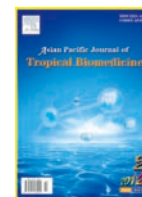




Contents lists available at ScienceDirect

## Asian Pacific Journal of Tropical Biomedicine

journal homepage: [www.elsevier.com/locate/apjtb](http://www.elsevier.com/locate/apjtb)

Document heading doi:10.1016/S2221-1691(11)60213-X © 2012 by the Asian Pacific Journal of Tropical Biomedicine. All rights reserved.

# Strategies for production of active eukaryotic proteins in bacterial expression system

Orawan Khow, Sunutcha Suntrarachun\*

Queen Saovabha Memorial Institute, Thai Red Cross Society, Bangkok 10330, Thailand

## ARTICLE INFO

## Article history:

Received 25 July 2011  
 Received in revised form 20 August 2011  
 Accepted 1 September 2011  
 Available online 28 February 2012

## Keywords:

Recombinant protein  
 Expression  
*Escherichia coli*

## ABSTRACT

Bacteria have long been the favorite expression system for recombinant protein production. However, the flaw of the system is that insoluble and inactive proteins are co-produced due to codon bias, protein folding, phosphorylation, glycosylation, mRNA stability and promoter strength. Factors are cited and the methods to convert to soluble and active proteins are described, for example a tight control of *Escherichia coli* milieu, refolding from inclusion body and through fusion technology.

## 1. Introduction

Bacterial expression is the most common expression system employed for the production of recombinant proteins. The organism, especially *Escherichia coli* (*E. coli*), is easy to manipulate, inexpensive in culturing and fast in generation of a recombinant protein. However, since it is a prokaryotic based system, heterologous eukaryotic proteins expressed are not correctly modified, and it can also be difficult to facilitate the secretion of expressed protein in large amounts. Moreover, proteins expressed in large amounts might precipitate, forming inclusion bodies and large complex proteins could be difficult to propagate[1].

## 2. Factors affecting bacterial expression system

### 2.1. Codon usage/bias

Most of amino acids have been encoded by more than one codon and all available amino acid codons are bias utilized by each organism. Transfer RNA (tRNA) of cells reflects the codon bias of its mRNA. Observation of codon usage in *E. coli* reveals that highly expressed genes exhibit greater codon bias than poorly expressed ones and the frequency of synonymous codons used reflects the abundance of their cognate tRNAs. These can imply that heterologous genes with abundant codons, rarely used in *E. coli*, may not be efficiently expressed in *E. coli* and may lead to translation error. Codon bias becomes highly prevalent problems when rare codons in the transcripts form clusters such as doublets or triplets accumulation that is large in quantities. Translation error arised from rare codon bias includes mistranslational amino acid substitutions, frameshifting events or premature translational termination[2–4].

### 2.2. Protein folding

Expression of recombinant proteins in *E. coli* is mainly directed to three different locations *i.e.* the cytoplasm, the periplasm, and the growth medium (through secretion). Expression in the cytoplasm is preferred since the production yields are usually high. Cytoplasmic folding is often enhanced at low temperatures thus the use of cold-

\*Corresponding author: Sunutcha Suntrarachun, Research and Development, Queen Saovabha Memorial Institute, Thai Red Cross Society, Bangkok 10330, Thailand, 1871 Rama IV Rd., Patumwan, Bangkok 10330, Thailand.

Tel: 66-2-2520161-4 ext 161

Fax: 66-2-2540212

E-mail: [sunutcha@yahoo.com](mailto:sunutcha@yahoo.com)

Foundation Project: This work was financially supported by Queen Saovabha Memorial Institute, The Thai Red Cross Society.

inducible promoters may facilitate this process. However, this is often accompanied by misfolding and segregation into insoluble aggregates known as inclusion bodies. Aggregation can be reduced to minimum through the control of parameters such as temperature, expression rate and host metabolism. Though formation of inclusion body renders easier protein purification, there is no guarantee that the *in vitro* refolding will generate large amounts of biologically active products. To release recombinant proteins into the periplasm and the growth medium, many systems have been studied. As such a approach is complicated, the systems have not been commercialized[1,2,5].

