

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

**WEB OF SCIENCE™**Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com

Plasmids for Optimizing Expression of Recombinant Proteins in *E. coli*

Ahmed Mahmoud Al-Hejin, Roop Singh Bora and Mohamed Morsi M. Ahmed

Additional information is available at the end of the chapter

Abstract

Plasmids are important vectors for the transfer of genetic material among microbes. The transfer of plasmids causes transmission of genes involved in pathogenesis and survival, to the host bacteria leading to their evolution and adaptation to diverse environmental conditions. A large number of plasmids of varying sizes have been discovered and isolated from various microorganisms. Plasmids are also valuable tools to genetically manipulate microbes for various purposes including production of recombinant proteins. *Escherichia coli* is the most preferred microbe for production of recombinant proteins, due to rapid growth rate, cost-effectiveness, high yield of the recombinant proteins and easy scale-up process. Several plasmids have been designed to optimize the expression of heterologous proteins in *E. coli*. In order to circumvent the issues of protein refolding, the codon usage in *E. coli*, the absence of post-translational modifications, such as glycosylation and low recovery of functionally active recombinant proteins, various plasmids have been designed and constructed. This chapter summarizes the recent technological advancements that have extended the use of the *E. coli* expression system to produce more complex proteins, including glycosylated recombinant proteins and therapeutic antibodies.

Keywords: plasmid, recombinant protein, promoter, codon usage, molecular chaperones, fusion tags

1. Introduction

Plasmids are defined as extrachromosomal double-stranded circular DNAs within a cell that have the capability to replicate independently of chromosomal DNA. Plasmids are found in many microorganisms including bacteria, archaea and some eukaryotes such as yeast [1].

The advent of new DNA sequencing technologies has successfully determined the complete sequence of 4602 plasmids; most of these plasmids, that is, 4418 are from bacteria, 137 plasmids are identified from archaea and 47 plasmids are identified from eukaryotes [2]. The size of plasmids can vary between 1 and 200 kb, and they more often harbour genes encoding proteins that confer selective advantage to host cells under adverse conditions. Some of the genes are known as resistance genes which confer resistance to certain antibiotics. Genes involved in synthesis of antibiotics and various kinds of toxins are also localized on plasmids. Some of the genes present on plasmids encode for various virulence factors that assist microbes to colonize host and escape from its defence mechanisms. Plasmids also harbour genes that empower bacteria to fix nitrogen.

Plasmids can be present in a single bacterial cell in varying number which may range from one to few hundreds. The usual number of plasmids that are present in an individual cell is termed as copy number and is governed by the size of plasmid and the regulation of replication initiation. Large-size plasmids are present in low copy number and exist as one or very few copies in single bacterium. Such single-copy plasmids employ *parABS* and *parMRC* systems, termed as partition system, to equally segregate a copy of plasmid to each daughter cell upon cell division [3].

Independent replication of plasmids requires the presence of a region of DNA that can serve as an origin of replication. A self-replicating unit is termed as a replicon. A classical bacterial replicon comprises of gene for plasmid-specific replication initiation protein (*Rep*), *DnaA* boxes, AT-rich region and repeating units called iterons [4]. Large-size plasmids also harbour genes required for their replication, while smaller plasmids employ host replicative machinery to undergo replication.

Plasmids can be transmitted from one bacteria to another by the process of conjugation, and it has been reported that approximately 14% of the currently known plasmids are conjugative [5, 6]. Conjugation is a very efficient mechanism to transfer genes among microbes and thus facilitate the rapid evolution and adaptation of microbes to various adverse environmental conditions [7]. This transfer of genes among bacteria is one mechanism of horizontal gene transfer and is responsible for spread of antibiotic resistance among pathogenic microbes [8, 9].

Plasmids are very commonly used as vectors in the field of genetic engineering for the purpose of cloning and expression of desired genes. Various types of plasmids are now available commercially for cloning and expression of foreign genes in a wide variety of host including *E. coli*, yeast and mammalian cells [10]. The desired genes to be cloned and expressed are inserted into suitable plasmid. For cloning purpose, plasmid vectors are designed to contain a site, known as multiple cloning site (MCS) or polylinker site, which allows insertion of heterologous genes. The multiple cloning site contains several commonly used restriction sites for Type III restriction enzymes. The plasmid vectors contain an origin of replication, which allows its replication in bacterial host. Besides, these plasmids also harbour a gene that confers resistance to a specific antibiotic such as ampicillin or kanamycin which is used as a selectable marker. After the insertion of desired heterologous gene into multiple cloning site of the plasmid vector, the constructed plasmid is introduced into bacterial cells by the process of transformation. The transformed cells are further exposed to selective growth

