Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 313 (2004) 89–96
BBRC
www.elsevier.com/locate/ybbrc

Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance☆

Xiaoqiu Wu,[a] Hans Jörnvall,[a] Kurt D. Berndt,[b,c] and Udo Oppermann[a,*]

[a] *Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[b] *Center for Structural Biochemistry, NOVUM, Karolinska Institutet, SE-141 57 Huddinge, Sweden*
[c] *Department of Natural Sciences, Södertörns Högskola, Box 4101, SE-141 04 Huddinge, Sweden*

Received 14 November 2003

Abstract

Expression patterns in *Escherichia coli* of two small archaeal proteins with a natural content of about 30% rare codons were analyzed. The proteins, a histone-like protein from *Sulfolobus shibatae* (Ssh10), and a glutaredoxin-like protein from *Methanobacterium thermoautotrophicum* (mtGrx), were produced with expression plasmids encoding wild-type genes, codon-optimized synthetic, and GST-fusion genes. These constructs were expressed in BL21 (DE3), its LysS derivative, and modified strains carrying copies for rare codon tRNAs or deletions in the RNAseE gene. Both Ssh10 and mtGrx expression levels were constitutively high in BL21(DE3) and its derivatives, with the exception of the LysS phenotype, which prevented high level expression of the Ssh10 wild-type gene. Surprisingly, a codon-optimized mtGrx gene construct displayed undetectable levels of protein production. The translational block observed with the synthetic mtGrx gene could be circumvented by using a synthetic mtGrx-glutathione *S*-transferase (GST) fusion construct or by in vitro translation. Taken together, the results underscore the importance of mRNA levels and RNA stability, but not necessarily tRNA abundance for efficient heterologous protein production in *E. coli*.

Keywords: Rare codon gene; RNA stability; Heterologous protein expression; Fusion protein; Archaeal proteins; Structural genomics

The current availability of completed genome sequences from numerous organisms makes the structural and functional analysis of all gene products a demanding goal. Heterologous expression of proteins thus plays a central role in functional genomics initiatives [1]. Several eukaryotic and prokaryotic systems for expression of foreign genes exist, and make the production of a desired protein in many instances a straightforward procedure. Overexpression of target genes in *Escherichia coli* is often the method of choice, because of extensive knowledge of *E. coli* genetics, availability of versatile vector systems, and host strains, the ease of use, low costs, and high expression levels, exceeding often more than 30% of total cellular protein [2–4]. However, despite the many advantages of *E. coli*, high-level expression is not routinely achieved. Among the many reported causes preventing efficient heterologous protein production in *E. coli* are biased codon usage, gene product toxicity, solubility, mRNA secondary structure, and mRNA stability [5]. In addition, rare codon gene expression can lead to translational errors as a result of ribosomal stalling at a position requiring incorporation of amino acids coupled to minor tRNAs, or even at sites requiring major tRNAs, but which are depleted because of overutilization of a particular amino acid [6,7]. The mistranslational events related to rare tRNAs are observed as codon misreadings and as processing errors and they manifest themselves as amino acid substitutions or frameshift events. Specifically, the rare arginine (AGG, AGA, CGG, and CGA), leucine (CUA),

☆ *Abbreviations:* Ssh10, DNA binding protein with mass of 10kDa from *Sulfolobus shibatae*; mtGrx, glutaredoxin-like protein from *Methanobacterium thermoautotrophicum*.

* Corresponding author. Fax: +46-8-33-74-62.
E-mail address: udo.oppermann@mbb.ki.se (U. Oppermann).

0006-291X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2003.11.091

isoleucine (AUA), and proline codons (CCC) often lead to frameshift errors and ultimately to undesired products [6–12]. Accordingly, several systems are now available alleviating many of the problems through tight transcriptional regulation, or through heterologous expression of tRNAs to circumvent problems with rare codon usage [6,7].

In a previous study we reported on the use of gene fusion systems as a considerable alternative to enhance protein production and rapid purification of thermostable peptides exhibiting a high proportion of rare codons [13]. In the present study we extended these investigations by comparing different strategies to enhance rare codon gene expression. We compared expression levels quantitatively and analyzed the protein products of wild-type, fusion-tagged, and synthetic genes where the majority of rare codons were substituted with preferred *E. coli* codons. In particular, we were interested in which factors influence the expression of heterologous genes. As model rare codon genes we investigated Ssh10, a small basic DNA binding protein from the archaeon *Sulfolobus shibatae* [14], and mtGrx, a glutaredoxin-like protein from the hyperthermophilic prokaryote *Methanobacterium thermoautotrophicum* [15].

