Biochemical and Biophysical Research Communications 371 (2008) 380–384

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

A synergistic effect of suppressive CGG codon in +2 position and downstream CAT repeats for efficient heterologous protein expression in *Escherichia coli*

Vaishnavo Rabindranath Pai, Vadivel Murugan *

Genetic Engineering Unit, Centre for Biotechnology, Anna University, Sardar Patel Road, Guindy, Chennai 600 025, India

ARTICLE INFO

Article history:
Received 6 April 2008
Available online 25 April 2008

Keywords:
Erythropoietin
Signal sequence
CAT repeats
Translational enhancer
Translational initiation
NGG codons

ABSTRACT

The negative effect of NGG codons at +2 position has been well documented for the down regulation of recombinant protein expression in *Escherichia coli*. But this is not true when certain specific sequences are present in the downstream of NGG codons. This has been proved in our study while expressing human Erythropoietin (EPO) in *E. coli* GJ1158. Towards this, nine recombinant constructs were made and their expression profile was compared. In our results, we found that the suppressive nature of NGG codon (GGG, CGG) in the +2 position was overcome by imposing a downstream CAT repeat motif. The expression of EPO levels is higher in the constructs having the combination of both CGG codon at 2nd position and CAT repeats than the other constructs having either CGG or CAT repeat alone. In addition, it is also interesting to note that increasing number of CAT repeats shows increased expression levels.

Considerable research has been carried out to investigate the role of the codon window in the translation initiation region on recombinant protein expression in *Escherichia coli*. The presence of NGG codons, including the GGG codon in the +2, +3 or +5 positions is known to play an inhibitory role in recombinant expression [1]. Consecutive AGG or AGA/AGG codons in the early coding region are known to significantly lower gene expression [2]. Codon AAA in the +2 position is known to enhance expression levels [3]. The effect of all the 64 codons in the +2 position on the expression of beta galactosidase gene in *E. coli* has been analyzed [4]. A consensus sequence present in the downstream box of *E. coli* comprising of five codons—5′-ATG AAT CAC AAA GTG-3′ and identified as a translation-enhancing element has been introduced in the pCOLD vector series [5].

In our study, we have analyzed the role of codons downstream of the start site on the expression of Human Erythropoietin (EPO) in *E. coli*. EPO is a sialoglycoprotein hormone and is the prime regulator of RBC production [6]. Desialylated EPO (Asialo EPO) has the same EPO receptor binding and neuroprotective properties [7]. In order to exploit this property of asialo EPO, we ventured to express EPO in *E. coli*. The precursor polypeptide is composed of 193 amino acids with a 27 amino acid signal sequence [6]. The mature sequence of EPO is composed of 166 amino acids, two disulfide bonds and is of molecular weight approximately 18.6 kDa in its unglycosylated form [6]. In order to achieve soluble expression of EPO, *E. coli* GJ1158, a salt inducible strain with a lower propensity for the formation of inclusion bodies [8] was used.

Failure to express unfused precursor EPO led us to investigate the effect of suppressive codon signature NGG in the +2 position. A total of nine constructs were made in order to study the sequences or codons in the N-terminal region of both vector and gene influencing EPO expression. From our results, we found that the 'suppressive' codons in fact helped to enhance expression when they were succeeded downstream by the CAT repeat motif—in this case the Hexahistidine coding sequence.

Materials and methods

Bacterial strains and plasmids. Escherichia coli GJ1158 was obtained from the Centre for Cellular and Molecular Biology (CCMB), India. *Escherichia coli* GJ1158 [8], derived from *E. coli* B strain BL21, is a salt inducible strain with a pro U promoter. The expression vector used in this study is pRSET A obtained from Invitrogen Life Technologies, USA (see Supplementary information Fig. S1).

