Original Article

Heterologous Proteins Production in *Escherichia coli*: An Investigation on the Effect of Codon Usage and Expression Host Optimization

Hasan Mirzahoseini, PhD.*, Samaneh Mafakheri, M.Sc., Nafiseh Soltan Mohammadi, M.Sc., Somayeh Enayati, M.Sc., Nahid Mortazavidehkordi, B.Sc.

Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran

** Corresponding Address: P.O. Box: 1316943551, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran*
Email: mirzahoseini@pasteur.ac.ir

Received: 24/Oct/2009, Accepted: 3/May/2010

Abstract

Objective: The production of heterologous proteins in *Escherichia coli* is strongly affected by codon bias. This phenomenon occurs when the codon usage of mRNA coding for the foreign protein differs from that of the bacterium. The ribosome pauses upon encountering a rare codon and may detach from mRNA, thereby the yield of recombinant protein production reduces. The aim of this study is to investigate the effect of these codon numbers reductions on the recombinant protein production.

Materials and Methods: Since most amino acids are encoded by more than one codon, codons were changed in order to their usage in a special host such as *E. coli* without any transformation in amino acids sequence. Silent mutations in 5' codons of human basic fibroblast growth factor cDNA carried out by site-directed mutagenesis and the expression level of the recombinant protein is analyzed by means of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot.

Results: Expression level in mutant and wild-type genes indicated a considerable difference. In contrast with the remarkable bands of wild-type gene in all the strains particularly in codon plus strain, there were no significant bands related to mutant gene in SDS-PAGE analysis.

Conclusion: Because of the same conditions of mutant and wild-type genes during the translation and transcription, this significant difference may relate to mRNA efficiency for translation. Our results indicate that increased stability of 5' mRNA secondary structures in *E. coli* prevents efficient translation initiation. Furthermore, wild-type gene significant bands in codon plus strain support the hypothesis that the possible elimination of translational pauses that increase translation rate leads to over expression.

Keywords: Codon, Usage, Gene Expression, hbFGF

Cell Journal(Yakhteh), Vol 12, No 4, Winter 2011, Pages: 453-458

Introduction

Escherichia coli is the system of choice for over expressing of heterologous proteins (1). As a host, this bacterium has numerous advantages, including: inexpensive culture conditions, very well known genetic background, easy manipulation and amenability to high density fermentation procedures (1-4). Although *E. coli* has a remarkable capacity to produce large quantities of proteins, recombinant proteins are often expressed in low amount when the codon usage of the mRNA encoding the foreign protein differs from that of the *E. coli* host (4, 5). This problem can arise due to product toxicity, mRNA instability, lack of posttranslational modification, saturation of the folding machineries of the host cell and cofactors deficiency (6). In addition, depletion of low-abundance tRNAs occurs if the foreign mRNA contains many codons that are rare in *E. coli*. This deficiency may lead to amino acid mis-incorporation and/or polypeptide truncation following by the effect on the heterologous protein expression levels and quality (7). Two alternative strategies are utilized to remedy codon bias. One approach is site-directed mutagenesis of the target sequence for the generation of codons reflecting tRNA pool in the host system (8, 9) and another is the co-transformation of the host with a plasmid harbouring a gene encoding tRNA cognate to the problematic codons (10). While codon optimization is a cumbersome and expensive process, modifying host availability of the rare tRNAs is an easier approach. This methodology guides to the commercialization of bacterial strains carrying plasmids containing extra copies of problematic tRNAs genes.

The aim of this work is to analyze the expression of codon optimized the codons of human basic fibroblast growth factor (*hbfgf*) cDNA in *E. coli* BL21(DE3)-pLysS strain (BL) commonly used for

protein expression, a codon bias-adjusted strain, BL21-Codon Plus (DE3)-RIL strain and OrigamiB (DE3) to evaluate both codon optimization and host modification. As the secondary structure at the translation initiation region (TIR) of the mRNA that plays a crucial role in the efficiency of gene expression (11) site directed mutagenesis is applied in this region by a long primer to study the affect of codon usage content of the mentioned cDNA.

