed with affinity purified GFP standards (0 to 2,740 ng) spiked with 400 µl of washed JM109(DE3) cells harboring pET28A plasmid. dH2O was added to a final volume of 2.4 ml prior to fluorescence measuring. Results were normalized by multiplying by the molecular weight of the protein and dividing by that of GFP (see Table 1) and are expressed in femtograms per cell (fg/cell). The standard error associated with this method was 6%, as evaluated by three independent calibration curves. In order to verify if the presence of cell membranes and other proteins was interfering with the GFP fluorescence that was measured, one assay was performed where readings

<table>
<thead>
<tr>
<th></th>
<th>ZZ-GFP</th>
<th>ZZ-Proinsulin</th>
<th>Z-Proinsulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td>367</td>
<td>215</td>
<td>157</td>
</tr>
<tr>
<td>Mol. weight (kDa)</td>
<td>41.5</td>
<td>24.1</td>
<td>17.4</td>
</tr>
<tr>
<td>% Positive</td>
<td>10.9</td>
<td>9.3</td>
<td>8.3</td>
</tr>
<tr>
<td>% Negative</td>
<td>15.0</td>
<td>14.0</td>
<td>13.4</td>
</tr>
<tr>
<td>% Hydrophobic</td>
<td>26.2</td>
<td>23.7</td>
<td>24.2</td>
</tr>
<tr>
<td>Theoretical πl</td>
<td>5.3</td>
<td>5.1</td>
<td>5.0</td>
</tr>
<tr>
<td>Aliphatic index</td>
<td>77.6</td>
<td>80.0</td>
<td>80.8</td>
</tr>
<tr>
<td>% α-Helix</td>
<td>29±5</td>
<td>30±5</td>
<td>47±9</td>
</tr>
<tr>
<td>% Extended strand</td>
<td>19±4</td>
<td>9±4</td>
<td>10±6</td>
</tr>
<tr>
<td>% Random coil</td>
<td>50±10</td>
<td>40±6</td>
<td>42±7</td>
</tr>
</tbody>
</table>

Table 1. In silico analysis results.
were made prior to and after cell disruption by sonication. The difference between these values was less than 10%, indicating that the effect of cell material was not significant on this assay.

**Periplasmic Fraction Analysis.** A periplasmic extract was prepared by an osmotic shock procedure from 600 µl of cell culture, as previously described [20]. Samples of periplasmic extract (0.1 to 0.05 µl, obtained by serial dilution in dH2O) were dissolved in 100 µl of coating buffer (15 mM Na2CO3, 30 mM NaHCO3, pH 9.6) and adsorbed onto the wells of Maxisorp microplates (Nalge Nunc, Copenhagen, Denmark). Affinity purified GFP was used as a standard to create a calibration curve by loading up to 10 ng of purified protein per well diluted in coating buffer.

The plate was incubated for 1 h at 37°C and overnight at 4°C. Samples were removed and the wells filled with 350 µl of phosphate-buffered saline (PBST) with 0.05% Tween 20 (PBST). The plate was incubated for 2 h at room temperature and wells were then washed with PBST (1 quick wash followed by 4 two-minute washes). The first antibody, a mouse anti-GFP monoclonal antibody (Santa Cruz, Santa Cruz, CA, U.S.A.) was then added (1:1,000 dilution in PBST). A h incubation for 1 h at room temperature, the wells were washed with PBST as before. Anti-mouse Ig, horseradish peroxidase-linked from sheep (Sigma) (100 µl in a 1:2,000 dilution in PBST) was used as the detection antibody. The plate was incubated with the conjugate for 1 h at room temperature and washed as before. A volume of 200 µl of substrate solution (TMB, Calbiochem) was added and the colorimetric reaction developed for 10 min in the dark, at room temperature. The reaction was stopped by adding 50 µl of 2 M H2SO4 to each well and the absorbance at 490 nm was read with a multiplate reader. The standard error associated with the method was 20%, as evaluated by diluting a reference sample in three different concentrations and loading each of them into two independent plates. Nonspecific binding of the primary antibody was assayed by comparing the signal obtained in a well containing coating buffer only and another well containing 0.1 µl of periplasmic extract from cells devoid of plasmid. Results varied 10%, indicating that binding of the primary antibody was specific. Regarding the secondary antibody, two wells containing 0.1 µl of periplasmic extract from cells devoid of plasmid were incubated with both antibodies in one case and only with the secondary antibody in the other case (PBST was used instead of the primary antibody). Results varied 9%, indicating that the binding of the secondary antibody was also specific.

Periplasmic secretion values were normalized as before and are presented in femtograms per cell (fg/cell). Secretion efficiency was calculated by dividing the production level on the periplasm by the total production (the sum of periplasmic and cytoplasmic protein).
RESULTS AND DISCUSSION

The effects of protein size, amino acid composition, growth phase, and culture medium composition were analyzed on the production of three GFP fusion proteins (ZZ-GFP, Z-GFP, and GFP) and two Proinsulin fusions (ZZ-Proinsulin and Z-Proinsulin) using the Staphylococcal Protein A (SpA) leader for translocation. Three secretion vectors were constructed for the production of ZZ-GFP, Z-GFP, and GFP using a high copy number plasmid (ColE1 origin), the constitutive SpA promoter, the SpA leader sequence, and a deletion derivative. The amino acid alignment of the produced proteins is outlined in Mergulhão et al. [21], replacing Proinsulin by GFP.

