Protein Expression and Purification xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

The effect of rare codons following the ATG start codon on expression of human granulocyte-colony stimulating factor in *Escherichia coli*

Zeinab Karimi, Navid Nezafat, Manica Negahdaripour, Shiva Hemmati, Younes Ghasemi *

Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
Pharmaceutical Sciences Center, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Article history:
Received 6 May 2015
and in revised form 27 May 2015
Accepted 29 May 2015
Available online xxxx

Keywords:
Granulocyte-colony stimulating factor
Rare codon
Recombinant protein

ABSTRACT

Presence of the rare codons resulted from the difference in codon usages among organisms is considered as an obstacle to heterologous gene expression. This is especially important for the expression of the genes with eukaryotic origin in *Escherichia coli*. The N-terminus of human granulocyte colony stimulating factor (hG-CSF) contains amino acids whose coding sequences belong to the rare codons in *E. coli*. In this study, the effect of rare codons on hG-CSF expression level was evaluated through introducing silent mutations in the 5′-end of the coding sequence. *E. coli* BL21 (DE3) was used as an expression host. The constructs with the rare codons at the positions following the ATG initiation site of hG-CSF elevated the expression level up to 53–56% of the total cell proteins. This effect may be explained either by the rare codons effects on the early elongation region to reduce ribosome traffic jams in the rest of transcript or by their impacts on reduction of GC content at the beginning region. Mfold RNA server and prediction of the 5′ mRNA secondary structure showed the less stable mRNA secondary structure is, the more hG-CSF expression level would be. However, the minimum free energy of the secondary structure individually, could not indicate this correlation between all constructs. This finding seems empirically important in designing the synthetic genes for production of the recombinant protein in *E. coli*.

1. Introduction

Although *Escherichia coli* is the most commonly used expression host for recombinant proteins and biopharmaceuticals, there are numerous obstacles in its heterologous expressions, especially when there is a difference in their codon usage from that of *E. coli* [1]. A set of codons including AGG, AGA, CGG and CGA (arginine, R), CUA and CUC (leucine, L), AUA (isoleucine, I), CCC, CCU and CCA (proline, P), UGU and UGC (cysteine, C), ACA (threonine, T), UCA, AGU, UCG and UCC (serine, S), GGA and GGG (glycine, G) have been defined as rare codons in *E. coli* codon usage [2–6]. However, roles of rare codons in *E. coli* genome are not completely clear; the presence of these codons specially at 5′ end of heterologous messenger may lead to low expression level and translational errors such as amino acid substitutions, frame shifts or premature translational termination arising from ribosomal stalling at positions of codons related to low frequency cognate tRNA [7–11]. On the other hand, Kudla et al. in their investigation on the impact of synonymous codon usage on gene expression, emphasized on the stability of mRNA folding near the ribosomal binding site rather than codon bias on protein levels of individual genes in *E. coli* [12]. Thermodynamically stable mRNA secondary structure near the start codon is a barrier that the ribosomal system must overcome to form a stable initiation complex and initiate translation efficiently [13–15].

The hG-CSF responsible for proliferation, survival and differentiation of neutrophils plays an important role in treating congenital or acquired neutropenias [16]. A mature hG-CSF is a 174 amino acid glycoprotein which is secreted from producing cells in the body after cleavage of 30 amino acids signal peptide from N-terminus of a 204 amino acids precursor. In order to achieve the proper expression of the recombinant protein in *E. coli*, 174 amino acids coding sequence should be inserted into the plasmid vector, and transformed to *E. coli* competent cells [17]. Although the final product has an additional N-terminal methionine without

Abbreviations: CAI, codon adaptation index (an index of codon usage bias towards the codon usage of highly expressed genes); hG-CSF, human granulocyte colony stimulating factor; MFE, minimum free energy; rhG-CSF, recombinant hG-CSF; SD, Shine-Dalgarno.