medium containing the specific antibiotic. The cells that contain the introduced plasmid will be able to survive and grow in selective medium as they carry the plasmid with the antibiotic resistance gene. This strategy is employed to clone and express heterologous proteins in *E. coli* for the large-scale production of recombinant proteins for therapeutics and wide variety of functional studies [10]. In this chapter, we summarize the recent technological advancements in the field of molecular biology that have extended the use of the *E. coli* expression system to produce more complex proteins, including glycosylated recombinant proteins and therapeutic antibodies.

2. *E. coli* as an expression system for production of recombinant proteins

Escherichia coli is a very commonly used, robust and cost-effective expression system for large-scale production of recombinant proteins. *E. coli* is genetically well characterized and is easy to handle and manipulate genetically. Its faster growth rate, inexpensive culture media, high expression levels and easy scale-up process provide major advantages for large-scale production of recombinant proteins for therapeutic purposes or various functional studies [11]. *E. coli* was successfully used to manufacture recombinant human insulin in 1982 for treating diabetes patients. It is important to note that insulin is a heterodimer and entails oxidative protein folding to attain a functionally active 3D conformation. The success accomplished with high level expression of recombinant human insulin validated the significance of *E. coli* expression system for large-scale production of recombinant proteins. Besides human insulin, several recombinant proteins for therapeutic applications, including human growth hormone, interferon $\alpha 2a$ and $\alpha 2b$, glucagon, urate oxidase, granulocyte colony-stimulating factor and parathyroid hormone, have been successfully manufactured using *E. coli* expression system [12]. Although *E. coli* has been used extensively for expression of heterologous proteins, it is still not possible to determine the optimum production conditions for all the proteins. Expression conditions that are optimal for one protein may not be ideal for another protein. One of the major issues while producing heterologous proteins in *E. coli* is differences in codon usage between the two organisms. This difference in codon usage could cause errors in translation leading to low expression levels of recombinant proteins [13]. The other parameters that can affect the protein expression are choice of promoters, growth conditions and hydrophobicity of proteins.

3. Promoter

Promoter is a very critical region in plasmid vectors, used for the expression of heterologous proteins. Promoter is a stretch of DNA that is involved in the initiation of transcription of a gene and is located upstream of the transcription initiation site of gene. Promoters are normally 100–1000 base pairs in size. In *E. coli* and other bacterial species, promoter encompasses two short DNA sequences that are 10 nucleotides (termed as the Pribnow Box) and

35 nucleotides upstream from the transcription initiation site. The consensus sequence at –10 region is ‘TATAAT’ and the consensus sequence at –35 region is ‘TTGACA’. This promoter sequence is recognized by RNA polymerase which leads to initiation of transcription. Several plasmids with strong or weak promoter are now available to express heterologous proteins in *E. coli*. Some of the commonly used promoters in *E. coli* expression vectors include T7 promoter, derived from bacteriophage T7, *E. coli* lac promoter, its improved modified version lacUV5 and Tac promoter produced from the combination of trp and lac promoters (**Table 1**). Another important promoter is the trc promoter, which is originated from lacUV5 and trp promoters. The potency of a promoter is governed by the frequency of transcription initiation which is regulated by the affinity of RNA polymerase for the promoter sequence. The T7 promoter is very strong as compared to *E. coli* promoters due to high frequency of transcription initiation and efficient processivity and hence it is routinely used for very high expression