Experimental

Cloning of Ssh10 and mtGrx expression constructs. The Ssh10 gene was obtained from genomic DNA of *S. shibatae* strain DSM 5389 (Deutsche Sammlung für Mikroorgansimen, Germany) by PCR using gene-specific primers. A wild-type gene of mtGrx was obtained by assembling a synthetic gene [16]. These PCR products were cloned into the pET15b vector (Novagen) as *Nco*I/*Bam*HI fragments. This resulted in expression constructs without His-tag and protease site fusion, usually encoded by the vector. GST-tagged fusion constructs of the wild-type genes were made using the pGEX4T-1 vector (Amersham–Pharmacia Biotech). Synthetic, untagged expression constructs of Ssh10 and mtGRX were obtained by optimization of rare codons and RNA secondary structure using the codon database (www.kazusa.or.jp/codon) for codon adaptation in *E. coli*, and the program mfold (http://bioweb.pasteur.fr/seqanal). Synthesis was achieved by PCR using established protocols [16]. Untagged synthetic genes were cloned into the pET15 vector at *Nco*I sites (as above) and GST-tagged constructs were made in pGEX4T-1 (also as above). The codon frequencies and free energy values for RNA secondary structure formation of the Ssh10 and mtGRX wild-type and synthetic genes are given in Table 1. DNA sequence analysis was carried out on ABI 370 instruments.

Heterologous expression of Ssh10 and mtGRX in different hosts. All expression constructs described above (wild-type, synthetic and GST-fusion) were analyzed for expression levels by transformation into differentially modified *E. coli* host strains. Strains BL21(DE3), BL21(DE3)LysS, and an RNAse E deficient strain (BL21 Star (DE3)) were from Novagen, and a strain carrying copies of tRNA genes argU, ileY, and leuW for codons AGG/AGA, AUA, and CUA, respectively (BL21 (DE3)-codonplus-RIL), was from Stratagene. For expression studies, chemically competent cells were transformed with 50 ng plasmid and grown overnight on agar plates with appropriate antibiotics. Fresh colonies were inoculated at 37 °C into 12 ml LB medium with appropriate antibiotics and induction was started with 1 mM IPTG

Table 1
Characteristics of Ssh10 and mtGrx genes used in this study

Gene	Total codons	Rare codons[a]									Secondary structure ΔG (kcal/mol)	
		AGG[b]/AGA[b] (Arg)	CTA[b] (Leu)	ATA[b] (Ile)	CGG (Arg)	CCA/CCT (Pro)	AGT/TCA/TCC (Ser)	GGG/GGA (Gly)	ACA (Thr)	TGT/TGC (Cys)		
Ssh10_wt	32/97	8/8	2/2	5/11	0/8	4/4	5/9	6/6	2/6	0/0	—	
Ssh10_syn	7/97	0/8	0/2	0/11	1/8	1/4	4/9	1/6	0/6	0/0	—	
mtGrx_wt	21/85	2/3	1/2	5/9	1/3	5/5	1/2	3/4	1/2	2/2	−68.1	
mtGrx_syn	2/85	0/3	0/2	0/9	0/3	1/5	0/2	0/4	0/0	1/2	−62.3	

[a] Codons used by *E. coli* at a frequency of <1% are arbitrarily defined as rare codons [31].
[b] Documented rare codons causing expression problems in *E. coli* [5].

when OD_{600} reached 0.8. After overnight induction at 24 °C, the cells were harvested by centrifugation, resuspended in 1× PBS, and lysed by sonication. The lysed samples were heated at 85 °C and pelleted at 20,000*g* for 30 min. The supernatant was decanted and reserved for later analysis, and the pellet was resuspended in SDS buffer (insoluble fraction). In each experiment at least three colonies were used for each construct expression. Independent experiments were carried out minimally four times. In initial experiments parameters were screened for optimal expression conditions regarding inductor concentration, growth temperature, culture volume, and induction time.

In vitro translation and analysis. In vitro translation of circular plasmid DNA was carried out using the *E. coli* T7 S30 Extract system for circular DNA (Promega) in the presence of [^{35}S]methionine. Plasmid DNA (2 μg, 2.5 μl amino acid mixture minus methionine, 10 μl S30 Premix without amino acids, 0.5 μl [^{35}S]methionine (1200 Ci/mmol at 15 mCi/ml, Amersham–Pharmacia Biotech, Uppsala, Sweden), 7.5 μl T7 S30 extract, and nuclease free water to a final volume of 25 μl were mixed and incubated at 37 °C for 2 h. The resulting products were analyzed by SDS–PAGE. Following gel staining with Coomassie blue, the gel was dried by a vacuum gel drier. The fixed gel was exposed to a phosphorimaging screen (Molecular Dynamics, Sunnyvale, CA, USA) and analyzed using the Image QuaNT software (Molecular Dynamics, Sunnyvale, CA, USA).

Northern blot analysis. Cells, grown and induced as above, were harvested and total RNA was isolated from each sample using the RNeasy Mini kit (Qiagen). A digoxigenin-labeled, double strand DNA probe was synthesized using the PCR DIG Probe Synthesis Kit (Roche). Unincorporated nucleotides were removed using a Qiaquick PCR purification kit (Qiagen). The isolated total RNA was electrophoresed on a 1.5% formaldehyde–agarose gel and the quality of the RNA preparation was evaluated by visualization with ethidium bromide. Transfer of the RNA to a nylon membrane Hybond-N (Amersham Biosciences), prehybrization, and hybridization were carried out following the DIG Luminescent Detection Kit protocol (Roche).