* Corresponding author at: Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, P.O. Box 71345-1583, Shiraz, Iran.
E-mail address: ghasemiy@sums.ac.ir (Y. Ghasemi).

http://dx.doi.org/10.1016/j.pep.2015.05.017
1046-5928/

any glycosylation, it shows proper biological activity. After removing the signal peptide, the TPLG residues are appeared in N-terminal region immediately downstream of the starting methionine. A majority of triplets coding for these residues are located in the rare codon table of *E. coli*. So the aim of our investigation was to understand which 5′ coding sequence is better to achieve a high level of recombinant hG-CSF expression in *E. coli*. What is the effect of low frequency codons in this region?

2. Materials and methods

2.1. Construction of recombinant plasmids

The hG-CSF coding sequence was optimized according to *E. coli* codon usage by using optimizer web servers, JCat and GenScript rare codon analysis tool [18,19]. The synthesized sequence was received from Cinnagen Co. (Tehran, Iran) in pGH cloning vector (Generay Biotech Co., Ltd. China) in the restriction site of *Sma*I. Plasmid pET-21a (Novagen, USA), with an IPTG inducible T7 promoter was used in this study to construct pET21a/hG-CSF recombinant plasmid. Four recombinant plasmids with the difference in the 5′ region of hG-CSF gene were constructed. Four forward primers, zG, kG, dG, pG, with different GC content and CAI but synonymous T, P, L and G triplets were used to make the silent mutagenesis immediately downstream of ATG (Table 1). In this study two forward primers were designed, and the others were synthesized with the nucleic acid sequences mentioned in the previous studies [20,21]. Polymerase chain reaction (PCR) was used to generate mutant forms of hG-CSF gene. Every 20 µl PCR reaction mixture contained 0.6 U *Pfu* DNA polymerase (Fermentas, Lithuania), 1× *Pfu* Buffer with $MgSO_4$, 200 µM of each dNTP and 20 pmol of each forward and reverse primers. The plasmid pGH-synthetic hG-CSF was used as DNA template and the reverse oligonucleotide was 5′-CGGGATCCATTACGGCTGAGCCAGATG-3′ with *Bam*HI restriction site at 5′ region. Amplifications were programmed in 25 cycles including 1 min at 94 °C, 1 min at 65 °C and 1 min at 72 °C. The final extension was done for 10 min at 72 °C. The hG-CSF amplicons were cloned in pET21a between *Nde*I and *Bam*HI restriction sites by using T4 DNA ligase (Fermentas, Lithuania). Cloning and transformation procedures were performed according to the standard protocols [22]. The *E. coli* TOP10F′ competent cells were transformed with the recombinant vectors and screened on Luria–Bertani (LB) agar supplemented with 100 µg/ml ampicillin as antibiotic marker. Nucleotide sequencing of the constructs was carried out using T7-promoter and T7-terminator primers to verify the correctness of silent mutations.

2.2. Expression and verification of rhG-CSF

E. coli BL21 (DE3) competent cells, as an expression host strain, were transformed with each of the recombinant constructs. LB broth media containing 100 µg/ml ampicillin were inoculated with one colony of each recombinant strain and incubated at 37 °C shaker incubator. When absorbance at 600 nm reached 0.5, protein expression was induced by addition of isopropyl-b-D-thiogalacto pyranoside (IPTG) at a final concentration of 1 mM, and the cells were harvested by centrifugation after 4 h incubation at previous condition. Total cell protein analysis of the harvested bacterial cells was done by using 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), according to the method of Laemmli [23].