Plasmid	Promoter	Origin of replication	Antibiotic selection marker	Fusion tags	Size of the fusion tag	Protease cleavage site	Vendor
pGEX-2 T	tac	pBR322	Ampicillin	GST	26 kDa	Thr	Pharmacia
pGEX-3X	tac	pBR322	Ampicillin	GST	26 kDa	Xa	Pharmacia
pGEX-6P	tac	pBR322	Ampicillin	GST	26 kDa	Pre	Pharmacia
pMAL-c2X	tac	ColE1	Ampicillin	MBP	42 kDa	Xa	New England Biolabs
pMAL-c2E	tac	ColE1	Ampicillin	MBP	42 kDa	EK	New England Biolabs
pQE-30/31/32	T5-lac	ColE1	Ampicillin	N-6XHis	1 kDa	None	Qiagen
pTrxFus	PL	ColE1	Ampicillin	TRX	11.6 kDa	EK	Invitrogen
pET-14b	T7	pBR322	Ampicillin	N-6XHis	1 kDa	Thr	Novagen
pET-19b	T7-lac	pBR322	Ampicillin	N-6XHis	1 kDa	EK	Novagen
pET-29a-c(+)	T7-lac	pBR322	Kanamycin	C-6XHis	1 kDa	Thr	Novagen
pET-41a-c(+)	T7-lac	pBR322	Kanamycin	GST	26 kDa	Thr	Novagen
pET-43a-c(+)	T7-lac	pBR322	Kanamycin	NusA	55 kDa	Thr	Novagen
pET-SUMO	T7	pBR322	Kanamycin	SUMO	11 kDa	SUMO protease	Invitrogen

GST, Glutathione S-transferase; MBP, maltose-binding protein; TRX, thioredoxin; NusA, N-Utilization substance A; N-6XHis, N-terminus 6X-histidine; C-6XHis, C-terminus 6X-histidine; SUMO, small ubiquitin-like modifier (SUMO); Thr, Thrombin; Xa, Factor Xa; Pre, PreScission; EK, Enterokinase;

Source: <http://wolfson.huji.ac.il/expression/bac-strains-prot-exp>; <http://www.emdmillipore.com/life-science-research/novagen>; <http://www.invitrogen.com/1/3/stratagene-products> <https://www.neb.com>

Table 1. List of some commonly used plasmids for improving expression levels and solubility of recombinant proteins in *E. coli*.

levels of recombinant proteins. However, in some instances, large-scale production of recombinant proteins can lead to its accumulation as insoluble aggregates also known as inclusion bodies, which can cause poor yield of biologically active recombinant proteins. In such cases, the use of expression vector containing a weak promoter such as the *trc* promoter instead of the T7 promoter can enhance protein solubility [14].

4. Ribosomal binding site

Ribosomal binding site is a very crucial component in plasmids commonly used for expression of recombinant proteins in *E. coli*. It is comprised of translation initiation codon, that is, AUG and the Shine-Dalgarno (SD) sequence and is required for efficient translation initiation [15]. Shine-Dalgarno sequence is localized 7–9 nucleotides upstream from the initiation codon and the consensus SD sequence is AAGGAGG. Various factors are known to affect translation initiation such as secondary structure of ribosomal binding site, consensus SD sequence, varying number of thymine and adenine and also the nucleotides upstream and downstream of the initiation codon AUG [16]. The translation efficiency is enhanced by the presence of more number of adenine and thymine in the ribosomal binding site. Highly expressed genes are found to contain adenine after the initiation codon [17]. Park and colleagues designed various variants of 5'-untranslated region (UTR) comprising of SD sequence and the AU-rich region, using PCR-based site-directed mutagenesis and analysed their impact on protein expression levels [18]. Such a strategy of modifying 5'UTR could be of immense value to improve translation efficiency and obtain high expression level of recombinant proteins. Incorporation of simple variations in 5'UTR could be exploited to optimize the expression of heterologous proteins [18, 19]. Another important factor that could affect the translation efficiency is the secondary structures in mRNA which could lead to variations in protein expression levels. The RNA helicase DEAD protein of *E. coli* can be exploited to remove secondary structures in mRNA. It had been demonstrated that co-expression of DEAD protein increased the expression of β -galactosidase from T7 promoter by several fold, implying that DEAD-box protein is involved in stabilizing the mRNA [20, 21]. This property of DEAD-box protein can be used to enhance the expression of genes which are poorly expressed due to secondary structures of mRNA. Protein translation is a highly efficient process as any error during translation could cause mutations, misincorporation of amino acids and low expression levels and hence can severely affect the quality of recombinant proteins produced in *E. coli*.

