Overexpression in *Escherichia coli*, Purification and Characterization of *Thermoanaerobacter tengcongensis* Elongation Factor G with Multiple Rare Codons

Liqiang Zhang, Peng Xue and Hongjie Zhang*

Center of System Biology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China

**Abstract:** The *fusA* gene encoding a thermophilic protein EF-G with multiple rare codons was cloned from *Thermoanaerobacter tengcongensis* (*Tte*EF-G) and overexpressed in *Escherichia coli* by cotransferring a RIG plasmid to overcome the potential codon-bias problem originated from Arg, Ile and Gly. The recombinant protein was identified by MALDI-TOF-MS for molecular mass with approximation of 76 kDa and by trypsin digestion coupled LC-MS/MS for peptide sequence coverage of 61.3%. The *in vivo* complementary assay indicates that *Tte*EF-G could significantly rescue the *E. coli* LJ14 (*frr*) at the non-permission temperature of 42°C in the bi-transformant of *TteRRF* and *Tte*EF-G. This study indicated that coexpression of rare codons’ cognate tRNA is a useful method for protein overexpression in *E. coli*.

**Keywords:** Elongation factor G, RIG plasmid, gene expression, purification and characterization, *Thermoanaerobacter tengcongensis*.

**INTRODUCTION**

Elongation factor G (EF-G) is an essential factor for protein synthesis in bacteria and in eukaryotic organelles, promoting the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome [1, 2]. EF-G from bacteria has been a drug target for treating bacterial infection for several decades [3, 4]. Beyond that it has been identified as an essential protein in concerted interaction with ribosome recycling factor (RRF) during disassembly of the posttermination complex, freeing the ribosome for the next round of translation [5-7]. The interaction between EF-G and RRF has showed strong species specificity. To date, all reported EF-Gs are highly homologous in amino acid sequence, and only *Thermus thermophilus* EF-G crystal structure has been resolved in different forms [8-10]. There are four domains in EF-G, each well separated, and domains III and IV have been identified to participate in the specific interaction between EF-G and RRF domain II [6].

Previous work from our and foreign laboratories has showed that thermophilic proteins in relation with protein biosynthesis are good candidates for elucidating the mechanism of peptide biosynthesis. One advantage is that they can tolerate extensive mutation or truncation to a certain degree while keeping the protein folded in the assay temperature range in a temperature sensitive strain *in vivo*. This property has endowed it a powerful tool for providing useful structural and functional insight on protein biosynthesis. For example the expression and function study of *Tte*RRF and two chimeras with *Eco*RRF in *E. coli* have been served for the successful identification of each domain’s function in RRF [7]. Another example is that C-terminal deletion showed enhanced complementary activity of *ThhRRF in vivo* [11, 12]. In order to elucidate the post-termination complex disassembly mechanism, both RRF and EF-G structural information is essential. As a first step in understanding the structural bases of the species-specific interaction between *Tte*EF-G and *Tte*RRF, it is necessary to obtain sufficient quantities of purified *Tte*EF-G.

There are two *fusA* genes encoding EF-G like proteins in *T. tengcongensis* genome [13]. Gene *fusA* (NCBI locus: NP_623834) encodes a 690 amino acid peptide chain, and Gene *fusA2* (NCBI locus: NP_623870) encodes a 700 amino acid peptide chain. Homologous analysis indicates that *fusA* encoded protein has share 83.9% similarity with *Thh*EF-G, significantly higher than that of *fusA2* encoded protein, which shares 66% similarity with *Thh*EF-G. A significant characteristic is that the EF-G is encoded extensively by rare used codons in protein expression system of *Escherichia coli*. In this paper *Tte*EF-G was highly expressed in *E. coli*, and its biological function has been identified. The results indicate that *Tte*EF-G encoded by *fusA* gene has elongation factor G’s property and could rescue the temperature sensitive *E. coli* strain at the non-permission temperature when *Tte*EF-G and *Tte*RRF were co-transformed into that cells.

