J. Microbiol. Biotechnol. (2015), 25(7), 953–962
http://dx.doi.org/10.4014/jmb.1412.12079

Review

jmb

Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives

Mohammed N. Baeshen[1], Ahmed M. Al-Hejin[1], Roop S. Bora[1,2], Mohamed M. M. Ahmed[1,3]*, Hassan A. I. Ramadan[1,4], Kulvinder S. Saini[1,2], Nabih A. Baeshen[1], and Elrashdy M. Redwan[1,5]

[1]*Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, 21589, Saudi Arabia*
[2]*Department of Biotechnology, Eternal University, Baru Sahib-173 101, Himachal Pradesh, India*
[3]*Nucleic Acids Research Department, Genetic Engineering and Biotechnology Research Institute (GEBRI), City for Scientific Research and Technology Applications, Alexandria 21934, Egypt*
[4]*Cell Biology Department, Genetic Engineering and Biotechnology Division, National Research Centre, Dokki-Cairo 12311, Egypt*
[5]*Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City for Scientific Research and Applied Technology, New Borg AL-Arab, Alexandria 21934, Egypt*

Received: December 30, 2014
Revised: February 13, 2015
Accepted: February 13, 2015

First published online
March 3, 2015

**Corresponding author*
Phone: +966-5400-49118 ;
Fax: +012-640-0000;
E-mail: mmmahmed6@yahoo.ca

pISSN 1017-7825, eISSN 1738-8872

Escherichia coli is the most preferred microorganism to express heterologous proteins for therapeutic use, as around 30% of the approved therapeutic proteins are currently being produced using it as a host. Owing to its rapid growth, high yield of the product, cost-effectiveness, and easy scale-up process, *E. coli* is an expression host of choice in the biotechnology industry for large-scale production of proteins, particularly non-glycosylated proteins, for therapeutic use. The availability of various *E. coli* expression vectors and strains, relatively easy protein folding mechanisms, and bioprocess technologies, makes it very attractive for industrial applications. However, the codon usage in *E. coli* and the absence of post-translational modifications, such as glycosylation, phosphorylation, and proteolytic processing, limit its use for the production of slightly complex recombinant biopharmaceuticals. Several new technological advancements in the *E. coli* expression system to meet the biotechnology industry requirements have been made, such as novel engineered strains, genetically modifying *E. coli* to possess capability to glycosylate heterologous proteins and express complex proteins, including full-length glycosylated antibodies. This review summarizes the recent advancements that may further expand the use of the *E. coli* expression system to produce more complex and also glycosylated proteins for therapeutic use in the future.

Keywords: *E. coli*, optimized protein production, biopharmaceuticals, codon usage, molecular chaperones

Introduction

The drug discovery and development process entails the expression of large numbers of recombinant proteins in copious amounts and properly folded as 3D structures. These therapeutic recombinant proteins are generally used for the treatment of various diseases, or are used as a target protein for screening of new and novel drugs. *E. coli* is one of the most desirable host for the expression of several recombinant proteins owing to its rapid growth rate, easier genetic manipulations, and high level of recombinant protein synthesis rates [63, 65]. In a few instances, very high levels of expression are achieved; that is, up to 30% of total cellular protein. *Escherichia coli* was the first expression host that was used for manufacturing a biopharmaceutical, which resulted in the regulatory approval of human insulin in 1982 for the treatment of diabetes. The approval of bovine growth hormone (bGH) in 1994 set a new standard for manufacturing heterologous proteins from *E. coli* for therapeutic use (Table 1). It is even more impressive that

Table 1. List of biopharmaceuticals produced in *E. coli*.