5. Codon usage and plasmid containing tRNA genes cognate to the rare codons

Codon usage is a major issue while expressing heterologous proteins, particularly human proteins in *E. coli*. There are marked differences in codon usage between *E. coli* and humans. Codons that are found to be very common in human and other eukaryotic genes are very rare

in *E. coli*. Presence of rare codons in heterologous genes can lead to errors in translation and cause low expression levels of recombinant protein in *E. coli*. The presence of rare codons in heterologous genes might cause translational errors due to ribosomal stalling at these positions. These translational errors include frame-shift mutations, amino acid substitutions or premature translation termination [22]. Some of the rare codons in *E. coli* that cause problems in recombinant proteins are AGA, CGG, CGA, AGG (arginine), AAG (lysine), GGA (glycine), CUA (leucine), AUA (isoleucine) and CCC (proline) [23]. In *E. coli*, CGG is a rare arginine codon which occurs at a frequency of 0.54%. McNulty and colleagues demonstrated that the presence of large number of rare arginine codon CGG in p27 protease domain from Herpes Simplex Virus 2 (HSV-2) resulted in the synthesis of recombinant protein of molecular weight that was 3 kDa higher than the actual molecular weight when expressed in *E. coli* [22]. The resultant increase in molecular weight was found to be due to the +1 frame-shift mutation at one of the CGG codons at the C-terminus of the viral protein. Besides, glutamine residues were misincorporated instead of arginine due to misreading of CGG as CAG [22].

Various strategies have been designed to circumvent the problem of codon bias in *E. coli* for enhancing the production of authentic, biologically active heterologous recombinant proteins. One strategy is to synthesize the full-length gene based on codon usage, but the high cost of gene synthesis is a major drawback. Another strategy requires site-directed mutagenesis of the foreign gene to generate codons which correlates with the tRNA pool of *E. coli*. However, this process is very expensive and time-consuming. Another approach is to co-transform the *E. coli* with plasmid containing the tRNA gene cognate to the rare codons. By increasing the copy number of rare tRNA genes, *E. coli* strains can be designed to complement the codon usage frequency in the foreign gene. This strategy is very feasible and cost-effective and highly efficient for expression of heterologous genes harbouring large number of rare codons. McNulty and colleagues carried out the co-expression of argX gene which codes for the cognate tRNA for rare arginine codon CGG, with the p27 protease domain of HSV-2 in order to circumvent the problem of codon bias. It was observed that the co-expression of cognate tRNA gene for CGG codon resulted in abolition of both frame-shift mutation and glutamine misincorporation and enhanced the expression levels of authentic recombinant protein by up to sevenfold [22]. This study clearly suggested that supplementation of the cognate tRNA for the rare codons such as CGG can alleviate the CGG codon bias in *E. coli* and hence lead to accurate and efficient synthesis of recombinant proteins. This strategy is now routinely being employed for several difficult-to-express heterologous recombinant proteins containing rare codons, in *E. coli*. Currently, several plasmids such as pRARE plasmids are commercially available which harbour genes encoding for tRNA cognate to rare codons. Another important feature in these pRARE plasmids is the presence of p15A replication origin, which facilitate their maintenance in the presence of compatible ColE1 origin of replication, commonly present in several *E. coli* expression vectors. Moreover, several *E. coli* strains are now commercially available that carry plasmids containing tRNA genes for cognate rare codons, such as BL21(DE3) CodonPlus-RIL and Rosetta (DE3). Tegel and colleagues analysed the expression of several human proteins in *E. coli* strain Rosetta (DE3) harbouring pRARE plasmid and observed that the total yields of the 35 recombinant proteins out of 68 proteins tested were enhanced significantly [24].

6. Plasmids carrying molecular chaperones for optimization of protein folding

Production of recombinant proteins for therapeutic purposes or various functional studies requires a robust and cost-effective expression system which can synthesize heterologous proteins in soluble form. Although *E. coli* expression system is always a preferred choice for expression of recombinant proteins, accumulation of foreign proteins as insoluble aggregates, also called as inclusion bodies, is a major problem. Recovery of proteins from these inclusion bodies is a very cumbersome process which entails denaturation and renaturation steps to obtain recombinant protein in properly folded and soluble form. However, this extraction process causes tremendous loss of proteins and further reduces the total yield of biologically active recombinant proteins. One approach to enhance the solubility of heterologous protein and reduce the formation of inclusion bodies is to employ molecular chaperones. It is now known that molecular chaperones assist the nascent polypeptide to fold properly during the process of protein synthesis and thus prevent protein aggregation. Few molecular chaperones are found to improve folding and solubilization of misfolded protein, while other chaperones are involved in prevention of protein aggregation [25–27]. The commonly used molecular chaperones in *E. coli* are GroEL, GroES, DnaK, DnaJ and Trigger factor (**Table 2**). These cytoplasmic chaperones can be employed either individually or in combination of different chaperones to enhance protein solubility and prevent formation of inclusion bodies [10, 26, 28, 29]. The GroEL-GroES chaperone combination is highly efficient to enhance protein refolding and also prevent protein degradation. It has been shown that Trigger factor interacts with GroEL and increases GroEL-substrate binding to improve protein folding [30]. Some chaperones such as heat shock proteins IpbA and IpbB prevent aggregation of heat denatured proteins [31]. It is advisable to test different combinations of molecular chaperones to identify the most efficient combination for improving the solubility of heterologous recombinant proteins. Co-expression of molecular chaperones Skp and FkpA in *E. coli* had been shown to improve the solubility of antibody fragments [32]. Combination of GroEL-GroES chaperones was found to be very efficient in production of anti-B-type natriuretic peptide single-chain antibody (scFv), as 65% of the expressed protein was in soluble form, which was almost 2.4-fold more than the one obtained in the absence of chaperones [33].