**MATERIALS AND METHODS**

**Bacterial Strains, Plasmids and Chemicals**

*Thermoanaerobacter tengcongensis* MB4 is a rod-shaped, gram-negative bacterium that propagates optimally at 75 °C. It was isolated recently from a hot spring in Tengcong, China [13]. It was used to prepare the genomic DNA for amplifying the *T. tengcongensis* EF-G gene in this study. *E. coli* LJ14 is a MC1061 strain (*frr*), of which the chromosomal wild-type *frr* allele is replaced with a mutant allele (*V117D*) [14], allowing *E. coli* LJ14 to grow at 30°C but not at 42°C. The strain was used for the *in vivo* activity assays of the concerted interaction of *Tte*EF-G with *Tte*RRF by com-
proliferation analysis, pET-28a is a T7 promoter initiated high expression vector for cloned gene. It was used for the pTEF-G expression. pQE-60 is a vector with a strong T5 promoter purchased from QIAGEN. pSTV-28 was purchased from Takara, Japan, which encodes Cm scored, and was used to cotransform pTEF-G gene into E. coli LJ14 strain for in vivo complementarity assay. RIG plasmid was a kind gift from W.G.J. Hol [15]. Restriction endonucleases and DNA ligation kit were from NEB and Takara, respectively. All other reagents were of analytic grade.

Cloning of the EF-G Gene from T. tengcongensis

Based on the putative EF-G gene sequence of T. tengcongensis [13] identified by bioinformatics, a forward, 5′-AGACAGCCATATGGCAAGGAATTTCAGCTTAGA TAAGTTTAGG-3′ and a reverse, 5′-GCAAGGATCTTTT ATTGGCTGATAAT ATTTGCTCAGC-3′ primer containing NdeI and BamHI sites (bold), respectively, were designed to amplify the T. tengcongensis EF-G gene by PCR using T. tengcongensis genomic DNA as a template. The PCR product was inserted into the NdeI and BamHI sites of pET-28a after digestion with the same enzymes, resulting in pET28-TteEF-G for expression of TteEF-G. The EF-G gene was also cloned into the pSTV-28 vector between the KpnI and BamHI sites, resulting in pSTV-TteEF-G for the complementation assay.

Expression and Purification of TteEF-G

pET-28-TteEF-G plasmid and RIG plasmid were cotransformed into E. coli BL21(DE3)pLysS cells. Transformants were grown in LB medium supplemented with 50 μg/ml kanamycin and 25 μg/ml chloramphenicol overnight at 37°C for small scale culture. The overnight culture (20 ml) was then diluted 50-fold with the same medium, and incubated at 37°C until a turbidity of 0.8 at 600 nm was obtained prior to induction by 0.5 mM IPTG for 4 h. The cell pellet was collected by centrifugation at 6,000 × g for 20 min at 4°C, resuspended in 45 ml of chelating buffer (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 10 mM imidazole), and frozen at –20°C until use. Because twenty amino acid residues including a His6-tag and a thrombin cleavage site were fused in frame to the N-terminus when the TteEF-G gene was cloned into pET28a vector, the TteEF-G protein was first expressed as a His6-tagged protein. To purify the His6-tagged TteRRF protein, the above bacterial cells were thawed at 37°C, ultrasonicated on ice for 1 min and then the supernatant (40 ml) was recovered after the lysed bacteria were centrifuged at 37,000 g for 30 min. The supernatant was applied directly onto a 5 ml of washing buffer (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 80 mM imidazole). Finally, the column was washed with elution buffer (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 200 mM imidazole), and monitored by absorbance at 280 nm. The major peak of protein was collected and then dialyzed against distilled water at 4°C. The purified fusion protein was then incubated with thrombin (Sigma T4648) in the ratio of 10 units of thrombin mg–1 protein in cleavage buffer (20 mM Tris-HCl, pH 8.0, 2.5 mM CaCl2, 10 mM MgCl2) at 4°C for 8 h. After digestion, the cleavage mixture was applied onto the second metal chelating chromatography column to remove the fusion partner. Further purification was performed using Sephacry S100 column after the N-terminal peptide was removed by thrombin cleavage. Protein purity was checked by SDS-PAGE.