2.3. Immunoblotting

E. coli cell extract proteins fractionated on a 15% SDS–PAGE were electro-transferred to a PVDF membrane by semi-dry transfer apparatus (Trans-Blot® SD Cell; 221BR 44652, BioRad). After blocking the membrane in 1% w/v non-fat dried milk in phosphate buffered saline for 1 h at room temperature and then overnight at 4 °C, it was incubated in blocking solution containing0.1–0.2 µg/ml rabbit polyclonal antibody to human G-CSF (ab9691, Abcam) with shaking 2 h at room temperature. The membrane was washed in five washes of TBST (PBS + 0.05% Tween 20), 10 min each. The membrane was further soaked 1 h in goat anti-rabbit IgG-peroxidase conjugated antibody (Sigma–Aldrich, USA) diluted in 1.0% non-fat dried milk-PBST at room temperature with shaking. Following five more washes with PBST, the membrane was developed with 3, 3′-diaminobenzidine (DAB) solution (Sigma- Aldrich, USA).

2.4. Expression level analysis

To compare the capability of different recombinant constructs in hG-CSF expression, 10 ml of LB media containing 10 µg/ml ampicillin in 50 ml Erlenmeyer flasks were inoculated with 500 µl of a pre-cultured cell ($OD_{600} \approx 0.5$) suspension of each rhG-CSF bacterial strains to get the optical density of 0.1 at 600 nm. After 2 h of incubation at 37 °C and 150 rpm, IPTG induction was done, followed by 4 h incubation at the same conditions. Then the pellets were collected by centrifugation of 1 ml of each bacterial solution for 1 min at 10,000 rpm. The harvested pellets were resuspended in 200 µl Laemmli sample buffer 6.25 mM Tris–HCl, pH 6.8, 2% SDS, 20% glycerol, 0.2% β-mercapto ethanol and 0.001% bromophenol blue and analyzed by SDS–PAGE. The level of protein expression was quantified by gel densitometry

Table 1
The oligonucleotides used for introducing the silent mutations in the 5′ end of hG-CSF gene.

Name of sequence	Sequence (5′-3′)									GC (%)	CAI (%)	MFE[c] (kcal/mol)	Level of expression (%)
	M +1	T +2	P +3	L +4	G +5	P +6	A +7	S +8					
hG-CSF[a]		ACC	CCC[r]	CTG	GGC	CCT[r]	GCC	AGC	TC	78.3	0.20	−5.60	<1.0[d]
kG[b]	ATG	ACT	CCG	CTG	GGT	CCG	GCT	AGC	AG	65.4	0.96	−10.0	40.5 ± 2.5
zG[b]	ATG	ACT	CCA[r]	TTA	GGT	CCT[r]	GCT	AGC	AG	50.0	0.31	−4.40	56 ± 1.0
dG[b]	ATG	ACA[r]	CCA[r]	TTA	GGA[r]	CCT[r]	GCT	AGC	AG	50.0	0.10	−1.0	53 ± 0.2
pG[b]	ATG	ACA[r]	CCA[r]	CTG	GGT	CCA[r]	GCT	AGC	AG	57.7	0.39	−3.50	55 ± 2.0

[a] The 5′ end of human granulocyte colony-stimulating factor (hG-CSF) isoform b mRNA. It is downstream of signal sequence. Its NCBI reference sequence is NM_172219.2.
[b] The oligonucleotides kG, zG, dG and pG were used as forward primers in PCR amplifications to introduce the silent mutations in the 5′ end of hG-CSF gene. The sequence 5′-ACTTACAT-3′ was designed before ATG to make *Nde*I restriction site in the 5′ end of the primers. The primers kG and zG were designed in this study and the primers dG and pG were synthesized with the nucleic acid sequences mentioned in the previous studies [20,21].
[c] The minimum free energy (MFE) of 40-nucleotides stretched from SD to position +26.
[d] The hG-CSF expression level from the wild type sequence was reported to be less than 1% of total cell proteins in *E. coli* [26,27].
[r] Rare codon.

Please cite this article in press as: Z. Karimi et al., The effect of rare codons following the ATG start codon on expression of human granulocyte-colony stimulating factor in *Escherichia coli*, Protein Expr. Purif. (2015), http://dx.doi.org/10.1016/j.pep.2015.05.017