Biopharmaceutical products	Therapeutic indication	Year of approval	Company
Humulin (rh insulin)	Diabetes	1982 US	Eli Lilly
IntronA (interferon α2b)	Cancer, hepatitis, genital warts	1986 US	Schering-Plough
Roferon (interferon α2a)	Leukemia	1986 US	Hoffmann-La-Roche
Humatrope (somatropin rh growth hormone)	hGH deficiency in children	1987 US	Eli Lilly
Neupogen (filgrastim)	Neutropenia	1991 US	Amgen Inc.
Betaferon (interferon β-1b)	Multiple sclerosis	1993 US	Schering Ag
Lispro (fast-acting insulin)	Diabetes	1996 US	Eli Lilly
Rapilysin (reteplase)	Acute myocardial infraction	1996 US	Roche
Infergen (interferon alfacon-1)	Chronic hepatitis C	1997 US	Amgen
Glucagon	Hypoglycemia	1998 US	Eli Lilly
Beromun (tasonermin)	Soft sarcoma	1999 EU	Boehringer Ingelheim
Ontak (denileukin diftitox)	Cutaneous T-cell lymphoma	1999 US	Seragen Inc.
Lantus (long-acting insulin glargine)	Diabetes	2000 US	Aventis
Kineret (anakinra)	Rheumatoid arthritis	2001 US	Amgen
Natrecor (nesiritide)	Congestive heart failure	2001 US	Scios Inc.
Somavert (pegvisomant)	Acromegaly	2003 US	Pharmacia NV
Calcitonin (recombinant calcitonin salmon)	Post menopausal osteoporosis	2005	Upsher-Smith Laboratories
Lucentis (ranibizumab)	Wet age-related macular degeneration	2006 US	Novartis
Preotact (human parathyroid hormone)	Osteoporosis	2006 EU	Nycomed Danmark
Krystexxal (rh urate oxidase, PEGylated)	Gout	2010	Savient
Nivestim (filgrastim, rhGCSF)	Neutropenia	2010	Hospira
Voraxaze (glucarpidase)	Lowering of toxic level of methotrexate conc. in patients with impaired renal function	2012	BTG International
Preos (parathyroid hormone)	Osteoporosis, hypoparathyroidism	2013 EU	NPS Pharmaceuticals

All data obtained from corporate websites and http://www.fda.gov/.
rh, recombinant human; G-CSF, Granulocyte-colony stimulating factor; EU, European union; US, United States of America.

both insulin and bovine growth hormone require oxidative protein folding, and, in addition, insulin is a heterodimer. Success achieved with recombinant insulin and growth hormone highlights the versatility and cost-effectiveness of *E. coli*-based production.

However, in *E. coli*, there is a probability of translational

errors due to the presence of a large number of rare codons in the heterologous gene(s). In the case of therapeutic proteins, these errors, even at low levels, can cause adverse immunogenic responses in humans. These cellular errors during protein translation may impact the tertiary structure and thus affect the biological activity of the recombinant protein. In this review, we will be focusing on the "desired modulation" of key biochemical parameters required to obtain properly folded and overexpressed biologically active therapeutic proteins in *E. coli*.

Codon Usage in *E. coli*

In *E. coli*, codon usage is manifested by the level of cognate aminoacylated tRNAs in the cytosol. Major codons are generally present in genes that are expressed at a high level; on the other hand, rare codons are commonly encountered in genes that are expressed at low levels. Codons that are rare in *E. coli* are found to be abundant in eukaryotic genes [31]. Expression of heterologous genes harboring rare codons can result in translational errors, due to ribosomal stalling at positions where amino acids coupled to rare codon tRNAs have to be incorporated [43]. In addition, translational errors due to the presence of rare codons in heterologous genes might include amino acid substitutions, frame-shift mutations, or premature termination of translation [31, 69]. Kane *et al*. [30] had reported in-frame, two amino acids "hops" at a rare AGA codon. It had also been reported that protein quality is affected to a great extent by codon bias, due to the incorporation of lysine for arginine at the AGA codon [4, 67]. Hence, expression of recombinant proteins, even at very high levels, is of no use if the quality of the protein is compromised as a result of translational errors. The most problematic rare codons in *E. coli* are AGA, AGG, CGG, CGA, CGG (arginine), AUA (isoleucine), GGA (glycine), CUA (leucine), CCC (proline), and AAG (lysine) [85]. It has been documented that rare arginine codons, AGG and AGA, occur at frequencies of 0.14% and 0.21%, respectively, in *E. coli* [31].

