Protein Expression and Purification 74 (2010) 248–256

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Influence of the second amino acid on recombinant protein expression

Louis Bivona, Zhongcheng Zou, Nicole Stutzman, Peter D. Sun *

Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, MD 20852, USA

ARTICLE INFO

Article history:
Received 12 May 2010
and in revised form 15 June 2010
Available online 20 June 2010

Keywords:
Recombinant expression
Second codon position
Igα
Recombinant protein production
Codon usage
mRNA structure
Mutational analysis
Escherichia coli expression system

ABSTRACT

Factors affecting protein expression have been intensely studied to the benefit of recombinant protein production. Through mutational analysis at the +2 amino acid position of recombinant Igα, we examined the effect of all 20 amino acids on protein expression. The results showed that amino acids at the +2 position affected 10-fold in the recombinant protein expression. Specifically, Ala, Cys, Pro, Ser, Thr, and Lys at the +2 position resulted in significantly higher expression of recombinant Igα than other amino acids, while Met, His and Glu resulted in greatly reduced protein expression. This expression difference depended on the amino acid instead of their codon usage. Consistent with the mutational results, a statistically significant enrichment in Ala and Ser at the +2 position was observed among highly expressed *Escherichia coli* genes. This work suggests a general approach to enhance protein expression by incorporating an Ala or Ser after the initiation codon.

Published by Elsevier Inc.

Introduction

The efficiency of translation initiation plays a key role in determining the amount of protein expressed in all organisms. Initiation in *Escherichia coli* has been intensively studied and involves multiple *cis*- and *trans*-acting factors, each contributing to the overall efficiency of translation. Several important *cis*-acting factors involved in translation initiation include the Shine–Dalgarno sequence and its spacing relative to the start codon [1–3], the initiation codon itself (AUG, GUG, rarely AUU or CUG) [4,5], and the non-random distribution of bases located both upstream and downstream of the start codon [6,7].

The downstream region immediately to the 3′ side of the initiation codon has been the focus of many studies related to protein expression. It has been previously shown that changes at the second amino acid position can lead to a 15-fold difference in expression [8]. This work was expanded by Stenstrom et al., in which it was shown that the AAA triplet is the most prevalent codon at the second amino acid position [9]. In addition, they showed that the AAA codon resulted in higher gene expression at the +2 position. However, it has also been reported that tandems of AGA or AGG codons promote favorable translational efficiency [10]. In highly expressed *E. coli* genes, it was found that guanosine is most frequently represented at the first codon position [11]. In addition to finding codons that positively affect protein expression, codons consisting of the sequence NGG at the second amino acid position showed marked decreases in expression [12]. Others have found that identical pairs of all four of the CGN triplets result in drastically inefficient translation [10]. In general, early studies favor nucleotide specific determinants for protein expression, namely codon bias.

Combining a mutational analysis on recombinant protein expression and statistical analysis on highly expressed *E. coli* genes, we show that individual amino acids at the +2 position can greatly affect their protein expression, and that the preference appears to be in amino acid rather than codon usage. The results suggest a general approach to increase protein expression level by a simple insertion of serine or alanine at the second position.

Materials and methods

Mutation of second amino acid

The pET30 vector containing the extracellular domain of Igα was kindly provided by Pavel Tolar. Mutations were made to the second codon of Igα by using Stratagene's QuikChange® II Site-Directed Mutagenesis Kit. Primers were designed by using Stratagene's QuickChange® Primer Design Program (Supplemental Table 1), and synthesized by Integrated DNA Technologies (IDT)[1]. After transformation of the mutated plasmid into XL1-Blue Supercompetent *E. coli* cells, the cells were spread on Luria–Bertani (LB)

* Corresponding author. Fax: +1 301 402 0284.
E-mail address: psun@nih.gov (P.D. Sun).

[1] *Abbreviations used:* IDT, Integrated DNA Technologies; LB, Luria–Bertani; IPTG, isopropyl β-D-1-thiogalactopyranoside.

1046-5928/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.pep.2010.06.005

agar plates supplemented with 25 μg/ml kanamycin overnight at 37 °C. Colonies from the transformation were inoculated at 37 °C in 20 ml of LB broth supplemented with 25 μg/ml kanamycin, and plasmid DNA were extracted using Qiagen's MiniPrep kit. The mutations were confirmed by DNA sequencing (ACTG Inc.). The above procedure was also performed for the mutagenesis of CXCL10.

Protein expression in E. coli

Mutant Igα or CXCL10 plasmids were transformed into *Escherichia coli* BL21 (DE3) cells (Novagen) for protein expression. Individual colonies from the transformation were inoculated in 20 ml of LB broth and grown at 37 °C for overnight. Fifty microliters of the overnight cultures were inoculated in 5 ml of antibiotic free LB media, and induced at an OD_{600} between 0.6 and 0.8 with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h. The cells were then centrifuged and the pellets were resuspended in B-PER® (Bacterial Protein Extraction Reagent) (Pierce) lysis buffer supplemented with 40 μg/ml DNase at room temperature for 10 min (4 ml of BPER per 1 g of wet cell pellet). The lysates were centrifuged at 13,000 rpm for 5 min. As Igα forms inclusion bodies in *E. coli*, the supernatants were removed and the pellets were resuspended in 350 μl of water. Igα expression was evaluated by SDS-gel electrophoresis using Laemmli Buffer (Sigma) and 4–12% NuPAGE Bis–Tris Gels (Invitrogen) for 1×, 2×, and 4× dilutions of the insoluble cell lysate fractions to avoid saturation of Igα band (Fig. 2). The SDS-gels were stained with Coomassie blue, destained, and scanned for intensity analysis using the ImageJ densitometry program. The intensity of Igα was normalized against a 40 kDa endogenous bacterial band to take account the SDS-gel sample loading affect.