Materials and Methods

Primer designing and site directed mutagenesis by polymarase chain reaction (PCR)

A long primer was designed to modify 5' end of *hbfgf* cDNA by site directed mutagenesis PCR using a wild type *hbfgf* cDNA as a template which was a gift from Dr. Seno (Table 1b). Moreover, this project was approved by Pasteur Institute of Iran. The mutant *hbfgf* cDNA encoding a variant of 18 kilo Dalton of the corresponding protein which was amplified during PCR containted 11 codons changes in favor of the codon usage of the host bacteria (*E. coli*). Changed codons were selected based on *E. coli* codon usage table from San Diego State University (http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/in-vitro-genetics/codon-usage.html). mRNA secondary structure and total free energy of each wild type and mutant *hbfgf* cDNAs were predicted by Vienna RNA Web Servers. The designed upstream (forward) mutagenic primer, the downstream (reverse) primer and the wild-type gene sequences are shown in table 1. For *in vitro* amplification, Gradient PCR was used in order to determine the optimum annealing temperature at a range of 59°C to 70.5°C.

Subcloning of mutant hbfgf

PCR product was cloned into pGEM-T easy vector (Promega, USA) and transformed into Top10f cells (Novagen, UK) for reproduction. α-complementation phenomenon (blue/white screening assay) was used for isolation of cells containing cloned cDNA (12). The white colonies containing our desire mutant cDNA were chosen among the blue colonies and their plasmids were extracted using plasmid purification kit (Qiagen, Germany). Mutant *hbfgf* cDNA was isolated from T-vector by *Nde*I (Fermentas, Lithuania) and *Bgl*II (Fermentas, Lithuania) digestion and ligated into *Nde*I and *Bam*HI sites of pET-22b expression vector (Novagen, UK).

Escherichia coli strains

Escherichia coli strains comprised Top10f (Novagen, UK), BL21 (DE3)-PlysS (Novagen, UK), BL21CodonPlus (DE3)-RIL (Stratagene, USA) and OrigamiB (DE3) (Novagen, UK) were selected for reproduction of new construct and expression of hbFGF.

Expression and identification

At the first stage, the new construct (named pET-1010) was transformed into Top10 host cells for reproduction. Afterwards, the new construct was isolated and transformed into the BL21 (DE3) pLysS, BL21-Codon Plus (DE3)-RIL and OrigamiB (DE3) for expression. Transformants were grown on Luria Bertani (LB) agar plates containing 50 μg/ml of ampicillin. The cells only bearing the expressed plasmid were inoculated into LB medium containing related antibiotic (50 μg/ml of ampicillin) for plasmid selection. When the absorbance value at 600 nm reached to 0.5, recombinant protein production was induced by adding 1 mM isopropy1 β-D-1-thiogalactopyranoside (IPTG) to the culture media for approximately 5 hours. Subsequently, the cells were collected by centrifugation and suspended in 250 μl of 50 Mm Tris-HCl buffer (pH=7.2). The suspension was kept on ice, then cells were sonicated (3 times, 30 seconds each time) and centrifugated for seperateing the cell proteins from the other cell particles.

Table 1: Wild-type gene, forward mutagenic primer and reverse primer sequences

a) Wild-type Gene Sequence
5’ATG CCC GCC TTG CCC GAG GAT GGT GGA TCG GGC GCC ATA CCG CCC GGC CAC TTC 3’ (ATG: Start codon)
b) Forward mutagenic primer sequence with *NdeI* Restriction Enzyme Site
5’ TACATATG CCG GCG CTG CCG GAA GAC GGT GGT TCT GGT GCG ATC CCG CCC GGC CAC TTC 3’ (CATATG: *NdeI*; ATG: Start codon)
c) Reverse primer sequence with *Bgl* II Restriction Enzyme Site
5'T ACT ATT AGA TCT TGG CCA TTA AAA TCA GC 3' (AGA TCT: *BGl* II)