Protein purification. Fusion proteins with the GST tag were purified by affinity chromatography using glutathione–Sepharose (Amersham–Pharmacia Biotech) as described [8]. Untagged Ssh10 and mtGrx were analyzed after heat treatment of the lysate (80 °C, 20 min), followed by a centrifugation step (20 min, 20,000*g*) resulting in removal of precipitated proteins and yielding a pre-purified fraction amenable to densitometry or MALDI spectrometry. Further purification was achieved on MonoQ (mtGrx) or Fast Flow CM Sepharose (Ssh10) columns (Amersham–Pharmacia Biotech).

SDS–PAGE and determination of expression levels. Expression levels and purity of protein samples were assessed by SDS–PAGE in 8–16% Tris–glycine or 4–12% NuPage gels (Novex), stained with colloidal Coomassie dye (Novex). Protein concentration of apparently pure samples was determined after hydrolysis in 6 M HCl by amino acid quantification on a Biochrome analyzer (Amersham–Pharmacia Biotech). For comparison of expression levels, SDS–PAGE gels were scanned with a Bio-Rad GS-710 Calibrated Imaging Densitometer, and relative levels of expressed proteins were determined using the PDQuest software package (Bio-Rad) with purified Ssh10 and mtGrx (for wild-type and synthetic gene constructs) or BSA (for GST fusion gene analysis) as standards. For comparative purposes, culture aliquots corresponding to identical cell densities were collected by centrifugation and lysed by sonication after lysozyme treatment (30 min, room temperature), followed by collection of supernatant (20 min, 20,000*g*).

Protein analysis. mtGrx and Ssh10 were analyzed by peptide mass mapping after in-gel digestion using MALDI-TOF (Voyager DE, PE Biosystems) and by total mass determination using ESI mass spectrometry (Q-TOF, Micromass). N- and C-terminal sequences of PVDF blotted, SDS–PAGE separated mtGrx and Ssh10 containing fractions from the different expression conditions were determined with Procise N and C protein sequencers (PE Biosystems). Circular dichroism spectra were recorded using an AVIV 202SF spectropolarimeter (AVIV Associates).

Results

Expression constructs encoding wild-type and codon optimized synthetic genes (labeled as $_{wt}$ and $_{syn}$, respectively) of the two small thermostable proteins Ssh10 and mtGrx were prepared and used for experiments, complemented with identical parental gene constructs with or without the glutathione *S*-transferase (GST) fusion tag. All constructs were analyzed for expression levels in *E. coli* BL21 strains with different genetic backgrounds. We used the BL21 (DE3) and BL21(DE3)LysS strains to determine basal expression levels and to circumvent possible problems of toxicity by using a tightly controlled operator when using the pET vectors. Interestingly, we found that Ssh10 but not mtGrx levels were reduced in the LysS strain, indicating interference with this T7 suppressor phenotype (Table 2, Figs. 1 and 2). The data obtained from BL21(DE3) were then compared to expression experiments using the modified strains BL21star(DE3) and BL21Codonplus(DE3)-RIL. These strains were chosen to analyze the influence of RNA stability and codon bias on heterologous expression of the two wild-type rare codon genes. For each construct, expression trials were performed by variation of induction time, inducer concentration, and growth temperature (data not shown), and optimal expression levels obtained in the different bacterial strains and conditions were used for further analysis.

Table 2
Comparison of relative Ssh10 and mtGrx expression levels

Host strain	Non-tag expression				Fusion expression			
	Ssh10_wt	Ssh10_syn	Grx_Hwt	Grx_Hsyn	Ssh10_wt	Ssh10_syn	Grx_Hwt	Grx_Hsyn
BL21(DE3) LysS	<1.0	5.6 ± 0.5	11.5 ± 0.6	ne	1.5	nd	3.8	2.0
BL21(DE3)	17.8 ± 1.0	16.0 ± 2.1	13.4 ± 3.1	ne	nd	nd	nd	nd
BL21Star(DE3)	25.8 ± 0.7	19.3 ± 4.9	11.0 ± 1.6	ne	1.9	nd	3.8	1.6
BL21Codonplus (DE3)-RIL	13.9 ± 2.5	18.6 ± 0.5	11.6 ± 0.2	ne	1.4	nd	3.5	1.9

Results represent yields from optimal expression conditions (200 ml culture volume, 24 °C, 16 h, 1 mM IPTG). Values are yields (mg protein/L culture) based on densitrometric analysis of expressed protein. *Abbreviations:* nd, not determined; ne, no expression detectable; wt, wild-type; syn, synthetic gene.