Cloning of the human EPO gene in pRSET vector. Standard recombinant DNA techniques were used for the subcloning of human EPO gene in pRSET vector. For details of primers, restriction enzymes, preparation of template for PCR, refer Supplementary information Table S1. A summary of all the constructs is given in the Table 1. The clones are represented with the following notation—clone 1: $pAESS^{+}$, clone 2: $pAEV_3{}^{+}H_6{}^{+}LS^{+}XP^{+}EK^{+}SS^{-}$, clone 3: $pAESS^{-}$, clone 4: $pAEV_3{}^{+}H_6{}^{+}LS^{+}XP^{+}EK^{+}SS^{+}$, clone 5: $pAESS^{+}GGG^{-}$, clone 6: $pAEV_3{}^{+}H_6{}^{+}SS^{-}$, clone 7: $pAEV_3{}^{+}H_6{}^{-}SS^{-}$, clone 8: $pAEV_3{}^{-}H_6{}^{+}SS^{-}$,

* Corresponding author. Fax: +91 44 22350229.
E-mail address: vmurugan@annauniv.edu (V. Murugan).

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.04.079

Table 1
A schematic representation of the clones developed in this study

	Schematic representation of the recombinant constructs used	Salient features at the N terminus
1	pAESS^{+} CAT ATG GGG GUG **EPO SIGNAL SEQ** **EPO MATURE SEQ**	EPO signal sequence present, no fusion tag, +2, +3 codons of signal sequence shown
2	pAE V$_3^{+}$H$_6^{+}$LS^{+}XP^{+}EK^{+}SS^{-} CAT ATG CGG GGT TCT **6x HIS- LS – XP- EK** **EPO MATURE SEQ**	Complete fusion tag encoded by pRSET vector, no signal sequence, +2, +3, +4 codons of vector shown
3	pAESS^{-} CAT ATG GCC CCA CCA CGC **EPO MATURE SEQ** *Bam*HI site	No signal sequence, no fusion tag, +2, +3, +4, +5 codons of mature EPO shown , CAC repeat in bold.
4	pAE V$_3^{+}$H$_6^{+}$LS^{+}XP^{+}EK^{+}SS^{+} CAT ATG CGG GGT TCT **6x HIS-LS-XP-EK** **EPO SIGNAL SEQ** **EPO MATURE SEQ** *Bam*HI site	Complete fusion tag encoded by pRSET vector followed by EPO signal sequence, +2, +3, +4 codons of vector shown
5	pAESS^{+}GGG^{-} CAT ATG GUG **EPO SIGNAL SEQ** **EPO MATURE SEQ** GGG at +2 position removed	Mutated EPO signal sequence, codon at second position GGG deleted
6	pAE V$_3^{+}$H$_6^{+}$SS^{-} CAT ATG CGG GGT TCT **6x CAT** **EPO MATURE SEQ**	Modified fusion tag, only early vector codons and 6X Histidine codons, no signal sequence
7	pAE V$_3^{+}$H$_6^{-}$SS^{-} CAT ATG CGG GGT TCT GCC CCA CCA CGC **EPO MATURE SEQ**	Modified fusion tag, only early vector codons, no histidine tag, Mature EPO- 1st four codons shown, CAC repeat in bold
8	pAE V$_3^{-}$ H$_6^{+}$SS^{-} CAT ATG **6X CAT** **EPO MATURE SEQ**	Modified fusion tag, only 6X Histidine codons, no signal sequence, no early vector codons
9	pAESS^{+His} CAT ATG GGG GTG CAC **5X CAT** **EPO SIGNAL SEQ** **EPO MATURE SEQ** 5x CAT repeat introduced after the 4th codon in signal sequence	5X CAT repeat introduced after the 4th codon CAC of EPO signal sequence to give 6X repeat, +2, +3, +4 codons of EPO signal sequence shown.