Plasmid	Chaperones	Promoter	Antibiotic marker	Inducer	Vendor
pGro7	<i>groES-groEL</i>	<i>araB</i>	Chloramphenicol	L-Arabinose	TaKaRa
pKJE7	<i>dnaK-dnaJ-grpE</i>	<i>araB</i>	Chloramphenicol	L-Arabinose	TaKaRa
pG-KJE8	<i>dnaK-dnaJ-grpE</i> <i>groES-groEL</i>	<i>araB Pzt-1</i>	Chloramphenicol	L-Arabinose Tetracycline	TaKaRa
pTf16	Trigger factor (tig)	<i>araB</i>	Chloramphenicol	L-Arabinose	TaKaRa
pG-Tf2	<i>groES-groEL-tig</i>	<i>Pzt-1</i>	Chloramphenicol	Tetracycline	TaKaRa
pBB540	<i>grpE, clpB</i>	Lac	Chloramphenicol	IPTG	Addgene

Table 2. List of plasmids carrying molecular chaperones for improving protein folding in *E. coli*.

It has been demonstrated that the periplasm of *E. coli* presents an ideal environment to express complex therapeutic proteins including antibody fragments. The oxidizing condition and Dsb protein family in the periplasm provide ideal environment for proper disulphide bond formation and folding of recombinant proteins. Moreover, very few host proteins are present in the periplasm, which leads to high yield of purified recombinant proteins. Therapeutic antibodies such as Lucentis and Cimzia (Fab fragments) and few full-length aglycosylated antibodies and scFvs have been successfully produced by periplasmic expression in *E. coli* [34, 35]. Yim and colleagues employed the endoxylanase signal peptide to produce large amount of granulocyte colony-stimulating factor (GCSF) at 4.2 g/l in the periplasm of *E. coli* [36]. Co-expression of periplasmic chaperones can be exploited to improve the expression of properly folded protein in periplasm of *E. coli*. Overexpression of periplasmic chaperones DsbA and DsbC was found to enhance the efficiency of the assembly of the heavy chain and light chain of antibody in the periplasm of *E. coli* and drastically increased the production of full-length antibody from 0.1 to 1.05 g/l [37]. Co-expression of anti-CD20 scFv antibody with the periplasmic chaperone Skp resulted in the enhanced yield as well as antigen binding of antibody [38]. Lee and colleagues developed a highly efficient *E. coli* expression system to produce full-length antibody, by modifying 5'UTR sequence and co-expressing periplasmic chaperone DsbC that resulted in very high yield of light and heavy chains and improved assembly in the periplasm [39]. These successful studies demonstrated that it is possible to produce complex therapeutic proteins including therapeutic monoclonal antibodies, through proper engineering of *E. coli*.