***The forward primer is flanked by NdeI site and the reverse primer is flanked by BglII site.**

SDS-PAGE using 15% acrylamide was applied to analyze the induced as well as uninduced lysate cells based on the method of laemmli (13). In addition, the samples were boiled with loading buffer including 2-Mercaptoethanol as a detergent and EDTA as a protease inhibitor. Identification of hbFGF was based on SDS-PAGE assay followed by the densitometer scanning; therefore, we used Quantity-One 1D-Analysis software (Bio-Rad, USA).

Western blotting assay was utilized in order to confirm hbFGF protein band. Briefly, the proteins separated on the gel were transformed onto a 0.2 µm pore size nitrocellulose membrane (schlecher & Schuell, USA). Rabbit anti-hbFGF anti serum (Sigma, USA) were served as the primary antibody and secondary antibody, horse radish peroxidase conjugated goat-anti-rabbit antibody (Sigma, USA) was added, then followed by incubation at room temperature for two hours. The immunoreaction was visualized using 0.5 mg/ml diaminobenzidine HCl and 0.01% H_2O_2.

Results

Construction of pET-1010

The PCR products were analyzed on 1.5 % agarose gel stained with ethidium bromide. The expected size of the target amplification fragment was 450 bp (Fig 1).

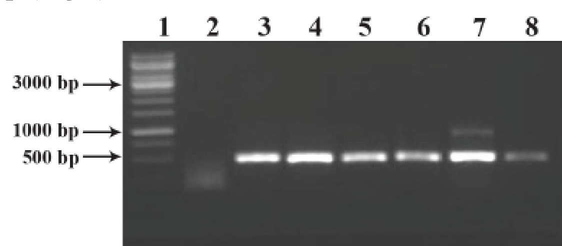

Fig 1: Mutant hbfgf amplification by gradient PCR, Lane 1: 1 kb DNA ladder (Fermentas, Lithuania), Lane 2: Blank sample, Lanes 3 to 8, gradient PCR, temperatures from 59°C to 70.5°C.

The band corresponded to 450 bp was purified from agarose gel and cloned into the pGEM-T Easy vector. Mutant *hbfgf* cDNA was isolated from T-vector by *Nde*I and *Bgl*II digestion and ligated into *Nde*I and *Bam*HI sites of pET-22b expression vector. The correct insertion of the new construct was verified by the restriction digest and plasmid sequencing. Briefly, the new construct (pET-1010) was analyzed by digestion with two different restriction enzymes, *Hind*III and *Bgl*II. If the ligation process was successful, we should have a 600 bp band in agarose gel according to pET-22b restriction sites (Fig 2), otherwise, we were able to see a 200 bp band. Moreover, Nucleotide sequence of the mutant *hbfgf* was determined by Gene Fanavaran Company (Tehran, Iran).

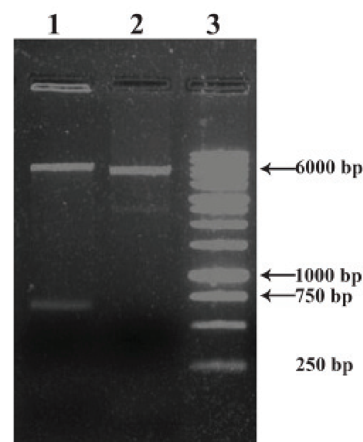

Fig 2: Gel electrophoretic analysis of pET-1010 in compare with pET-22b for confirming the correct ligation of mutant hbfgf cDNA in pET-22b vector. Lane 1: pET-2010 is digested with Hind III & Bgl II, Lane 2: pET22b is digested with Hind III, Lane 3: 1 kb DNA Ladder (Fermentas, Lithuania). The 600 bp band in lane 1 shows the right cloning of hbfgf cDNA in pET-22b vector.