### 2.3. Protein phosphorylation and glycosylation

*E. coli* has limited eukaryotic post-translational machinery function, which is considered as a key disadvantage for producing the eukaryotic phosphoproteins *i.e.* serine/threonine/tyrosine protein kinases. To overcome these obstacles, co-expression of modified mammalian enzymes such as protein methylases and acetylases and their substrates from single or two separate plasmid vectors in the same *E. coli* may result in the production of recombinant proteins that closely resemble native eukaryotic proteins[2].

Glycosylation is another complex process of post-translational modification. It is responsible for the formation of cellular glycans which are often attached to proteins and lipids. Glycosyltransferase and glycosidases are enzymes responsible for glycosylation of many proteins. Glycoproteins, which are commonly distributed in eukaryotic cells, are rarely presented in prokaryotic organisms because cellular organelles essential for glycosylation are missing in these organisms[4,6–8].

### 2.4. Stability of mRNA

The stability of mRNA affects expression rates. The average half-life of mRNA in *E. coli* at 37 °C ranges from seconds to the maximum at 20 min and the expression rate depends directly on the inherent stability of mRNA. Degradation of mRNA by RNases can be protected through RNA folding, ribosomes and stability modulation by polyadenylation. Recombinant expression systems with mRNA stability enhancement is commercially available, for example, Invitrogen BL21 star strain, containing a mutation in the gene encoding RNaseE[1,8,9].

### 2.5. Promoter strength

Recombinant expression plasmids require strong transcriptional promoter to enable high-level gene expression. Promoter must be induced using either thermal or chemical means and the most common inducer is the sugar molecular isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). However, IPTG is not suitable for large scale production of human therapeutic proteins because it is toxic

and expensive[2,9–11].

## 3. Strategies for improving the expression of active and soluble protein

### 3.1. Tight control of the *E. coli* cellular milieu

Expression of soluble proteins can be regulated through many factors that the host cell normally use in controlling of toxic protein expression[6,8,12–14].

#### 3.1.1. Modification of *E. coli* host strain

The strain or genetic background of host strain is important for recombinant protein expression. Expression strains should be deficient in harmful proteases, but should stably maintain the expression plasmid and confer the relevant genetic elements to the expression system. DE3. *E. coli* BL21 is an example of the most common host and it has been proven outstanding in application for standard recombinant expression. It can grow efficiently in minimal media as non-pathogenic bacterium that cannot survive to cause diseases in host tissues[9,11,15,16].

#### 3.1.2. Modification of media composition

Production of recombinant protein requires nutrients for bacterial growth and there is a limited control on the growth parameters. This process often leads to changes in substrate depletion, pH, and concentration of dissolved oxygen as well as accumulation of inhibitory substances from various metabolic pathways. These changes are not beneficial for the production of either soluble or correctly folded active protein. Proper and efficient protein folding might require specific cofactors in the growth media such as metal ions. Addition of these essential factors to the culture media could considerably increase the yield as well as the folding rate of the soluble proteins[4,17].

#### 3.1.3. Expression at lower temperatures

Protein expression in *E. coli* growing at low temperature has shown its success in improving the solubility of proteins that are difficult to express as soluble proteins. Expression at low temperature conditions leads to the increase of stability and correct folding patterns due to the fact that hydrophobic interactions determining inclusion body formation are temperature dependent. Moreover, any expression associated with toxic phenotype observed at 37 °C incubation conditions, will be suppressed at low temperatures. The increase of expression and activity of lower temperatures growth is associated with increased expression of chaperones in *E. coli*. Therefore, growth at a temperature range of (15–23) °C, could also lead to a significant reduction of expressed protein degradation[4,18,19].

#### 3.1.4. Co-expression of molecular chaperones

Molecular chaperones are proteins adapted to assist

*de novo* protein folding and /or facilitate expressed polypeptide's proper conformation attainment. Co-expression of molecular chaperone strategy has been adopted for prevention of inclusion body formation, leading to improving of solubility of the recombinant protein. Chaperones are working as a trigger factor assisting in recombinant protein refolding. These polypeptides continue to attain folding into the native state even after their release from the protein-chaperone complex. Moreover, some chaperons could also prevent protein aggregation[2,20–22].