7. Use of plasmids containing fusion tags to improve solubility

Another strategy to improve the solubility of recombinant proteins is to construct a fusion with a highly soluble protein. Several plasmid vectors are commercially available that carry fusion protein tags. Fusion tags technology can be used to increase protein expression, improve solubility as well as facilitate purification of recombinant proteins. Fusion tags are currently one of the most preferred methods to produce difficult-to-express heterologous proteins in *E. coli*. Some of the most commonly used fusion tags are maltose-binding protein (MBP), glutathione-s-transferase (GST), thioredoxin (TRX), NusA (N-Utilization substance A), ubiquitin (Ub), small ubiquitin-like modifier (SUMO) and split SUMO as shown in **Table 1** [40–43]. Marblestone and colleagues carried out a comparative study to evaluate the expression levels and solubility of three heterologous proteins fused to C-terminus of GST, MBP, NusA, Ub, TRX and SUMO fusion tags. TRX and SUMO fusion partners were found to enhance the expression levels of recombinant proteins as compared to other fusion tags, while SUMO and NusA were found to enhance the solubility of recombinant proteins as compared to other fusion tags [41]. Another study by Braun and colleagues analysed the expression of 32 human proteins of varying molecular weight ranging in size from 17 to 110 kDa using various fusion tags and demonstrated that GST and MBP fusion tags are very efficient in improving the expression levels and also total yield of recombinant proteins after purification was high as compared to other fusion partners [40]. Another study analysed the expression of 40 different heterologous proteins with various fusion tags and observed that MBP fusion tag was

very efficient in enhancing the expression levels and solubility of recombinant proteins as compared to other fusion tags [44]. The variations in the data from these comparative studies suggested that the various fusion tags vary in their efficiency for improving the expression and solubility of recombinant proteins, which may depend upon the amino acid compositions, number of disulphide bonds and hydrophobicity of the heterologous proteins. Hence, it is advisable to screen for the most efficient fusion tag for each desired heterologous protein to improve its expression and solubility. One major issue with using fusion tags for improving solubility of heterologous protein is the removal of fusion tags, as it may interfere with the functional activity of the recombinant proteins. To remove fusion tags, cleavage sites are introduced between the fusion tag and recombinant protein, which is recognized and cleaved by site-specific proteases such as factor Xa, thrombin protease or SUMO protease. However, cleavage of the fusion tags can result in lower yield of recombinant proteins. Hence, it is advisable to select the most efficient fusion tag and cleavage strategy to achieve the desirable high yield of authentic and biologically active recombinant proteins.

8. Future perspectives

E. coli is a preferred expression system for production of heterologous proteins due to its well-characterized genetics, ease of genetic manipulation, availability of several plasmid vectors and engineered host strains, low manufacturing cost, high yield of recombinant proteins as compared to other expression systems including yeast, mammalian cell lines, transgenic plants and transgenic animals. However, there are some limitations which need to be surmounted such as codon bias, protein folding and solubility issues and post-translational modifications. Several technological advancements have been made to address these issues. Plasmids such as pRARE plasmids have been designed that contain tRNA genes cognate to the rare codons. Co-transformation of these plasmids would increase the copy number of rare tRNA genes in *E. coli* host and thus would be able to complement the codon usage frequency in heterologous genes. This strategy is very cost-effective and more efficient for enhancing the expression levels of heterologous genes containing large number of rare codons. Solubility and proper folding of recombinant proteins can be achieved by using plasmids that contain genes encoding for molecular chaperones such as GroEL, GroES, DnaK, DnaJ and Trigger factor. Molecular chaperones are known to assist in proper folding of recombinant proteins and prevent formation of inclusion bodies. Similarly, fusion protein tags such as GST, MBP, NusA, Ub, TRX and SUMO can be exploited to improve the expression levels of difficult-to-express recombinant proteins and enhance their solubility. In addition, expression of recombinant proteins in periplasm of *E. coli* along with molecular chaperones provides various advantages such as improved solubility, proper protein folding, easier protein purification and higher yield of authentic and biologically active recombinant proteins. Some of the antibodies that have been approved for therapeutic use in humans such as Lucentis and Cimzia have been successfully produced in the periplasm of *E. coli*, thus confirming the commercial viability of this approach. One of the major drawbacks of *E. coli* expression system is the absence of post-translational modifications such as glycosylation which limits its utility for

production of complex glycosylated biopharmaceuticals. Wacker and colleagues discovered a novel N-linked glycosylation pathway in bacteria *Campylobacter jejuni* and demonstrated the successful transfer of glycosylation pathway in *E. coli* to generate a strain with capability to produce recombinant glycosylated proteins [45]. These various technological advancements have demonstrated that *E. coli* can be engineered specifically for each heterologous protein to obtain high yield of biologically active products.

