Biochemical and Biophysical Research Communications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

# An alternative method of enhancing the expression level of heterologous protein in *Escherichia coli*

Jun Yin [1], Hong Tian [1], Lichen Bao, Xin Dai, Xiangdong Gao *,[2], Wenbing Yao *

*State Key Laboratory of Natural Medicines, School of Life Science and Technology, China Pharmaceutical University, Nanjing 210009, China*

ARTICLE INFO

*Article history:*
Received 23 October 2014
Available online xxxx

*Keywords:*
Expression level
Translation initiation region
mRNA secondary structure
$\Delta G$
Translation efficiency

ABSTRACT

Though numerous strategy options are available for achieving high expression levels of genes in *Escherichia coli*, not every gene can be efficiently expressed in this organism. By investigating the relationship between the mRNA secondary structure of translational initiation region (TIR) and gene expression in *E. coli*, we establish a simple method to design sequences of appropriate TIR (from −35 to +36) that meet a specific expression level as we need. Using this method, overexpression of native human humor necrosis factor α and extracellular domain of Her2/neu protein (aa 23–146) in *E. coli* were achieved. Differences in expression appeared was mainly related to the efficiency of translation initiation and the stability of mRNA secondary structure, because the intracellular mRNA levels analyzed by real-time RT-PCR were quite similar. Our approach can overcome the steric hindrance of translation startup, and therefore promote translation smoothly to acquire high expression of exogenous protein.



## 1. Introduction

The expression of target protein in different hosts is unpredictable and challengeable. It can be affected by various factors, including the copy number and stability of the expression plasmid, the feature of target gene, the synthetic level and stability of mRNA, the structure and stability of the synthesized protein [1–3]. Many advantages of *Escherichia coli* have ensured that it remains a very valuable organism for the high-level production of recombinant proteins [4,5]. Despite the substantial knowledge on the genetics and molecular biology of *E. coli*, it is still difficult to express every gene efficiently in this organism [6]. In addition, no gold criteria have been established in *E. coli*, and strategies for the optimization of expression level differ greatly from gene to gene and are often decided by trial and error.

In the previous study, the native human tumor necrosis factor α (TNF-α) presents poor expression level in some strong prokaryotic expression vectors, even in pET vectors. After fused a $His_6$-Tag to the N terminal of human TNF-α gene, the expression of this protein was remarkably enhanced. The unique difference between the native TNF-α and $His_6$-TNF-α is the $His_6$-Tag which contains continuous 6 CAU codons just next to the initiation AUG codon. We considered that the remarkable increase of the expression level may be correlated to the enlargement of A/U base-pair ratio in translational initiation region (TIR). To enhance the expression of exogenous protein without changing the amino acid sequence, we focus on the relationship between the TIR and gene expression.

It is now clear that the various translation efficiencies of different mRNAs are predominantly due to the unique structural features at the 5′ end of each mRNA [7–9]. Meanwhile, the mRNA secondary structure of translation initiation region (TIR) plays a crucial role in the efficiency of gene expression [10–12]. The stability of mRNA secondary structure is quantified as the value of minimal folding free energy ($\Delta G$). An increase in the $\Delta G$ of 1.4 kcal/mol corresponds to the reduction by a factor of 10 in translation initiation rate [13]. Wu et al. used $\Delta G$ of the secondary structure as the index and chose a 73 bp fragment of TIR from +1 to +73, optimized the TIR of xylanase gene from *Thermotoga maritima*. Consequently, the xylanase was successfully overexpressed in *E. coli* [14]. In particular, Seo et al. found the $\Delta G$ of the downstream region (DR) was linearly associated with the relative expression level over a range of 4-fold change in *E. coli* [15]. Though the $\Delta G$ of TIR is found to be correlated with the relative expression level, there has been no straightforward approach that can direct us about how to design sequences of TIR that meet a specific expression level as we need.

In this study, we seek to establish a universal method that could regulate the expression level of target protein without changing

* Corresponding authors. Fax: +86 25 83271249 (X. Gao). Fax: +86 25 83302827 (W. Yao).
*E-mail addresses:* xdgao@cpu.edu.cn (X. Gao), wbyao@cpu.edu.cn (W. Yao).
[1] These authors contributed equally to this work.
[2] We designate this one to further communicate with the Editorial and Production offices.

http://dx.doi.org/10.1016/j.bbrc.2014.10.149
0006-291X/

the amino acid sequence by optimizing the mRNA secondary structure of TIR. Based on this method, we further develop a general tool, called OpTIR. Silent mutations were systematically introduced into specific codons at the 5′ end of exogenous gene to modulate the local secondary structure of the mRNA, recombined with the same vector to form different TIRs after optimization. By these results, we may get a series of sequences, which contain some specific expression levels as we need. Furthermore, we report the significant improvement of the expression levels of human TNF-α and Her2 (aa 23–146) in *E. coli* by using our method. Extracellular domain of Her2/neu protein (aa 23–146) containing two Her2 T-helper epitopes: p42 (aa 42–56), p98 (aa 98–114) [16] is likely to be chosen for an anti-Her2 vaccine to break the immunological tolerance. Thereby, we can optimize the gene sequence at the beginning phase of recombinant protein expression to avoid repeatedly designing the target gene from scratch for high expression.