Several strategies are employed to circumvent the issue of codon bias in *E. coli*. One approach is to synthesize the whole gene based on codon usage, which is currently a preferred method to improve the expression of heterologous proteins in *E. coli* [59, 61]. However, a major drawback is the high cost associated with total gene synthesis. Another approach involves the site-directed mutagenesis of the

Table 2. Commercially available *E. coli* strains to improve protein solubility.

Vendor	Bacterial strain	Characteristics
Stratagene	BL21 CodonPlus, BL21-CodonPlus(DE3)-RIL BL21 CodonPlus-RP BL21-CodonPlus(DE3)-RP-X BL21-CodonPlus(DE3)-RILP	- *E. coli* cells harbor extra copies of the *argU*, *ileY*, and *leuW* tRNA genes, which recognize rare codons AGA/AGG (arginine), AUA (iso-leucine), and CUA (leucine), respectively. BL21-RIPL carry extra copy of proL tRNA - Enhance expression of heterologous proteins
Novagen	Origami Origami B	- Carry mutations in thioredoxin-reductase and glutathione reductase genes - Improves disulfide bond formation in cytoplasm
Novagen	Rosetta Rosetta-pLysS Rosetta-gami-pLysS	- These *E. coli* strains carry tRNA for rare codons AUA, AGG, AGA, CUA, CCC, GGA - Designed to improve expression of eukaryotic proteins - Assist in disulfide bond formation
Lucigen	C41 (DE3) C43 (DE3)	- Efficient in expressing toxic and membrane proteins - Contain mutation that prevents cell death associated with expression of toxic recombinant proteins
NEB	Shuffle T7 Express	- Constitutively express disulfide bond isomerase (DsbC) - Promotes proper folding of recombinant proteins - Deficient in proteases Lon and OmpT
Genlantis	SoluBL21	- Optimized for expressing insoluble protein in soluble form - Efficient in expressing toxic proteins

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

heterologous gene sequence to generate codons, which reflects the tRNA pool of *E. coli*. However, this process may become very tedious and expensive, whenever several nucleotides need to be modified. Another strategy requires the co-transformation of *E. coli* strains with plasmid harboring a gene encoding the tRNA cognate to the rare codons [12]. By enhancing the copy number of the limiting tRNAs, *E. coli* can be manipulated to match the codon usage frequency in heterologous genes. Several commercial plasmids, such as pRARE, are now available for rare tRNA coexpression in *E. coli*. Moreover, these plasmids contain the p15A replication origin, which enables their maintenance in the presence of the ColE1 replication origin in the *E. coli* expression vectors. There are also several commercial *E. coli* strains available, as listed in Table 2, that harbor plasmids containing gene sequences encoding tRNA for rare codons, such as BL21(DE3) CodonPlus-RIL, BL21(DE3) CodonPlus-RP (Stratagene, USA), and Rosetta (DE3). Several studies have employed these excellent strategies to enhance the expression of heterologous proteins in *E. coli.* [13, 33, 45, 46, 71]. This approach has been successfully used to enhance the industrial production of human recombinant interferon. Co-transformation of *E. coli* BL21 (DE3) with the plasmid harboring human interferon-α2a gene and the *argU* gene that encodes for rare tRNAs for Arg (AGG/AGA) resulted in much higher levels of expression of IFN-α2a, and the recombinant protein constituted about 25% of the total proteins [26]. In another study, it was observed that production of human interferon alpha 2b protein was increased 9.5−11.5-fold by replacing the rare arginine codons with more frequently used codons [78]. The *E. coli* Rosetta (DE3) strain containing plasmid pRARE was used to express different human proteins, and it was shown that the yields of the recombinant protein were increased dramatically for about 35 of the 68 proteins tested [17, 75].

