Journal of Genetic Engineering and Biotechnology (2017) xxx, xxx–xxx

HOSTED BY

ELSEVIER

Academy of Scientific Research & Technology and National Research Center, Egypt

Journal of Genetic Engineering and Biotechnology

www.elsevier.com/locate/jgeb

ORIGINAL ARTICLE

# Chemical complexity of protein determines optimal *E. coli* expression host; A comparative study using Erythropoietin, Streptokinase and Tumor Necrosis Factor Receptor

**Subramani Ramkumar** *, **Vaishnavo Rabindranath Pai, Chinnathambi Thangadurai, Vidhya Priya Murugan**

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

Received 8 August 2016; revised 3 November 2016; accepted 19 December 2016

**KEYWORDS**

Erythropoietin;
Streptokinase;
TNFR;
Expression;
Recombinant;
Inclusion bodies

**Abstract** High throughput expression of proteins is often hampered by the failure of certain proteins to express in the particular *E. coli* host strain used for the study. The identification of a host strain compatible for a wide variety of proteins is desirable. In this study, the recombinant expression of therapeutic proteins Erythropoietin (EPO), Streptokinase (SK) and Tumor Necrosis Factor Receptor Extra cellular domain (TNFR ED) that vary widely in their chemical nature was studied in four different strains of *E. coli* namely BL21 (DE3), BL21 (DE3) pLys S, BL21 (DE3) Rosetta pLys S and GJ1158. Since there are no previous report for the analysis of expression and solubility of the above mentioned proteins we studied the same in various *E. coli* stains. Here we report that *E. coli* strain GJ1158 which uses salt induction was found to be the most suitable for overexpression of all the three proteins. Interestingly rare codons were found not to play any significant role in the expression. Protein toxicity and aggregation propensity were also studied. One of the major factors influencing expression was the tendency of the protein to aggregate which in turn influences folding and toxicity levels. The solubility of the proteins was inversely proportional to aggregation. Expression levels were in the order of TNFR ED < EPO < SK. In conclusion, it was observed that *E. coli* GJ1158, a strain known to decrease aggregation of proteins was found to be more suited for expression. This is the first time GJ1158 has been included in this kind of analysis for comparison of protein expression in various *E. coli* hosts.



## 1. Introduction

An efficient recombinant protein production strategy determines the friendliest host for the protein of interest in order to

* Corresponding author.
E-mail address: ramelango2002@gmail.com (S. Ramkumar).
Peer review under responsibility of National Research Center, Egypt.

http://dx.doi.org/10.1016/j.jgeb.2016.12.006

achieve high yields. When a large number of recombinant proteins are being screened for expression, the choice of the host for expression is of major importance. A host that is compatible for majority of the proteins is desired.

In our study, we have compared the expression of three major therapeutic proteins namely Streptokinase (SK), Erythropoietin (EPO) and Tumor Necrosis Factor Receptor extracellular domain (TNFR ED) in four different strains of *E. coli*. Streptokinase is a thrombolytic agent, its source being *Streptococcus*. TNFR ED is a human transmembrane receptor used in treatment of rheumatoid arthritis [2]. EPO is a human secretory glycoprotein used in chronic anemia therapy [9]. In this study on recombinant protein synthesis, a construct carrying the first domain of TNFR II namely, TNFR-ED was used for expression analysis. A number of differences are present between the proteins for eg. Post-translational modifications such as glycosylation, presence of disulfide bonds, number and composition of rare codons (Table 1). There are no disulfide bonds in SK, since there are no cysteine residues present in this protein. On the other hand human TNFR ED, which is pretty rich in cysteine residues, contains 12 disulfide bonds. Bridging these two above-mentioned proteins is human EPO having two disulfide bonds. Both TNFR ED and EPO are known to be glycosylated in humans; however in *E. coli* this particular posttranslational modification is not possible. Based on these modifications and composition of amino acids, the proteins can be said to be in the following order of chemical complexity – TNFR ED > EPO > SK. It is therefore interesting to analyze the heterologous expression behaviors of the above-mentioned proteins in different *E. coli* host strains. Interestingly, there are no previous reports of such comparison for the above-mentioned proteins.

The *E. coli* expression hosts that have been used are also different in that they have their own special purpose in recombinant protein expression. *E. coli* BL21 (DE3) is the commonly used host for expression of non-toxic proteins. *E. coli* BL21 (DE3) pLys S is also a common host, used however preferentially for the expression of toxic proteins. [16]. *E. coli* BL21 (DE3) Rosetta pLys S is also a stringent host providing tRNAs for rare codons. *E. coli* GJ1158 is unique in that it contains the T7 RNA Polymerase gene under the control of pro U promoter, which is induced by increased salt concentrations [14]. There have been many reports comparing expression of particular proteins in different hosts [3,4], however this is the first time *E. coli* GJ1158 has been included in this kind of analysis. Here we have attempted to analyze the different expression levels of the therapeutic proteins EPO, TNFR ED and SK in different hosts and of the different proteins in same host.

## 2. Materials and methods

### 2.1. Bacterial strains and plasmids

For initial transformation of the ligation mixtures, the maintenance strain *E. coli* DH5α was used. The bacterial strains used to study expression were *E. coli* GJ1158, *E. coli* BL21 (DE3), *E. coli* BL21 (DE3) pLys S, *E. coli* BL21 Rosetta (DE3) pLys S. *E. coli* GJ1158 [14], derived from *E. coli* B strain BL21, is a salt inducible strain with a pro U promoter (Table 2). The strain was obtained from the Centre for Cellular and Molecular Biology (CCMB), India. *E. coli* BL21 (DE3), *E. coli* BL21 (DE3) pLys S, *E. coli* BL21 Rosetta (DE3) pLys S (Novagen,) are IPTG inducible strains in which the T7 RNA Polymerase gene is under the control of a Lac UV5 promoter.

The expression vector used in this study is pRSET A obtained from Invitrogen Life Technologies, USA. pRSET is a high copy number plasmid with a pUC origin of replication.

### 2.2. Recombinant plasmid construction

Standard recombinant DNA techniques were used for the cloning of human EPO gene, Streptokinase and TNFR ED in pRSET vector. Genes were cloned in the *Bam* HI-*Hind* III

**Table 1** Characteristics of the proteins used in this study.

| | Streptokinase | Erythropoietin | TNFRII-ED |
|---|---|---|---|
| Origin | Prokaryotic | Eukaryotic | Eukaryotic |
| Molecular weight (kDa) | 45 | 18.6 | 25 |
| No of amino acids in protein used in this study | 414 | 166 | 235 |
| Rare codons | 9 leucine<br>8 arginine<br>2 proline<br>2 isoleucine<br>No rare codon repeats | 8 arginine<br><br>1 prolines<br><br> | 13 prolines<br>6 arginines<br>Arg-Pro doublet<br>In two positions |
| Disulfide bonds | No cysteines | 2 | 12 |
| Glycosylation | None | 4 sites | 4 sites |
| Protein localization | Secretory in *Streptococcus* | Secretory in humans | Either part of receptor or shed into blood stream |
| Hydrophilicity(Kyle and Do Little) |  |  |  |
| Insolubility index http: //www.biotech.ou.edu/ | 57.3% chance of solubility | 72.6% chance of insolubility | 87.5% chance of insolubility |