### 3.2. Inclusion body folding

Inclusion bodies are intracellular protein aggregates which were observed when the target gene is over expressed in the cytoplasm of *E. coli*. Formation of inclusion bodies in recombinant expression systems occurs as a result of erroneous equilibrium between *in vitro* protein solubilization and aggregation and might lead to unfavorable protein folding[23–25].

#### 3.2.1. Refolding/resolubilization of *E. coli* inclusion body proteins

Recombinant proteins expressed as inclusion bodies in *E. coli* have been widely used for the commercial product of therapeutic proteins. The major drawbacks during the refolding of inclusion body proteins into more efficient, soluble and correct folded product are reducing of recovery. Other than that the requirement for optimization of refolding conditions for each target protein and the resolubilization procedure could possibly affect the activity of refolded protein. Therefore, the production of soluble recombinant protein remains a preferable alternative than the *in vitro* refolding procedures[1,9,23–25].

#### 3.2.2. Isolation/solubilization of inclusion bodies

Isolation of inclusion bodies can be done by lysozyme treatment along with EDTA before cell homogenization to facilitate cell disruption. Inclusion bodies are recovered by low speed centrifugation of bacterial cells that has been mechanically disrupted either by ultrasonication or high pressure homogenization. Bacterial cell envelop or outer membrane proteins may co-precipitate with the insoluble fractions as the inclusion body impurities. These contaminants can easily be removed by adding detergents such as Triton X-100 or low concentrations of chaotropic compounds. After removal of the impurities, inclusion bodies are solubilized by various concentrations of chaotropic agents such as urea or guanidinium hydrochloride. The latter is more favored due to its better chaotropic properties. Inclusion body proteins that were solubilized under mild denatured conditions are better in refolding yields and retaining of biological activities[4,9,23–25].

#### 3.2.3. Refolding of solubilized and unfolded proteins

The methods normally used for solubilization of

inclusion body could lead to non-native conformation of the expressed protein. This problem could be resolved by proper refolding procedures of target protein at low denaturant concentrations. Higher concentration of the unfolded protein often leads to decreased refolding yields, regardless of refolding method. So, it is desirable to keep the concentration of the initial un-folded protein to a minimum level if higher and correct refolding proteins are expected[1,3].

### 3.3. Active proteins production

#### 3.3.1. Production of fusion protein

In order to simplify the expression, solubilization and purification of recombinant proteins, a wide range of protein fusion partners have been developed. These fusion proteins usually include a partner or "tag" which may be linked to the target protein by a recognition site-specific protease. Most fusion partners are exploited for the purpose of specific affinity purification. Several different tags are commercially available and offer additional advantages such as protection of partner protein from intracellular proteolysis, enhanced solubility and they can also be used as specific expression reporters. The most popular affinity tags are poly-histidine (His6) tags, which are compatible with immobilized metal affinity chromatography (IMAC) and the glutathione S-transferase (GST) tag for purification through glutathione based resins. GST, 27 kDa, can be prohibitive due to the slow binding kinetics of GST to glutathione-sepharose resin and lead to loading of cell extracts extremely time consuming, especially when large cell culture volumes are being processed. Poly-histidine tags, on the other hand, are small and do not, in most cases, affect the folding of the attached protein. It also has very strong reversible binding attributes allowing for rapid single-step purification. Polyhistidine tags can be attached on either the N- or C- termini of recombinant proteins, but the optimal location depends on the folding and biochemical characteristics of the adjacent recombinant protein[14,20,26,27].

#### 3.3.2. Site-specific protease

The recombinant purified protein can be separated from its fusion partners, such as affinity tags, solubility enhancers or expression reporters, by site-specific proteases. These enzymes are highly efficient for the cleavage at the inserted recognition sequence. Selection of specific protease and its optimal cleavage conditions mostly depend upon the amino acid sequences of target protein. Therefore, fusion tags with protease cleavage sequences similar to the one present in recombinant protein should be avoided. Two serine proteases, factor Xa and thrombin, are widely used for site-specific fusion protein cleavage. Enterokinase and tobacco etch virus protease (TEV) are examples of specific proteases used in intracellular processing of fusion proteins[9,25].

## 4. Conclusion

Although many alternative organisms and expression systems are now available for recombinant protein production, bacteria such as *E. coli* continue to be the most attractive host for the production of heterologous proteins. At present, certain post-translational modification cannot be achieved in *E. coli*, factors related to improvement of recombinant protein expression in prokaryotic system have been reviewed. New approaches for production of complex eukaryotic proteins in a prokaryotic expression system might become available in the near future.