Statistical analysis of second amino acid position

The 200 highest expressing *E. coli* genes were listed by Ishihama et al. [13], in which they used a quantitive mass spectrometry approach to determine cytosolic protein concentrations. For comparison, 200 genes were selected at random using the online *E. coli* genome database, EcoCyc [14]. The amino acids at the second position were counted and graphed. These calculations were also performed for the 3rd, 4th, 5th, and 40th amino acids.

RNA structure prediction

RNA structures were predicted between the Shine-Dalgarno sequence and +24 nucleotides from the initiation codon using the RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Results

Serine mutation at the second amino acid position increased Igα expression

pET plasmids remain the most widely used series of prokaryotic vectors for recombinant protein expression (Novagen, Inc.), partly because the presence of various multiple cloning sites with different conveniently built-in affinity tags. In the existing pET vectors, the +2 amino acid immediately after the initiation methionine varies among Gly, Ala, Ser, Asp, Asn, Pro, His, Lys, and Arg. While the choice of a specific pET vector is primarily based on the availability of unique cloning sites and desirable affinity tags, we have occasionally cloned our target genes into two or more pET vectors. We observed in several cases that significant variations in recombinant protein expression existed among different pET vectors. Specifically, the pET vectors with +2 amino acids being Ala and Ser appeared to result in 2- to 5-fold higher expression of the same recombinant protein. When the mature sequence of a human B cell co-receptor, Igα, was cloned into a pET-30a vector using the 5′-NdeI site, which resulted in a simple addition of the initiation Met to the N-terminal leucine of the mature Igα sequence, this wildtype Igα was expressed at low levels in BL21 (DE3) strain of *E. coli* (Fig. 1). When the wildtype Igα construct was transformed into a Rosetta strain of BL21 cells, it resulted in a 1.7-fold increase in Igα expression, consistent with observed improvements in expression using the Rosetta strain of bacteria [15]. A subsequent mutation of Leu to Ser at the +2 position further increased the Igα expression by 2-fold over the wildtype in BL21 Rosetta cells (Fig. 1). Thus, the combination of Rosetta cells and the Ser mutation resulted in almost a 4-fold increase in Igα expression.

Effect of other amino acids at +2 position on the expression of Igα protein

To further evaluate whether other amino acids at the +2 position also influence the Igα expression in *E. coli*, we systematically replaced the +2 position Leu of Igα with all 19 amino acids using QuikChange® II Site-Directed Mutagenesis and analyzed the effect of mutations to the protein expression. After mutagenesis, the plasmids were transformed into *E. coli* BL21 (DE3) cells. Individual colonies were cultured in LB broth and induced with 1 mM IPTG at OD_{600} between 0.6 and 0.8. In all cases, individual colonies of the same mutants resulted in consistent levels of protein expression with no significant clone variations. For example, three clones from each of the Glu, His and Met mutations as well as two clones from the Ser and Val mutations showed little variation in their Igα expression (Fig. 2A and B). However, substantial variations in Igα expression were observed among different mutants (Fig. 2C and D, Table 1). To quantify the differences in mutant Igα expression, serial dilutions of the wildtype and mutant expression samples were analyzed by SDS-gel electrophoresis and the intensities of

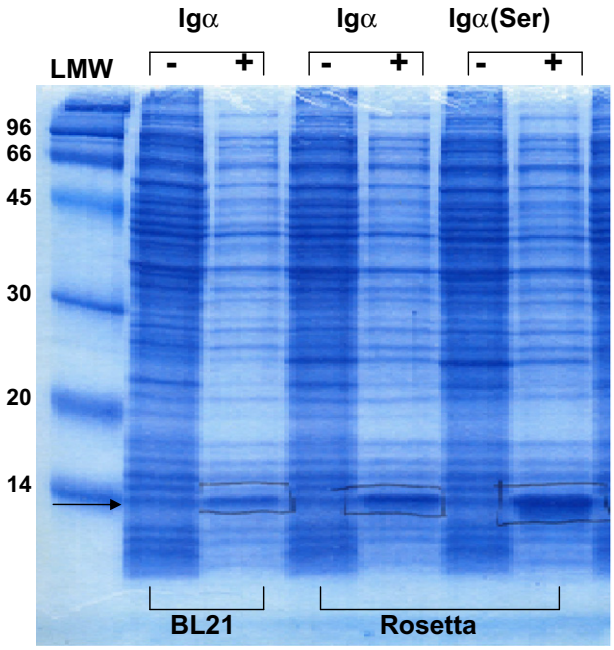

Fig. 1. Initial results indicating that substituting a serine at the second amino acid position results in higher protein expression in *E. coli*. The lanes represent the uninduced (−) and induced (+) samples of Igα, from left to right, in BL21 (DE3), BL21 Rosetta cells and Leu to Ser mutation in BL21 Rosetta cells. The low molecular weight standard is shown on the left.