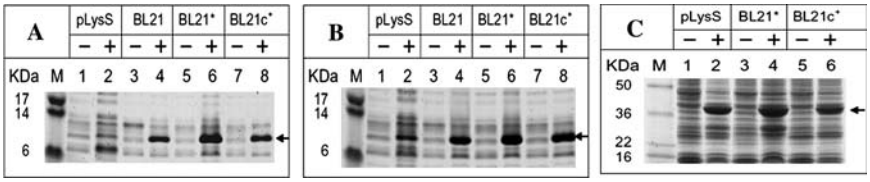

Fig. 1. Comparison of expression levels of wild-type (A), synthetic gene (B), and GST- (C) Ssh10 constructs in different *E. coli* hosts. Analyzed on Novex Tris–glycine SDS–PAGE or NuPAGE Novex Bis–Tris Gel, Coomassie staining. Arrows point to the target protein. (A) pET15b/Ssh10_wt non-fusion construct in different strains on 4–12% NuPAGE. Lanes: M—SeaBluePlus2 (Pre-Stained Protein Standard, Invitrogen); 1—(uninduced); 2—(induced) BL21(DE3)LysS; 3—(uninduced), 4—(induced), BL21(DE3); 5—(uninduced), 6—(induced), BL21star(DE3); 7—(uninduced), and 8—(induced), BL21-codonplus(DE3)-RIL. (B) pET15b/Ssh10_syn non-fusion construct in different strains (4–12% NuPAGE). Lanes: M—marker SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3)LysS; 3—(uninduced); 4—(induced), BL21(DE3); 5—(uninduced); 6—(induced), BL21star(DE3); 7—(uninduced); and 8—(induced), BL21-codonplus(DE3)-RIL. (C) GST-tagged Ssh10_wt on 8–16% SDS–PAGE. Lanes: M—SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3)LysS; 3—(uninduced); 4—(induced), BL21star(DE3); 5—(uninduced); and 6—(induced), BL21-codonplus(DE3)-RIL.

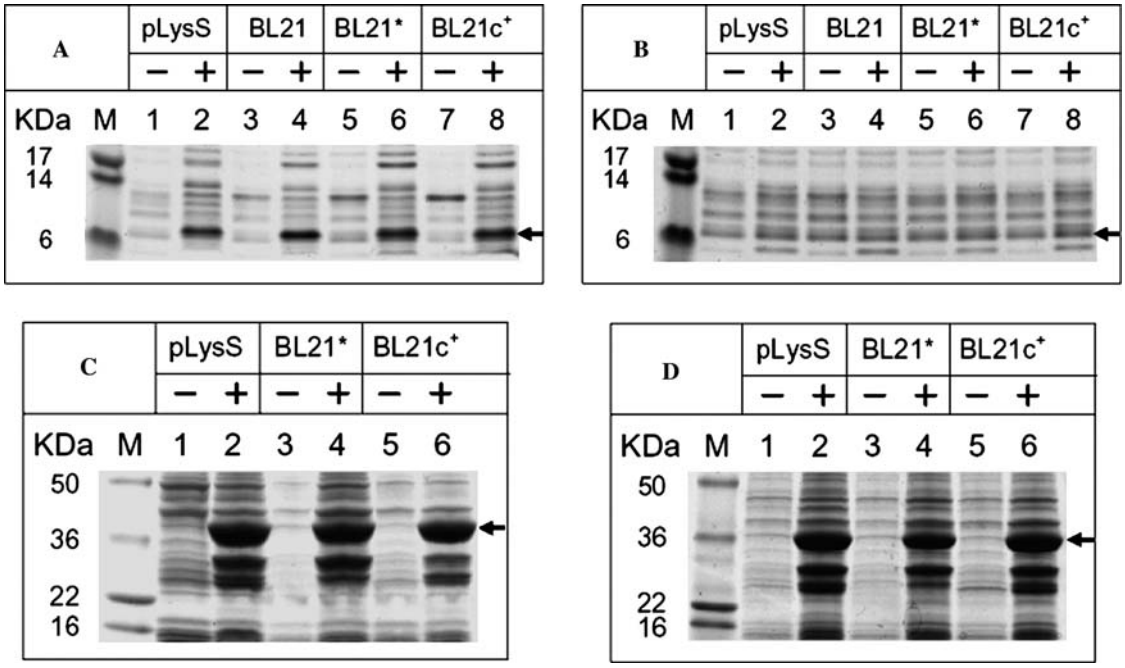

Fig. 2. Comparison of expression levels of wild-type (A), synthetic gene (B), and GST_wt (C) or GST_Syn (D) mtGrx constructs in different *E. coli* hosts. Analyzed on Novex Tris–glycine SDS–PAGE or NuPAGE Novex Bis–Tris Gel. Arrows point to the target protein. (A) pET15b/mtGrx non-fusion construct in different strains on a 4–12% NuPAGE. Lanes: M—SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3)LysS; 3—(uninduced); 4—(induced); BL21(DE3); 5—(uninduced); 6—(induced), BL21star(DE3); 7—(uninduced); and 8—(induced), BL21-codonplus(DE3)-RIL. (B) mtGrx_syn non-fusion construct in different strains analyzed on 4–12% NuPAGE. Lanes: M—SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3)LysS; 3—(uninduced); 4—(induced), BL21(DE3); 5—(uninduced); 6—(induced), BL21star(DE3); 7—(uninduced), and 8—(induced), BL21-codonplus(DE3)-RIL. (C) GST-tagged mtGrx_wt analyzed on 8–16% SDS–PAGE. Lanes: M—SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3)LysS; 3—(uninduced); 4—(induced), BL21star(DE3); 5—(uninduced); and 6—(induced), BL21-codonplus(DE3)-RIL). (D) GST-tagged synthetic Grx_H on 8–16% SDS–PAGE. Lanes: M—SeaBluePlus2; 1—(uninduced); 2—(induced), BL21(DE3); 3—(uninduced); 4—(induced), BL21star(DE3); 5—(uninduced); and 6—(induced), BL21-codonplus(DE3)-RIL.