Expression of wild type and mutant hbfgf genes

pET-1010 was transformed into three *E. coli* strains: BL21(DE3) PlySs, BL21-CodonPlus (DE3)-RIL and OrigamiB(DE3), for expression. Cells containing the expressed construct were grown and induced by IPTG. Comparing the result from SDS-PAGE to the wild-type gene revealed that a codon of the optimized gene construct displayed undetectable levels of the protein production (Fig 3, Lanes 3, 5 and 7).

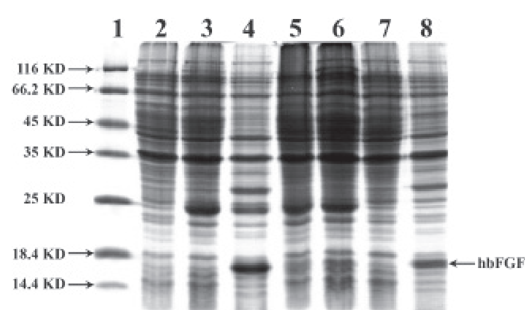

Fig 3: Comparison of expression levels of wild-type and mutant genes in three different strains of E. coli. The arrows in the left show molecular weight. Lane 1: Protein marker, Lane 2: OrigamiB(DE3) cell without plasmid, Lane 3: induced BL21CodonPlus(DE3)-RIL cell containg pET-1010, Lane 4: induced BL21CodonPlus(DE3)-RIL cell containg pET-22b-wild type gene, Lane 5: induced BL21(DE3)PlySs cell containg pET-1010, Lane 6: induced BL21(DE3)PlySs cell containg pET-22b-wild type gene, Lane 7: induced OrigamiB (DE3) cell containing pET-1010, Lane 8: induced OrigamiB (DE3) cell containing pET-22b-wild type gene.

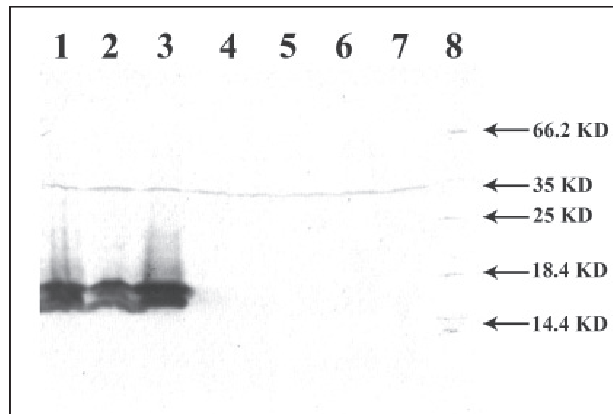

Fig 4: Western Blotting Analysis. Lane 1: induced Origami B (DE3) cells containing pET-22-b-wild type gene, Lane 2: induced BL21Codon Plus (DE3)-RIL cells containg pET-22b-wild type gene, Lane 3: induced BL21 (DE3) PlySs cells containg pET-22b-wild type gene, Lane 4: induced BL21Codon Plus (DE3)-RIL cells containg pET-1010, Lane 5: induced BL21 (DE3) PlySs cells containg pET-1010, Lane 6: induced Origami B (DE3) cells containing pET-1010, Lane 7: Origami B (DE3) cells without plasmid, Lane 8: Protein marker.