## Conflict of interest statement

We declare that we have no conflict of interest.

## Acknowledgements

We would like to thank Miss Pannipa Chulasugandha for revising the manuscript.

## References

- [1] Jana S, Deb JK. Strategies for efficient production of heterogenous proteins in *Escherichia coli*. *Appl Microbiol Biotechnol* 2005; **67**: 289–298.
- [2] Cabrita LD, Dai W, Bottomley SP. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* 2006; **1**: 6–12.
- [3] Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, et al. Design parameters to control synthetic gene expression in *Escherichia coli*. [Online] Available from: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0007002>. [Accessed on 20 March, 2011]
- [4] Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* 2008; **307**: 249–264.
- [5] Yin J, Li G, Ren X, Herrler G. Select what you need; a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol* 2007; **127**: 335–347.
- [6] Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 2009; **27**: 297–306.
- [7] Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci* 2009; **18**: 936–948.
- [8] Makino T, Skretas G, Georgiou G. Strain engineering for improved expression of recombinant protein in bacteria. *Microb Cell Fact* 2011; **10**: 32–41.
- [9] Sorensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 2005; **115**: 113–128.
- [10] Lewin A, Mayer M, Chusainow J, Jacob D, Appel B. Viral promoters can initiate expression of toxin genes introduced into *Escherichia coli*. *BMC Biotechnol* 2005; **20**: 19–27.
- [11] Samuelson JC. Recent developments in difficult protein expression: a guide to *E. coli* strains, promoters, and relevant host mutations. *Methods Mol Biol* 2011; **705**: 195–209.
- [12] Perrakis A, Romier C. Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview methods. *Methods Mol Biol* 2008; **426**: 247–256.
- [13] Valdez-Cruz NA, Caspeta L, Perez NO, Ramirez OT, Trujillo-Roldan MA. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and pR promoters. *Microb Cell Fact* 2010; **19**: 9–18.
- [14] Francis DM, Rage R. Strategies to optimize protein expression in *E. coli*. *Curr Protoc Protein Sci* 2010; **24**: 1–29.
- [15] Hunt I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr Purif* 2005; **40**: 1–22.
- [16] Saida F, Uzan M, Odaert B, Bontems F. Expression of high toxic genes in *E. coli*: special strategies and genetic tools. *Curr Protein Pept Sci* 2006; **7**: 47–56.
- [17] Peti W, Page R. Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif* 2007; **51**: 1–10.
- [18] Saida F. Overview on the expression of toxic gene products in *Escherichia coli*. *Curr Protoc Protein Sci* 2007.
- [19] Caspeta L, Flores N, Pérez NO, Bolívar F, Ramírez OT. The effect of heating rate on *Escherichia coli* metabolism, physiological stress, transcriptional response, and production of temperature-induced recombinant protein: a scale-down study. *Biotechnol Bioeng* 2009; **102**: 468–482.
- [20] de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol* 2007; **7**: 32–40.
- [21] Peleg Y, Unger T. Application of high-throughput methodologies to the expression of recombinant proteins in *E. coli*. *Methods Mol Biol* 2008; **426**: 197–208.
- [22] Burgess RR. Refolding solubilization inclusion body proteins. *Methods Enzymol* 2009; **463**: 259–282.
- [23] Mannall GJ, Titchener-Hooker NJ, Dalby PA. Factors affecting protein refolding yields in a fed-batch and batch-refolding system. *Biotechnol Bioeng* 2007; **97**: 1523–1534.
- [24] de Groot NS, Espargaró A, Morell M, Ventura S. Studies on bacterial inclusion bodies. *Future Microbiol* 2008; **3**: 423–435.
- [25] Eiberle MK, Jungbauer A. Technical refolding of proteins: do we have freedom to operate? *Biotechnol J* 2010; **5**: 547–559.
- [26] Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* 2005; **99**: 303–310.
- [27] Rabhi-Essafi I, Sadok A, Khalaf N, Fathallah DM. A strategy for high level expression of soluble and functional human interferon & as a GST-fusion protein in *E. coli*. *Protein Eng Des Sel* 2007; **20**: 201–209.