Expression of wild-type genes in different host strain backgrounds

Ssh10wt and mtGrxwt constructs were expressed in BL21(DE3)LysS, BL21(DE3), BL21*, and BL21c$^+$ strains (Figs. 1A and 2A, Table 2). Whereas low levels of Ssh10wt were obtained in BL21(DE3)LysS (around 1 mg/culture), expression was significantly enhanced using strains BL21(DE3) (17.8 mg/L), BL21* (26 mg/L), and BL21c$^+$ (14 mg/L culture) indicating importance of RNA stability of the Ssh10wt gene for high-level expression. In contrast to Ssh10wt, no reduction in expression levels of mtGrx in the LysS strain was observed, and mtGrx levels were similar in all strains tested, ranging from 11 to 13 mg/L culture (Table 2).

Expression of synthetic genes

A replacement of the majority of rare codons in the Ssh10 and mtGrx genes led to opposite effects on protein

expression levels (Figs. 1B and 2B, Table 2). A Ssh10 construct, where all problematic rare codons were replaced by frequently used *E. coli* codons (Table 1), was expressed to a level of 5–6 mg/L in BL21pLysS, to 16 in BL21 (DE3), and 19 mg/L in BL21* and BL21c+. In contrast to the data obtained with a codon optimized Ssh10 construct, we surprisingly found no expression of the untagged synthetic mtGrx construct, in any of the employed bacterial strains and under no expression screening condition (Fig. 2B, Table 2).

Expression of Ssh10 and mtGRX fusion genes

Expression constructs of wild-type Ssh10 and mtGrx genes as C-terminal fusion to GST were subjected to expression trials in the same *E. coli* strains as the untagged variants (Figs. 1C and 2C and Table 2). Expression levels achieved were between 1.4 and 1.9 mg/L culture (calculated as the amount of Ssh or Grx protein obtained) for Ssh10, and between 3.5 and 3.8 mg/L culture for mtGrx. Expression of the codon-optimized, synthetic mtGrx gene as C-terminal fusion to GST led to expression levels of about 1.6–2.0 mg/L culture, opposed to the undetectable levels of the untagged version described above (Figs. 2B and D, Table 2).

Northern blot analysis of synthetic mtGrx expression

To further analyze the lack of protein expression observed with the synthetic mtGrx construct, total RNA isolated from expression trials of the different strains was analyzed by Northern blotting, hybridized with a DIG-labeled mtGrx probe (Fig. 3). A clear specific signal is obtained in the induced conditions, demonstrating that no transcriptional block is responsible for the lack of protein expressed.

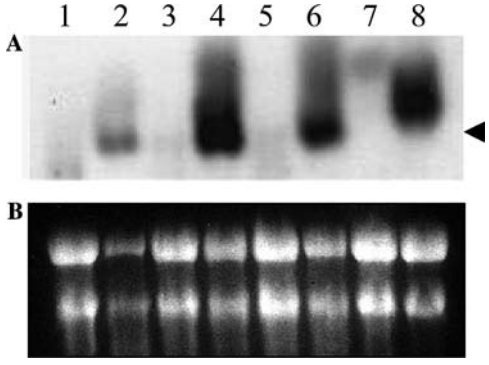

Fig. 3. Northern blot analysis of wild-type and synthetic mtGrx gene expression in different *E. coli* strains. Total RNA (20 μg) from different cultures was separated by denaturing agarose electrophoresis and expression was demonstrated by hybridization with a DIG-labeled mtGrx probe (A) an exposure time of 1 min. Lane 1, mtGrx_syn (uninduced), BL21(DE3)LysS; lane 2, mtGrx_syn (induced), BL21(DE3)LysS; lane 3, mtGrx_syn (uninduced), BL21star(DE3); lane 4, mtGrx_syn (induced), BL21star(DE3); lane 5, mtGrx_syn (uninduced), BL21-codonplus(DE3)-RIL; lane 6, mtGrx_syn (induced), BL21-codonplus(DE3)-RIL; lane 7, MtGrx_wt (uninduced), BL21-codonplus(DE3)-RIL; and lane 8, MtGrx_wt (induced), BL21-codonplus(DE3)-RIL. (B) Ethidium bromide staining of the blotted gel (A) to demonstrate loading and intact RNA.