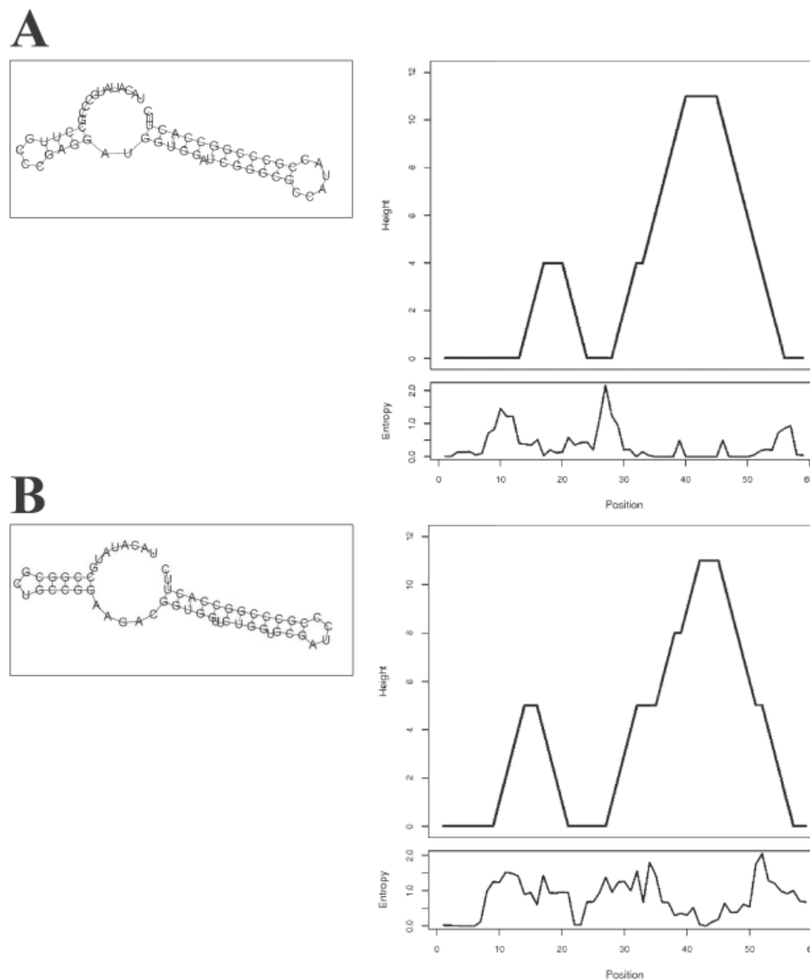

Fig 5: Illustration of RNA secondary structures. The theoretical ∆G values of RNA secondary structures at 5'end of wild type (A) and mutant hbfgf (B) were calculated by Vienna RNA Secondary Structure Prediction program (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) with thermodynamic algorithm and illustrated.

Furthermore, wild-type *hbfgf* cDNA was over expressed in the codon bias-adjusted strain and western blotting analysis that confirmed the obtained results from SDS-PAGE (Fig 4).

mRNA secondary structure and its total free energy

By inspecting the sequences of wild type and mutant *hbfgf* cDNA with Vienna online program, possible secondary structures were predicted (Fig 5). The total free energy at the 5' end of the mRNA was increased from -23.40 kcal/mol to -22.80 kcal/mol, implying that unstable RNA secondary structure at 5' and might cause mRNA inefficiency for translation.

Discussion

Many studies have questioned the importance of codon usage of 5' end of genes according to host cells for the initiation of translation (14-16). Matching the codon usage of recombinant genes to that of the expression host is a common strategy for increasing the expression of heterologous proteins in bacteria and it is always employed secondary structure of mRNA (17,18). However, we found that the altering codons to those preferred by *E. coli* led to significantly lower expression compared to the wild-type *hbfgf* gene, despite the presence of several rare *E. coli* codons in the *hbfgf* sequence.

Evidence indicates that these strategies do not necessarily result in the same degree of enhancement for different heterologous protein expression. In the some studies, modification of TIR according to the codon usage of host revealed higher level of expression (16, 19), but in the other experiments showed an opposite effect on the expression level (20). In particular, our results indicate that the increased stability of 5′ mRNA secondary structures in *E. coli* optimized transcript prevents efficient translation. In fact, ribosomes do not recognize mRNA secondary structures and only bind to single-stranded RNA. Hence, secondary structure of mRNA is a key factor in determining the efficiency of translation initiation in prokaryotes, in other words, translation efficiency is directly determined by the availability of unfolded TIR (21).