Protein Translation

Efficient translation initiation in *E. coli* requires a ribosomal binding site that includes the Shine-Dalgarno (SD) sequence and a translation initiation codon [69]. The Shine-Dalgarno sequence is generally located 7−9 nucleotides upstream from the initiation codon AUG [62]. It has been observed that translation initiation is more efficient from mRNAs containing the consensus SD sequence AAGGAGG. The secondary structure of ribosomal binding site is very critical for translation initiation, and the translation efficiency is further increased by the presence of a large number of thymine and adenine [35]. The efficiency of translation initiation is also affected by the nucleotide that follows the initiation codon, and it has been observed that adenine is very common in highly expressed genes [74]. Taken together, translation initiation is affected by various factors, including a consensus Shine-Dalgarno sequence, nucleotides upstream of the initiation codon, and the secondary structure of the ribosomal binding site [73]. It had also been reported that mRNA secondary structures that prevent ribosome binding might affect the protein expression to a great extent [11, 14]. It was observed that a single base change compromised the stability of the secondary structure near the SD region and resulted in a 500-fold change in the expression levels of the coat protein of RNA bacteriophage MS2. Park *et al*. [55, 68] have developed a very efficient and simple method for designing 5′-untranslated region (5′UTR) variants for tunable expression in *E. coli*. Since 5′UTR containing the SD sequence and the AU-rich sequence play a significant role in protein translation, the expression levels of recombinant proteins can be manipulated by incorporating simple variations in the 5′UTR. It has been found that secondary structures in mRNA could be disrupted by RNA helicases such as the DEAD protein of *E. coli*. It was shown that expression of the DEAD protein enhanced the expression of β-galactosidase by 30-fold from the T7 promoter, but there was no significant increase in the expression levels from the lac promoter [22]. However, in the absence of DEAD protein, β-galactosidase synthesis from a T7 promoter was found to be 10-fold less as compared with the expression from the lac promoter, even though the transcription rate was 10-fold higher. It was proposed that the DEAD-box protein plays an important role in stabilizing the mRNA [6, 23, 39]. The DEAD-box protein thus can be exploited to additionally improve the expression of genes with suspected problematic mRNA secondary structures.

In the bacterial genome, UAA is the most frequently used termination codon, followed by UGA and UAG. During translation, an error in reading the termination codon leads to extended protein synthesis until another termination codon is encountered in the mRNA. This read-through results in the synthesis of a larger peptide, with several additional C-terminal amino acids. It had been shown that replacing UGA with the UAA terminator in human IFN-α2b resulted in a 2-fold increase in protein expression level [66]. It had been shown that transcription terminators stabilize the mRNA by creating a stem loop structure at the 3′ end of the mRNA [49]. In *E. coli*, stop codon UAA is used more commonly for translation termination. The efficiency

of translation termination can be improved by adding consecutive stop codons or by using a prolonged UAAU stop codon [57]. Translation errors during protein synthesis can cause frame-shift mutations, premature truncation, lower expression, and misincorporation of amino acid, and thus adversely affect the quality of recombinant protein production in *E. coli* [3, 56, 67]. High expression levels and correct apparent molecular mass does not always ensure the translation integrity of the recombinant protein [4, 18, 52].