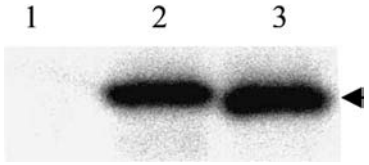

Fig. 4. In vitro translation of mtGrx_wt and mtGrx_syn using the T7 RNA polymerase based *E. coli* S30 system. Translation was carried out in the presence of [^{35}S]methionine, and analyzed by SDS–PAGE and autoradiography. Lane 1, pET15b vector, no insert; lane 2, pET15b/mtGrx_wt; and lane 3, pET15b/mtGrx_syn. The arrow indicates a mass of 9480 Da.

Coupled in vitro transcription and translation of synthetic and wild-type mtGrx

The untagged mtGrx-syn construct was used in an in vitro coupled transcription/translation reaction. The system was isolated from *E. coli*, driven by the T7 RNA polymerase for transcription, and the reaction with the mtGrx-syn Grx plasmid was compared to a translation reaction with a mtGrx-wt construct (Fig. 4). Specific ^{35}S-labeled Grx products are obtained, indicating successful transcription and translation with the in vitro system.

Protein analysis

Purified Ssh10 and mtGrx variant proteins were analyzed by N- and C-terminal sequence analysis, compositional amino acid analysis, MALDI-TOF, and ESI-TOF mass spectrometry to exclude erroneous translation in the rare codon gene products (Fig. 5). In all cases the sequences and masses experimentally obtained matched the theoretical values, indicating that translational misincorporations did not occur with Ssh10 and mtGrx genes under the conditions employed in our experimental systems. The folded nature of the expressed and purified proteins was assessed using circular dichroism spectroscopy. The absence of a minimum at 190 nm and presence of minimum in the 205–225 nm region are strong evidence of considerable secondary structure content which we interpret to be consistent with a folded native-like protein. It must be emphasized that due to the lack of functional assays for these proteins secondary structure determination is the strongest indication that these proteins are native-like.

Discussion

We have analyzed the heterologous expression characteristics in *E. coli* of two small thermostable proteins

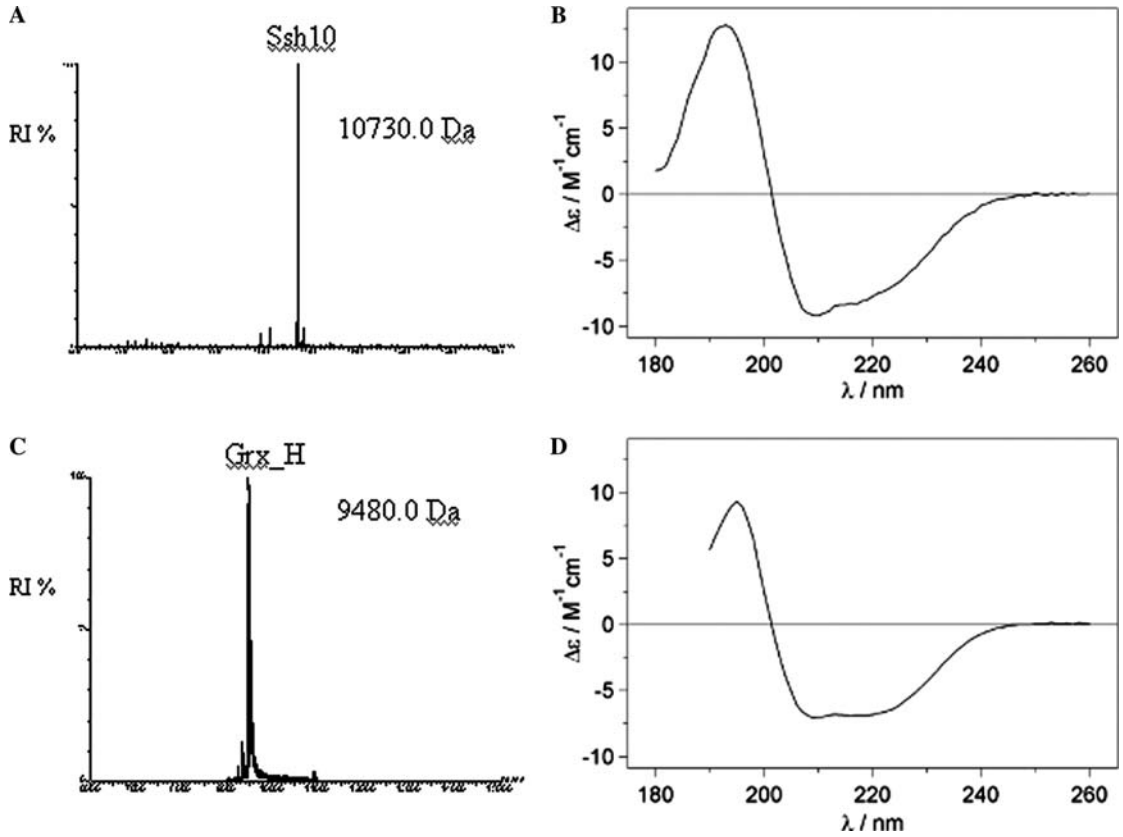

Fig. 5. Structural analysis of recombinant Ssh10 and mtGrx. (A,C) ESI-TOF spectra of recombinant Ssh10 (A) and mtGrx (C) demonstrating coinciding mass values for theoretical and experimentally determined proteins (Masses: $Ssh10_{exp}$: 10730.0 Da, $Ssh10_{theor}$: 10729.5 Da; Grx_{exp}: 9480.0 Da, and GRx_{theor}: 9480.9 Da). (B,D) Circular dichroism spectra of purified Ssh10 in H_2O and Grx_H in 50 mM phosphate, 50 μM EDTA at pH 7.0 at 25 °C demonstrating ordered secondary structures.