## 2. Materials and methods

### 2.1. Bacterial strains, plasmids and biochemical reagents

*E. coli* DH5α, *E. coli* BL21(DE3) and pET28a(+) vector were purchased from Invitrogen. Human TNF-α gene (GenBank NO. KM658597) and Her2 (aa 23–146) gene (GenBank NO. KM658598) were synthesized by Invitrogen (Shanghai, China). *E. coli* DH5α was used for cloning and amplifying plasmids and *E. coli* BL21 (DE3) was employed for expression. PrimerStar DNA polymerase and LA Taq DNA polymerase (Takara, China) were used for PCR. Restriction enzymes and goat anti-rabbit IgG conjugated alkaline phosphatase were purchased from Sangon Biotech (Shanghai) Co., Ltd. (China). Primers were synthesized by Invitrogen Corporation Shanghai Representative Office (China). For immunoblot analysis, rabbit anti-human TNF-α antibody was obtained from Anbo Biotechnology Co., Ltd. (USA).

**Table 1**
The appropriate codons for coding the 2–12th amino acids of target protein.

| Amino acid | Codon | Amino acid | Codon |
|---|---|---|---|
| F | UUU UUC | H | CAU CAC |
| L | UUA UUG CUU CUG | Q | CAA CAG |
| I | AUU AUC | N | AAU AAC |
| M | AUG | K | AAA AAG |
| V | GUU GUC GUA GUG | D | GAU GAC |
| S | UCU AGC AGU | E | GAA GAG |
| P | CCU CCA CCG | C | UGU UGC |
| T | ACU ACC ACA ACG | W | UGG |
| A | GCU GCC GCA GCG | R | CGU CGC |
| Y | UAU UAC | G | GGU GGC GGG |

### 2.2. Tool development based on ΔG of mRNA TIR

According to *E. coli* codon preference, we removed a portion of non-preference codons that are less frequently used so that the appropriate codons could be chosen to code the 2–12th amino acids of target protein (Table 1). Then, all possible combinations of mRNA sequences coding 2–12th amino acids of the target protein were calculated, recombined with the 35 bases upstream of the start codon in the vector and initiator codon AUG to form the various TIRs. We used ViennaRNA Package [17] to calculate $\Delta G$ of these TIRs, arraying the order based on their $\Delta G$ value. This tool provided the number of outputs depended on the index value of $\Delta G$ we set, making us easier to choose the $\Delta G$ that interested us for encoding the target protein in *E. coli*. The optimization of gene sequence mentioned above was achieved by using C language on Linux system (Online Resource 1). The calculation process was developed into a general tool named OpTIR (Fig. 1A).

### 2.3. Molecular cloning

DNA digestion and molecular cloning were performed following standard procedures [18], and DNA restriction and modification enzymes were purchased from Takara. Plasmids or DNA fragments were isolated from *E. coli* cells or agarose gels with TIANprep Mini Plasmid Kid, or TIANgel Midi Purification Kit, respectively.

### 2.4. RNA secondary structure prediction and gene mutagenesis

The secondary structures of mRNA in the TIRs of human TNF-α and Her2 (aa 23–146) were predicted using the algorithm of minimum free energy by online Vienna RNA Secondary Structure Prediction program (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). The selected regions for predicting were defined from the point −35 to +36 including the Shine-Dalgarno (SD) and the transcription start

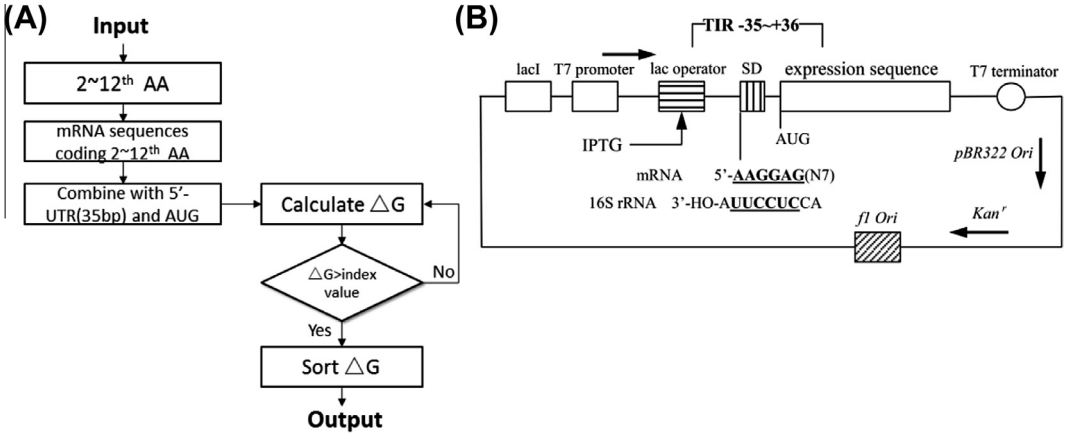


**Fig. 1.** (A) Schematic representation of the calculation process for the tool named OpTIR. (B) Schematic presentation of the features and sequence elements of the prokaryotic expression vector pET28a(+). The direction of transcription is indicated by the bold arrow. The part of TIR (from −35 to +36) consists of the SD sequence followed by an A+T-rich translational spacer that has an optimal length of approximately 7 bases. As shown, the 3′ end of the 16S rRNA interacts with the SD sequence during translational initiation. The repressor is encoded by a regulatory gene (lacI), which may suppress the activity of the lac operator. The activity of the lac operator is induced by the isopropylthiogalactoside (IPTG). The expression region of coding strand is transcribed by T7 RNA polymerase on account of T7 promoter. The transcription T7 terminator serves to stabilize the mRNA and the vector. In addition, an antibiotic resistance gene for kanamycin facilitates phenotypic selection of the vector, and the origin of replication (ori) determines the copy number of the vector. The various features are not drawn to scale.

Please cite this article in press as: J. Yin et al., An alternative method of enhancing the expression level of heterologous protein in *Escherichia coli*, Biochem. Biophys. Res. Commun. (2014), http://dx.doi.org/10.1016/j.bbrc.2014.10.149