Mass Spectrographic Analysis

For MALDI-TOF-MS analysis, a protein sample (10 pmol) was dissolved in 40 μl of a solution containing 50% MeCN and 0.1% trifluoroacetic acid for MALDI-TOF-MS analysis. A freshly prepared solution of α-cyano-4-hydroxy-cinnamic acid (10 mg/ml) in 70% MeCN and 0.1% trifluoroacetic acid was mixed with the same volume of the above sample solution. A total of 1.0 μl of sample mixture was applied to the MALDI target and allowed to dry. A MALDI-TOF mass spectrum was acquired on an AXIMA-CFP plus mass spectrometer equipped with a 337.1-nm nitrogen laser. Data from 100 laser shots were averaged for each spectrum.

LC-MS/MS analysis sample was first digested in SDS gel using reported conditions. In brief, protein band was soaked in wash condition (50% MeCN, 2.5 mM NH4HCO3) for 24 h at 4°C and then washed and dried. Freshly prepared trypsin solution (10 μl; 0.02 μg/μl trypsin in 50 mM NH4HCO3) was added to plug for in-gel protein digestion at 37°C for 16 h. Peptides were extracted by sonication in 10 μl of 50% MeCN, 0.1% Formic acid for 20 min and concentrated to dryness before use [16].

In Vivo Complementary Assay of the in Concerted Activity of TteEF-G

The in vivo assay was performed by monitoring complementation of the E. coli LJ14 (frr) with TteRRF and TteEF-G or with only TteRRF. Plasmid pQE-TteRRF and pQE-TteEF-G, which expresses TteRRF and TteEF-G, respectively, were cotransformed into E. coli LJ14 (frr), or as a contrast only pQE-TteRRF was transformed. Cultures were started with 0.06% (v/v) inoculum from freshly growth (at 30°C) overnight cultures, and growth at both permissive (30°C) and non-permissive (42°C) temperatures was monitored by recording culture turbidities at 600 nm at regular intervals [17].

RESULTS AND DISCUSSION

Construction of TteEF-G Expression Plasmid

As described in Materials and methods, the TteEF-G coding sequence was synthesized by PCR and cloned into
vector pET28, resulting in pET28-TteEF-G for the expression of the TteEF-G gene. The constructed protein contains a His$_{6}$-tag and thrombin cleavage site coding sequences in the expression vector, which simplify the purification of the TteEF-G protein. On this case three amino acid residues (GSH) was extended at EF-G’s N terminus.

**High Soluble Expression and Purification of the His$_{6}$-tagged TteEF-G Protein**

Fig. 1 lane A and lane B showed the expression of cloned TteEF-G gene in *E. coli* without or with IPTG induction, respectively. In both cases, there were no significant expressions of the cloned TteEF-G gene in *E. coli*. Checking the nucleotide sequence of TteEF-G, it could be found that certain amino acid are extensively encoded by rarely used codons by *E. coli* in highly expressed genes [15]. Fig. 2 and Table 1 indicate the rare codons in *fusA* from *T. tengcongensis*. Arg has been extensively encode by rare codon AGA/AGG with the appearance of 91.4%; Ile by rare codon AUA (ATA in DNA code) with the appearance of 52.6%, and Gly by GGA with the appearance of 42.4%. Moreover these rarely used codons by *E. coli* are appeared highly at the translation starting terminus and also appeared in doublets, and even tetralets (Table 1). Rare codons have been shown to greatly diminish expression of recombinant protein in *E. coli* because of translational stalling [18, 19]. One method of overcoming this codon bias is to re-engineer the TteEF-G gene to be overexpressed so that it uses the preferred codons of *E. coli*. Obviously this method is both costly and time consuming [20, 21]. Recently Baca and Hol tried to cotransform *E. coli* with a plasmid (RIG) carrying the rare codon’s cognate tRNA and could successfully express several parasite genes in *E. coli* highly [15]. In order to know whether this strategy could solve our problem, the TteEF-G gene was cotransformed with RIG-plasmid in *E. coli* BL21(DE). Expression of the His$_{6}$-tagged TteEF-G gene in *E. coli* was analyzed by SDS-PAGE. Fig. 1 lane C and lane D show the 10% SDS-PAGE pattern of the cell lysate from *E. coli* harboring pET28-TteEF-G before and after induction with IPTG for 4 h. A very intense staining protein band can be observed in the induced total cell lysate (Fig. 1, lane D), but not in the un-induced cell lysate (Fig. 1, lane C). Further examination indicates that the expressed TteEF-G is mainly in soluble form (data not shown). His$_{6}$-tagged TteEF-G was first purified to apparent homogeneity in one step by Ni$^{2+}$ chelating affinity chromatography (Fig. 1, lane E) except several weak bands with lower molecular weight. This His$_{6}$-tag was further cleaved by thrombin and then re-eluted by Ni$^{2+}$ chelating affinity column, and the unbound fraction was collected for further assay use.