derived from archaeal organisms. The results obtained highlight important issues in selections of strategies for heterologous expression of foreign genes. First, the amount of rare codons does not necessarily correlate with expression levels obtained and the existence of rare codons does not exclude *E. coli* as host for rare codon gene expressions per se. This apparently contradicts other observations that levels of rare Arg codons (AGA and AGG) in a gene are negatively correlated with expression levels in *E. coli* [17,18]. It was reported that levels of 3% AGG/AGA codons can severely impede expression levels [19], and occurrence of such codons in tandem can lead to frame-shifting, premature termination, and lowered fidelity of translation [11]. This is in contrast to the data obtained with Ssh10 and mtGrx, showing that wild-type expression levels were already high in BL21(DE3). Since mtGrx and Ssh10 contain significant amounts of rare AGA/AGG/CGG codons, which constitute 100% of the total Arg codons in both genes, it is concluded that high frequency of a particular rare codon is not sufficient to prevent high level expression in these two genes. This demonstrates that other factors affect protein expression, e.g., localization of rare codons within the gene, and this conclusion is paralleled by observations in other studies [20–22]. Interestingly, our data obtained on Ssh10 expression in BL21(DE3) and the LysS derivative show that the commonly used LysS suppressor phenotype might interfere with high level expression. We can exclude an interaction on the protein level since the protein expressed from the synthetic gene shows good expression levels. We conclude that the reduced transcriptional rate obtained with the LysS phenotype results in a decreased protein level. This is a well-known phenomenon, since in bacteria transcription and translation are tightly coupled. However, in the Ssh10 case differential RNA stability between the wild-type gene and the synthetic gene might play a central role, and thus link the different protein levels obtained with the two different constructs (wild-type vs synthetic in LysS). Interestingly, a clear correlation between decreased RNA stability and amount of rare codons has been established in yeast [23]. We therefore interpret our results in such a way that decreased RNA stability of rare codon genes is a major contributor to lowered expression levels.

Second, to compare expression strategies and possibly enhance expression levels for rare codon genes, we constructed synthetic genes of Ssh10 and mtGrx by

substituting the majority of rare codons with those frequently used by *E. coli*. Numerous examples of increased protein expression through altered codon usage are found in the literature, making this synthetic gene approach a common strategy for protein expression [24–26]. Despite the reported advantages in raising expression levels, we found the synthetic gene synthesis a work- and cost-extensive process, suggesting other alternative methods for high-throughput studies.

As indicated above, we did not find significant differences between wt and syn levels for Ssh10 in the non-LysS background. However, we realized to our surprise that codon optimization of the mtGrx gene led to undetectable expression levels. This effect has been observed in few other instances [27,28] and could be attributed to the propensity of the codon-optimized mRNA to form stable 5′ structures, which is compatible with our observations that N-terminal fusion of GST to the synthetic Grx gene led to good expression levels for the GST/pGEX vector series. However, calculations of the theoretical ΔG values for the synthetic and wild-type mtGrx yield −62.3 and −68.1 kcal/mol, respectively. Inspection of secondary structure predictions suggests that increased stability of the mRNA is not the determining factor for the observed differential expression. We suspect that specific secondary structures formed in vivo inhibit the translational process, in agreement with observations that rare codons located at the 3′ end can increase protein levels possibly due to specific secondary structure formation [29]. A further possible explanation might be the relationship between the folding efficiency of nascent proteins and the translational speed. Lowered translational rate might allow folding to take place, and substitution of the rare codons can then decrease the expression yield. It has been noted also that codon bias and expression problems are less prominent in longer genes than in shorter ones [30]. Independent of the underlying causes for the observed suppression of synthetic mtGrx expression, fusion of a 5′ sequence or in vitro translation is now found to be the means to circumvent the translational block observed in vivo with the synthetic mtGrx gene. This observation also underscores the robustness of systems utilizing fusion domains to enhance levels of heterologous protein. Besides increasing solubility and providing affinity tags for rapid purification or interaction applications, the fusion with an additional gene could possibly be advantageous to produce favorable secondary structures of a gene allowing increased protein production. Lower levels of heterologous proteins were obtained as fusion protein compared to wild-type levels, due to the portion of fusion protein produced in addition to the target protein. However, total protein levels are usually lower with vectors of the pGEX series containing tac promoters than with the T7 promoter based pET vectors. The point of reduced overall protein yield is especially important when in vivo labeling with costly stable isotopes is performed for subsequent NMR studies, and thus needs to be balanced against the obvious advantages of a fusion system, i.e., its robustness and ease of protein purification.