Expression and Secretion of GFP Fusion Proteins

Cytoplasmic production reached a maximum after the exponential phase for all the constructs (Figs. 1A and 1B). When considering the number of molecules per cell, the specific production of GFP was almost the double (6.7 × 10^4 vs. 3.5 × 10^5 molecules per cell) when compared with Z-GFP and ZZ-GFP. Among these three proteins, GFP is the smallest (see Table 1), but the fact that higher production values were obtained with this protein cannot be solely explained on the basis of its smaller size, as Z-GFP and ZZ-GFP, which have different sizes, have shown similar productions. Nevertheless, for all the constructs, the production values leveled off upon entry into the stationary phase (Figs. 1A and 1B). A decrease in production capacity upon entry into the stationary phase has been observed in the production of several recombinant proteins [8, 9, 27, 31] and can be explained by limitations in cell machinery components like initiation factors, elongation factors, RNA polymerase [8], ribosomes, and mRNA [9, 31] as well as the triggering of stress responses [12].

It has been shown that periplasmic GFP that has been translocated by the Sec system is not fluorescent because the protein is improperly folded when localized outside the cytoplasm [11]. Additionally, the Sec system is not compatible with the transport of folded proteins [22], and therefore, the fluorescent protein localized in the cytoplasm is not transported to the periplasm. Therefore, cytoplasmic production of GFP was assayed by fluorescence, since it has been demonstrated that fluorescence can be quantitatively related to the levels of recombinant protein produced [1]. However, an alternative method was developed to quantitate the periplasmic fraction, which is not fluorescent (assayed by ELISA).

Periplasmic production levels (Fig. 1C) peaked before the maximum value of cytoplasmic production was reached (Fig. 1B). This suggests that the transport capacity of the system might be saturated at this point. This result corroborates those previously obtained on the secretory production of ZZ- and Z-Proinsulin fusions [18, 21]. The maximum value of ZZ-GFP secretion was 25% and 59% higher than for Z-GFP and GFP, respectively. The relatively reduced level of GFP secretion when compared with the Z-fused proteins can partially be explained by the leader peptide that was used for its translocation. This leader peptide is a deletion derivative of the SpA signal sequence [21], and in silico analysis on the secretion probability of these three proteins revealed that using this derivative could cause a 17% reduction in translocation efficiency. After the maximum secretion value was attained, periplasmic protein dropped sharply (on average 2.7-fold) until maximum cytoplasmic production was reached. Even if translocation is not occurring at this phase, one would expect periplasmic production to level off as it was observed on the cytoplasm. Possible explanations for this observation are leakage of periplasmic contents to the culture medium and/or proteolytic degradation.

The secretion efficiency (Fig. 1D) was found to be independent of the size of the translocated peptide (27–42 kDa) considering the predicted decreased value for GFP. Higher secretion efficiencies were obtained on the exponential phase for all the constructs and the respective values decreased upon entry into the stationary phase (Figs. 1A and 1D).

Translocation Performance

In order to evaluate the factors that influence translocation performance, i.e., translocated amino acids per cell (aa/
cell), the data obtained from GFP fusions were compared with some new and previously obtained data from Proinsulin fusions [18, 21]. Comparison of ZZ-Proinsulin data in minimal M9 and in rich LB medium (Fig. 2) shows that maximum translocation values dropped from $9.7 \times 10^7$ in M9 to $3.7 \times 10^7$ aa/cell in LB (2.6-fold).

Regarding the effect of protein size on secretion, analysis of translocation performances for ZZ-Proinsulin and Z-Proinsulin (with 24 and 17 kDa, respectively) obtained of translocation performances for ZZ-Proinsulin and Z-proinsulin presented in Fig. 2, it seems that proteins that are prone to comparing the output of this analysis (Table 1) with data trends that were obtained are probably valid. Thus, by of algorithms was used to analyze all the proteins, the strand was predicted). However, since the same combination of algorithms was used accurately predicted the amount of α-helix, although the amount of predicted extended strand was lower than the known value (only about 30% of extended strand was predicted). However, since the same combination of algorithms was used to analyze all the proteins, the trends that were obtained are probably valid. Thus, by comparing the output of this analysis (Table 1) with data presented in Fig. 2, it seems that proteins that are prone to form α-helix structures were translocated with higher performance, and as the percentage of predicted extended strand and random coil increased, the maximum translocation values decreased. If such secondary structures are formed, then protein transport may be affected either by altered interactions with Sec components [16] or even by steric hindrance at the translocation site.

It has been suggested that the size of the exiting peptide can affect secretion performance [17, 28, 32] and that large cytoplasmic proteins may be physically impossible to translocate [3, 11]. The results presented here show that even for a 2.4-fold increase in the size of the exiting peptide, there was no effect on the secretion efficiency. We postulate that factors other than the size (e.g., formation of secondary structures or even charge distribution) were responsible for altered secretion performances, since the threading mechanism of the Sec pathway [22] should not impose a size limit constraint on its substrates [35].

Acknowledgments

F. J. M. Mergulhão acknowledges the receipt of a postdoctoral fellowship from the Operational Programme for Science Technology and Innovation, Ministério da Ciência e Tecnologia, Portugal.

References