Author details

Ahmed Mahmoud Al-Hejin^{1*}, Roop Singh Bora^{1,2} and Mohamed Morsi M. Ahmed^{1,3}

*Address all correspondence to: aalhejin@kau.edu.sa

1 Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

2 Department of Biotechnology, Eternal University, Baru Sahib, HP, India

3 Nucleic Acids Research Department, Genetic Engineering and Biotechnology Research Institute (GEBRI), City for Scientific Research and Technological Applications, Alexandria, Egypt

References

- [1] Funnell B, Phillips G. Preface. In: Funnell B, Phillips G, editors. *Plasmid Biology*. Washington, DC: ASM Press; 2004. xi
- [2] Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology*. 2015;**6**:242. DOI: 10.3389/fmicb.2015.00242
- [3] Salje J, Gayathri P, Lowe J. The ParMRC system: Molecular mechanisms of plasmid segregation by actin-like filaments. *Nature*. 2010;**8**:683-692
- [4] Rajewska M, Wegrzyn K, Konieczny I. AT-rich region and repeated sequences—The essential elements of replication origins of bacterial replicons. *FEMS Microbiology Reviews*. 2012;**36**(2):408-434
- [5] Frost LS, Koraimann G. Regulation of bacterial conjugation: Balancing opportunity with adversity. *Future Microbiology*. 2010;**5**:1057-1071. DOI: 10.2217/fmb.10.70
- [6] Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F. Mobility of plasmids. *Microbiology and Molecular Biology Reviews*. 2010;**74**:434-452. DOI: 10.1128/MMBR.00020-10
- [7] Aminov RI. Horizontal gene exchange in environmental microbiota. *Frontiers in Microbiology*. 2011;**2**:158. DOI: 10.3389/fmicb.2011.00158

- [8] Frost L, Leplae R, Summers A, Toussaint A. Mobile genetic elements: The agents of open source evolution. *Nature Reviews. Microbiology*. 2005;**3**:722-732. DOI: 10.1038/nrmicro1235
- [9] Sota M, Top E. Horizontal gene transfer mediated by plasmids. In: Lipps G, editor. *Plasmids: Current Research and Future Trends*. Norfolk, VA: Caister Academic Press, Horizon Scientific Press; 2008. pp. 111-181
- [10] Arya R, JSM S, Bora RS, Saini KS. Optimization of culture parameters and novel strategies to improve protein solubility. *Methods and Protocols, Methods*. 2015;**1258**:45-63. DOI: 10.1007/978-1-4939-2205-5_3
- [11] Rodriguez V, Asenjo JA, Andrews BA. Design and implementation of a high yield production system for recombinant expression of peptides. *Microbial Cell Factories*. 2014;**13**:65
- [12] Baeshen MN, Al-Hejin A, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS, et al. Production of biopharmaceuticals in *E. coli*: Current scenario and future perspectives. *Journal of Microbiology and Biotechnology*. 2015;**25**(7):953-962
- [13] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends in Biotechnology*. 2004;**22**:346-353
- [14] Samuelson JC. Recent developments in difficult protein expression: A guide to *E. coli* strains, promoters, and relevant host mutations. In: *Heterologous Gene Expression in E. coli*. *Methods in Molecular Biology*. 2011;**705**:195-209
- [15] Sørensen HP, Laursen BS, Mortensen KK, Sperling-Petersen HU. Bacterial translation initiation—Mechanism and regulation. *Recent Research Developments in Biophysics and Biochemistry*. 2002;**2**:243-270
- [16] Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews*. 2005;**69**:101-123
- [17] Stenstrom C, Jin H, Major L, Tate W, Isaksson LA. Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene*. 2001;**263**:273-284
- [18] Park YS, Seo SW, Hwang S, Chu HS, Ahn JH, Kim TW, et al. Design of 5'- untranslated region variants for tunable expression in *Escherichia coli*. *Biochemical and Biophysical Research Communications*. 2007;**356**:136-141
- [19] Seo SW, Yang JS, Cho HS, Yang J, Kim SC, Park JM, et al. Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Scientific Reports*. 2014;**4**:4515
- [20] Iost I, Bizebard T, Dreyfus M. Functions of DEAD-box proteins in bacteria: Current knowledge and pending questions. *Biochimica et Biophysica Acta*. 2013;**1829**:866-877
- [21] Linder P, Daugeron MC. Are DEAD-box proteins becoming respectable helicases? *Nature Structural Biology*. 2000;**7**:97-99