**Identification of the TteEF-G Protein**

MALDI-TOF-MS analysis shows two major peaks at m/z 76, 376.2 and 38, 205.4, corresponding to single and double protonated TteEF-G, respectively (Fig. 3). The molecular mass of TteEF-G obtained by mass spectrometry is consistent with the theoretical value calculated from amino acid composition of the protein (76, 955.95) using the software of Paws. The relative deviation of the measured value versus its theoretical value is less than 0.8%. This relative large deviation may partly be due to the measurement having surpassed the standard marker range of BSA, of which the average molecular mass is 66, 430.09 Da.

![Figure 1. 10% SDS-PAGE analysis of TteEF-G at different condition. Lane A. total protein extract from BL21 (DE3) harboring pET28-TteEF-G alone before IPTG induction. Lane B. total protein extract from BL21 (DE3) harboring pET28-TteEF-G alone after IPTG induction for 4.0 h. Lane C. total protein extract from BL21 (DE3) harboring pET28-TteEF-G and RIG plasmid before IPTG induction. Lane D. total protein extract from BL21 (DE3) harboring pET28-TteEF-G and RIG plasmid after IPTG induction for 4.0 h. Lane E. purified TteEF-G protein after chelating affinity chromato- graphy. Lane M. protein molecular weight marker (Fermentas) with different molecular weight, band 1, 118 kDa; band 2, 86 kDa; band 3, 47 kDa; band 4, 34 kDa; band 5, 26 kDa and band 6, 19 kDa, respectively.](image)
Figure 2. Sequence analysis of fusA gene from *T. tengcongensis*. The rare codons for *E. coli* (AGA/AGG for Arg, ATA for Ile and GGA for Gly) were impressed by boxes.

Table 1. Usage of Rare Codon for Arg, Ile and Gly in fusA from *T. tengcongensis*

<table>
<thead>
<tr>
<th></th>
<th>Arg</th>
<th>Ile</th>
<th>Gly</th>
<th>Number of Multiple rare codon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(AGA/AGG)</td>
<td>(AUA)</td>
<td>(GGA)</td>
<td>doubles</td>
</tr>
<tr>
<td>No. of total codons</td>
<td>35</td>
<td>57</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>No. of rare codons</td>
<td>32</td>
<td>30</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Percentage of rare codons</td>
<td>91.4%</td>
<td>52.6%</td>
<td>42.4%</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. MALDI-TOF-MS of TtEF-G. The sample (10 pmol) was dissolved in 50% acetonitril and 0.1% (v/v) trifluoroacetic acid and applied onto the target by mixing it with α-cyano-4-hydroxycinnamic acid as the matrix. The single and double charged molecular peaks were labeled out.

Figure 4. MS/MS spectrum acquired for the peptide. The parent ion selected is m/z=1191.712. The b-ions and y-ions were labeled out.