Taken together, these studies provide important examples of unexpected rare codon gene expression patterns and solutions to enhance protein production, useful for designing expression strategies in structural genomics initiatives. Whereas expression trials of wild-type, rare codon genes in various modified *E. coli* strains are the recommended first choice, a fusion gene approach is a considerable robust alternative preventing problems encountered by rare codon or specific RNA secondary structure formation. In this respect, the costly and tedious synthesis of synthetic genes appears to be the least desirable alternative for production of heterologous proteins.

Acknowledgments

Expert assistance by Monica Lindh, Ella Cederlund, Marie Ståhlberg, Irene Byman, and Maria Gustafson is gratefully acknowledged. This study was supported by the Swedish Science Foundation and Karolinska Institutet.

References

[1] J.R. Swartz, Curr. Opin. Biotechnol. 12 (2001) 195–201.
[2] G. Hannig, S.C. Makrides, Trends Biotechnol. 16 (1998) 54–60.
[3] S.C. Makrides, Microbiol. Rev. 60 (1996) 512–538.
[4] F. Baneyx, Curr. Opin. Biotechnol. 10 (1999) 411–421.
[5] J.F. Kane, Curr. Opin. Biotechnol. 6 (1995) 494–500.
[6] D.E. McNulty, B.A. Claffee, M.J. Huddleston, J.F. Kane, Protein Expr. Purif. 27, (2003) 365–374.
[7] T.L. Calderone, R.D. Stevens, T.G. Oas, J. Mol. Biol. 262 (1996) 407–412.
[8] B.Q. Wang, L. Lei, Z.F. Burton, Protein Expr. Purif. 5 (1994) 476–485.
[9] S.L. Martin, B. Vrhovski, A.S. Weiss, Gene 154 (1995) 159–166.
[10] E. Goldman, A.H. Rosenberg, G. Zubay, F.W. Studier, J. Mol. Biol. 245 (1995) 467–473.
[11] R.A. Spanjaard, J. van Duin, Proc. Natl. Acad. Sci. USA 85 (1988) 7967–7971.
[12] J. Sipley, E. Goldman, Proc. Natl. Acad. Sci. USA 90 (1993) 2315–2319.
[13] X. Wu, U. Oppermann, J. Chromatogr. B 786 (2003) 177–185.
[14] P. Forterre, F. Confalonieri, S. Knapp, Mol. Microbiol. 32 (1999) 669–670.
[15] S.C. McFarlan, C.A. Terrell, H.P. Hogenkamp, J. Biol. Chem. 267 (1992) 10561–10569.
[16] W.P. Stemmer, A. Crameri, K.D. Ha, T.M. Brennan, H.L. Heyneker, Gene 164 (1995) 49–53.
[17] U. Brinkmann, R.E. Mattes, P. Buckel, Gene 85 (1989) 109–114.
[18] T. Kleber-Janke, W.M. Becker, Protein Expr. Purif. 19 (2000) 419–424.
[19] M.J. Rogers, Parasitol. Today 12 (1996) 124.
[20] M.F. Alexeyev, H.H. Winkler, Biochim. Biophys. Acta 1419 (1999) 299–306.

[21] B. Andrews, H. Adari, G. Hannig, E. Lahue, M. Gosselin, S. Martin, A. Ahmed, P.J. Ford, E.G. Hayman, S.C. Makrides, Gene 182 (1996) 101–109.
[22] S.E. Gerchman, V. Graziano, V. Ramakrishnan, Protein Expr. Purif. 5 (1994) 242–251.
[23] D. Herrick, R. Parker, A. Jacobson, Mol. Cell. Biol. 10 (1990) 2269–2284.
[24] W. Pan, E. Ravot, R. Tolle, R. Frank, R. Mosbach, I. Turbachova, H. Bujard, Nucleic Acids Res. 27 (1999) 1094–1103.
[25] P. Prapunwattana, W. Sirawaraporn, Y. Yuthavong, D.V. Santi, Mol. Biochem. Parasitol. 83 (1996) 93–106.
[26] D.R. Casimiro, P.E. Wright, H.J. Dyson, Structure 5 (1997) 1407–1412.
[27] L.O. Nilsson, B. Mannervik, Biochim. Biophys. Acta 1528 (2001) 101–106.
[28] K.E. Griswold, N.A. Mahmood, B.L. Iverson, G. Georgiou, Protein Expr. Purif. 27 (2003) 134–142.
[29] Y.G. Gursky, R. Beabealashvilli, Gene 148 (1994) 15–21.
[30] L. Duret, D. Mouchiroud, Proc. Natl. Acad. Sci. USA 96 (1999) 4482–4487.
[31] K. Wada, Y. Wada, F. Ishibashi, T. Gojobori, T. Ikemura, Nucleic Acids Res. 20 (Suppl) (1992) 2111–2118.