In addition, in spite of the fact that codon usage optimization of the 5’end of mRNA has been applied to improve the efficiency of human protein production in *E. coli*; high level expression of human protein in the mentioned-bacteria is still a challenge that virtually depends upon each individual target genes (22). A possible explanation is that the effects vary based on the identity, density and location of the rare codons; in addition, it may involve complex effects on mRNA structure and its stability.

Conclusion

Although we tried to increase the abundance of tRNA with changing in codon usage in human basic fibroblast growth factor, the results underscore the importance of RNA stability, but not necessarily tRNA abundance for efficient heterologous protein production in *E. coli*. These results emphasize the idea that potential 5′ mRNA secondary structures should be considered along with codon usage when designing a synthetic gene for high level expression in *E. coli*.

Acknowledgments

The financial support from Pasteur Institute of Iran is gratefully appreciated. There is no conflict of interest in this article.

References

1. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2006; 72: 211-222.
2. Sorensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in Escherichia coli. J Biotechnol. 2005; 115: 113-128.
3. Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem. 2008; 307:249-264.
4. Makrides SC. Strategies for achieving high-level expression of genes in Escherichia coli. Microbial Rev. 1996; 60: 512-538.
5. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli, Curr Opin Biotechnol. 1995; 6: 494-500.
6. Idicula-Thomas S, Balaji PV. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. Protein Sci. 2005; 14: 582-592.
7. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004; 22: 346-353.
8. Calderone TL, Stevens RD, Oas TG. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. J Mol Biol. 1996; 262: 407-412.
9. Kane JF, Violand BN, Curran DF, Staten NR, Duffin KL, Bogosian G. Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of Escherichia coli. Nucleic Acids Res. 1992; 20: 6707-6712.
10. Dieci G, Bottarelli L, Ballabeni A, Ottonello S. tRNA-assisted overproduction of eukaryotic ribosomal proteins. Protein Expr Purif. 2000; 18: 346-354.
11. Hall MN, Gabay J, Debarbouille M, Schwarts M.A role for mRNA secondary structure in the control of translation initiation. Nature.1982; 295: 616-618.
12. Primrose SB, Twyman RM, Old RW. Principle of gene manipulation. 6th edition. London, UK. Blackwell publishing; 2001; 35.
13. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;

227: 680-685.
14. Wu X, Jornvall H, Berndt KD, Oppermann U. 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. Biochem Biophys Res Commun. 2004; 313: 89-96.
15. Kim S, Lee SB. Rare codon clusters at 5′-end influence heterologous expression of archaeal gene in Escherichia coli. Protein Expr Purif. 2006; 50: 49-57.
16. Makoff AJ, Oxer MD, Romanos MA, Fairweather NF, Ballantine S. Expression of tetanus toxin fragment C in E.coli: high level expression by removing rare codons. Nucleic acid res. 1989; 17: 10191-10202.
17. Zhang W, Xiao W, Wei H, Zhang J, Tian Z. mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli. Biochem Biophys Res Commun. 2006; 349: 69-78.
18. Stenström CM, Jin H, Major LL, Tate WP, Isaksson LA.Codon bias at the 3’ side of the initiation codon is correlated with translation initiation efficiency in Escherichia coli. Gene. 2001; 263: 273-284.
19. Jiang X, Li S, Zhou A, Li F, Xu X, Zhu D. Translation initiation region plays an important role in the expression of human thrombopoietin in Escherichia coli. Biochem Mol Biol Int. 1996; 39: 1109-1113.
20. Griswold KE, Mahmood NA, Iverson BL, Georgiou G. Effects of codon usage versus putative 5'-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm. Protein Expr Purif. 2003; 27: 134-142.
21. De Smit MH, van Duin J. Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. Proc Natl Acad Sci USA. 1990; 87(19): 7668-7672.
22. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O.Codon optimization can improve expression of human genes in Escherichia coli: A multigene study. Protein Expr Purif. 2008; 59: 94-102.