Table 2. Calculated Masses for b and y Ions from MS/MS Measurements of the 1191.712 Da Tryptic Fragment

<table>
<thead>
<tr>
<th>b-ion</th>
<th>MH⁺ calc.</th>
<th>Residue</th>
<th>y-ion</th>
<th>MH⁺ calc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>114.09</td>
<td>I</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>227.18</td>
<td>L</td>
<td>11</td>
<td>1077.59</td>
</tr>
<tr>
<td>3</td>
<td>342.10</td>
<td>D</td>
<td>10</td>
<td>964.51</td>
</tr>
<tr>
<td>4</td>
<td>399.22</td>
<td>G</td>
<td>9</td>
<td>849.48</td>
</tr>
<tr>
<td>5</td>
<td>470.26</td>
<td>A</td>
<td>8</td>
<td>792.46</td>
</tr>
<tr>
<td>6</td>
<td>569.33</td>
<td>V</td>
<td>7</td>
<td>721.42</td>
</tr>
</tbody>
</table>
Functional Characterization of TteEF-G

As reported recently, the concerted interaction between RRF and EF-G during disassembly of the translation posttermination complex showed strong species specificity. This provides a possible way to measure the biological activity of the expressed protein in semi-quantitative. In a previous publication from our group, we have reported that TteRRF showed very low complementary activity in a temperature sensitive E. coli LJ14 (frr°C) transformant at 42°C. Only after prolonged incubation the E. coli strain growth could be observed. In contrast, the strain with EcoRRF transformant could proliferate very well at the same condition. The low in vitro activity of TteRRF in E. coli is possibly due to the low efficient recognition between TteRRF and EcoEF-G. To investigate whether TteEF-G is able to complement E. coli LJ14 harboring TteRRF gene, we monitored the growth of transformants harboring vector alone (pQE-60) or the recombinants pQE-TteEF-G and pQE-TteRRF, which express TteEF-G and TteRRF, respectively, or the recombinant pQE-TteEF-G and pQE-TteRRF together. Although all the transformants grown in liquid cultures well at the permissive temperature (30°C) (data not shown), only the transformant harboring pQE-TteEF-G plus pQE-TteRRF could grow significantly after 12 h incubation at the non-permissive temperature (42°C), while at that time the transformant with only EcoEF-G grows less, indicating that the expressed protein TteEF-G could recognition TteRRF efficiently than EcoEF-G (Fig. 5). The in concerted interaction of TteEF-G and TteRRF disassemble the translational posttermination complex, showing one of TteEF-G’s biological functions.

It should be pointed out that in the in vivo activity assay system, the TteEF-G gene could also be expressed without the help of RIG-plasmid. Although the amount of TteEF-G is very low, it is enough to rescue the E. coli LJ14 strain from heat shot at 42°C. This result indicates that E. coli itself has the ability of synthesis the rare codon’s cognate tRNA, but its low amount limits the synthesis rate of the extrinsic gene and/or made it difficult to accumulate in cell due to degradation. The existence of RIG-plasmid is mainly to increase the rate of peptide synthesis, not to alter the peptide synthesis mechanism. This study also indicated that coexpression of rare codons’ cognate tRNA may be a widely applicable method for protein expression in E. coli.

![Figure 5](image-url)  
**Figure 5.** Growth curves of various transformants of E. coli LJ14 (frr°C) at the non-permissive temperature (42 °C). Vector indicates the transformants with pQE-60 and pSTV-28; TteEF-G, the transformants with pQE-60 and pSTV-TteEF-G; TteRRF, the transformants with pQE-TteRRF and pSTV-28; TteRRF and TteEF-G, the transformants with pQE-TteRRF and pSTV-TteEF-G, respectively. All the transformants were cultured at 42 °C with the same start value of A600nm = 0.04.

**ACKNOWLEDGEMENTS**

The author is in debt to Prof GZ Jing for his kindness help and financial support for this work. We thank Professor A. Kaji, Department of Microbiology, School of Medicine, University of Pennsylvania, PA, U.S.A. for providing E. coli LJ14 and MRE6000 strains.

**REFERENCES**