- [22] McNulty DE, Claffee BA, Huddleston MJ, Kane JF. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Expression and Purification*. 2003;**27**:365-374
- [23] Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, et al. Modified nucleoside dependent Watson-Crick and wobble codon binding by tRNA^{Lys}-sUUU species. *Biochemistry*. 2000;**39**:13390-13395
- [24] Tegel H, Tourle S, Ottosson J, Persson A. Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta (DE3). *Protein Expression and Purification*. 2010;**69**:159-167
- [25] Carrio MM, Villaverde A. Role of molecular chaperones in inclusion body formation. *FEBS Letters*. 2003;**537**:215-221
- [26] de Marco A. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. *Nature Protocols*. 2007;**2**:2632-2639
- [27] Ronez F, Arbault P, Guzzo J. Co-expression of the small heat shock protein, Lo18, with β -glucosidase in *Escherichia coli* improves solubilization and reveals various associations with overproduced heterologous protein, GroEL/ES. *Biotechnology Letters*. 2012;**34**:935-939
- [28] Cui SS, Lin XZ, Shen JH. Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948. *Protein Expression and Purification*. 2011;**77**:166-172
- [29] Levy R, Weiss R, Chen G, Iverson BL, Georgiou G. Production of correctly folded Fab antibody fragment in the cytoplasm of *Escherichia coli* *trxB* *gor* mutants via the coexpression of molecular chaperones. *Protein Expression and Purification*. 2001;**23**:338-347
- [30] Folwarczna J, Moravec T, Plchova H, Hoffmeisterova H, Cerovska N. Efficient expression of human papillomavirus 16 E7 oncoprotein fused to C-terminus of tobacco mosaic virus (TMV) coat protein using molecular chaperones in *Escherichia coli*. *Protein Expression and Purification*. 2012;**85**:152-157
- [31] Guzzo J. Biotechnical applications of small heat shock proteins from bacteria. *The International Journal of Biochemistry and Cell Biology*. 2012;**44**:1698-1705
- [32] Ow DSW, Lim DYX, Nissom PM, Camattari A, Wong VVT. Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. *Microbial Cell Factories*. 2010;**9**:22
- [33] Maeng BH, Nam DH, Kim YH. Coexpression of molecular chaperones to enhance functional expression of anti-BNPscFv in the cytoplasm of *Escherichia coli* for the detection of B-type natriuretic peptide. *World Journal of Microbiology and Biotechnology*. 2011;**27**:1391-1398
- [34] Jung ST, Kang TH, Kelton W, Georgiou G. Bypassing glycosylation: Engineering aglycosylated full-length IgG antibodies for human therapy. *Current Opinion in Biotechnology*. 2011;**22**:858-867

- [35] Nelson AL, Reichert JM. Development trends for therapeutic antibody fragments. *Nature Biotechnology*. 2009;**27**:331-337
- [36] Yim S, Jeong K, Chang H, Lee S. High-level secretory production of human granulocyte-colony stimulating factor by fed-batch culture of recombinant *Escherichia coli*. *Bioprocess and Biosystems Engineering*. 2001;**24**:249-254
- [37] Reilly DE, Yansura DG. Production of monoclonal antibodies in *E. coli*. In: Shire SJ, Gombotz W, Bechtold-Peters K, Andya J, editors. *Current Trends in Monoclonal Antibodies Development and Manufacturing*. New York: Springer; 2010. pp. 295-308
- [38] Mavrangelos C, Thiel M, Adamson PJ, Millard DJ, Nobbs S, Zola H, et al. Increased yield and activity of soluble single-chain antibody fragments by combining highlevel expression and the Skp periplasmic chaperonin. *Protein Expression and Purification*. 2001;**23**:289-295
- [39] Lee YJ, Lee DH, Jeong KJ. Enhanced production of human full-length immunoglobulin G1 in the periplasm of *Escherichia coli*. *Applied Microbiology and Biotechnology*. 2014;**98**:1237-1246
- [40] Braun P, Hu Y, Shen B, Halleck A, Koundinya M, Harlow E, et al. Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences*. 2002;**99**(5):2654-2659
- [41] Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Science*. 2006;**15**:182-189
- [42] Raran-Kurussi S, Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS One*. 2012;**7**:e49589
- [43] Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*. 2012;**7**:620-634
- [44] Shih YP, Kung WM, Chen JC, Yeh CH, Wang AH, Wang TF. High-throughput screening of soluble recombinant proteins. *Protein Science*. 2002;**11**(7):1714-1719
- [45] Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, et al. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 2002;**298**:1790-1793