Abbreviations: CAT ATG: NdeI site, CGG GGT TCT: early vector codons, 6X His: histidine tag, LS: gene 10 leader sequence, Xpress epitope represented as XP, enterokinase recognition site denoted by EK. Since the mature sequence of human EPO is common in all the clones, it is not represented in the nomenclature of the constructs; whereas detailed representation has been given to the preceding sequences. pAE: EPO gene in pRSET A vector, V_3: early vector codons 'CGG GGT TCT' immediately following the start codon, H_6: 6X Histidine tag codons, LS: gene 10 leader sequence, XP: X Press Epitope, EK: enterokinase recognition site, SS: EPO signal sequence, (−) and (+) in the superscript indicate the absence or presence, respectively, of features in the preceding sequence.

clone 9: pAESS^{+his}. In the nomenclature, pAE represents EPO gene in pRSET A vector, V_3: early vector codons 'CGG GGT TCT' immediately following the start codon, H_6: 6X Histidine tag codons, LS: gene 10 leader sequence, XP: X Press Epitope, EK: Enterokinase recognition site, SS: EPO signal sequence, − and + in the superscript indicate the absence or presence, respectively, of features in the preceding sequence. 'His' in the superscript of pAESS^{+his} indicates inclusion of His tag within the signal sequence of EPO.

Expression studies of human EPO from pRSET vector. GYE medium without NaCl in M9 salts mixture was used for the growth and induction of *E. coli* GJ1158. GYE medium comprises of 1.0% glucose, M9 salts, yeast extract, 2 mM magnesium sulphate and 1× trace metals mix. Stock solution comprised 1000× trace metal mixture contained 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM each of $MnCl_2$ and $ZnSO_4$, and 2 mM each of $CoCl_2$, $CuCl_2$, $NiCl_2$, Na_2MoO_4, and H_3BO_3 in Millipore water. Three milliliters cultures of *E. coli* GJ1158 harboring the recombinant plasmids were grown to an OD of 0.4 at 37 °C. An aliquot of 1 mL was withdrawn, harvested, to serve as a control of uninduced cells. The remaining culture was induced with 0.1 M NaCl and incubated for a period of three hours at 37 °C in shaker at 180 rpm. The induced cultures were then harvested by centrifugation at 5000*g* for 10 min. Cell pellets of both induced and uninduced cell cultures were then suspended in 1× PBS. The protein concentration was estimated by Bradford method [9]. To the protein samples, sample solubilising buffer containing SDS and β-mercaptoethanol was added, boiled for 10 min at 100 °C. The protein samples were then subjected to 12% SDS–PAGE analysis [10] for protein expression. Percentage of EPO expression in the total protein was estimated by using the spot densitometry tool in Alpha Innotech gel documentation system. Unstained Protein Molecular Weight Marker from Fermentas Life Sciences, Canada and Medium Range Protein marker from Amersham Biosciences, USA was used as standards for SDS–PAGE.

Results and discussion

Effect of EPO signal sequence and the pRSET fusion tag on EPO expression

Initially, we have made attempts to express the EPO gene with and without the native EPO signal sequence in *E. coli* GJ1158. Towards this, the recombinant constructs pAESS^{+} and pAESS^{-} were made (Fig. 1 and Table 1). The precursor form of EPO with the signal sequence failed to show expression in SDS–PAGE analysis (Fig. 2). The expected protein band corresponding to the size of approximately 21 kDa was not detectable both in uninduced and induced status of the culture. Variations in the conditions such as culture OD for induction, time period after induction, concentration of inducer and media composition did not yield detectable levels of expression in SDS–PAGE. On the other hand, the construct containing only the mature EPO sequence (without signal sequence) expressed the protein to moderate levels under the above conditions (Fig. 2). From the above results, we observed that the signal sequence at the N-terminus hampers the expression of precursor EPO from pAESS^{+}. Codons succeeding the translation start site have been shown to play a very important role in translation efficiency [1,11,12]. Especially, the second codon has been cited as a major determinant of the levels of gene expression [1]. In pAESS^{+}, wherein the precursor EPO has been cloned in the NdeI–HindIII site, the triplet GGG is positioned at the second codon. Earlier studies had shown that the presence of codons with