Molecular Chaperones to Optimize Protein Folding

One major issue in the production of biopharmaceuticals using *E. coli* host is the accumulation of heterologous proteins, mainly as insoluble aggregates in the form of inclusion bodies. Extraction of recombinant protein from these inclusion bodies is a tedious and cumbersome process that requires denaturation and several renaturing steps to obtain a soluble and properly folded recombinant protein. One strategy to improve protein solubility in *E. coli* is the use of molecular chaperones. Inclusion bodies in *E. coli* comprise misfolded aggregates of heterologous proteins. During protein synthesis, molecular chaperones interact with nascent polypeptide chains to prevent aggregation during the folding process. Some chaperones are shown to prevent protein aggregation, while other chaperones promote refolding and solubilization of misfolded proteins [5]. The most widely used and important cytoplasmic chaperones in *E. coli* are DnaK, DnaJ, GrpE, GroEL, GroES, and Trigger factor (Table 3). These chaperones have been used either singly or in combination to improve the protein solubility in *E. coli* [1, 9, 10, 38, 64]. The most efficient chaperone combinations that are widely used to improve protein refolding are GroEL-GroES and DnaK-DnaJ-GrpE [27, 50, 84, 80]. In addition, it was observed that DnaK-DnaJ-GrpE assists in the release of unfolded proteins and GroEL-GroES chaperones prevent the degradation of peptides [72]. Trigger factor has been shown to interact with GroEL and improves GroEL-substrate binding to facilitate protein folding [19]. Other chaperones such as ClpB (Hsp100), in association with DnaK, solubilizes protein and prevents aggregation [2, 47]. Other heat shock proteins, such as IpbA and IpbB, prevent aggregation of heat denatured proteins [20]. For improving the production of recombinant heterologous proteins in *E. coli*, different combinations of molecular chaperones should be analyzed to identify the most efficient one. It has been shown that coexpression of Skp and FkpA chaperones increased the solubility of antibody fragments in *E. coli* [53]. It was revealed that coexpression of GroEL–GroES resulted in production of about 65% of anti-B-type natriuretic peptide single chain antibody (scFv) in soluble form, which was 2.4-fold higher than in the absence of chaperones [40].

Another strategy involves the lowering of the growth temperature to enhance the solubility of recombinant proteins [32]. Growth at a lower temperature decreases the rate of protein synthesis and thus prevents the accumulation of folding intermediates in the cytosol. Moreover, it also reduces the aggregation of protein by preventing inter- and intramolecular hydrophobic interactions, thus minimizing the formation of inclusion bodies [15]. This approach has been very efficient in enhancing the solubility of various therapeutic proteins such as interferon-α-2, human growth hormones, and Fab fragments [24, 79]. In addition, to facilitate the production of soluble recombinant proteins in *E. coli*, it is crucial to optimize several other parameters, such as effect of medium composition, choice of expression vectors, choice of promoters, choice of expression hosts,

Table 3. Application of molecular chaperones to improve protein solubility in *E. coli*.

Chaperones	Plasmids carrying chaperones	Protein expressed	References
GroEL/GroES	pGro7, pGro6, pGro11	Interferon γ, ALDH3A1, pappilomavirus 16 E7 oncoprotein, XynB	[50] [19] [9, 27, 84]
DnaK/DnaJ/GrpE	pKJE7	ALDH3A1, lipase-Lip948,	[9, 19, 80]
GrpE, ClpB	pBB540	IL-6	[47]
Trigger factor	pBAD33-Tig	Fab antibody	[38]
Skp	pAR3	Fab antibody	[38]
sHSP-Lo18	pACYC18	β-Glucosidase	[64]

ALDH3A1, aldehyde dehydrogenase 3 family, member A1; xynB, xylanase B; IL-6, Interleukin-6; Fab, fragment antigen-binding; HSP, heat shock protein.

application of various fusion tags, rate of protein synthesis, inducer concentration, and duration of induction [16, 36, 54].

It has been suggested that the secretion of heterologous proteins into the periplasm of *E. coli* may provide an opportunity to produce complex therapeutic proteins. The Dsb protein family in the periplasm assists in disulfide bond formation, and the oxidizing environment in the periplasm facilitates the proper folding of heterologous proteins. Moreover, the periplasm contains very few host protein and low proteolytic activity, which facilitate the downstream processing to recover high yields of therapeutic proteins [41]. Various therapeutic proteins have been successfully manufactured by periplasmic secretion, such as commercialized Fab fragments (*i.e.*, Leucentis and Cimza) and some full-length aglycosylated antibodies and scFvs [29, 48]. Various secretion signals have been used to target heterologous proteins to the periplasm of *E. coli*, such as Omp A, LamB, PhoA, STII, PelB, and endoxylanase [7]. Yim *et al*. [86] exploited the endoxylanase signal peptide to obtain very high-levels of expression of granulocyte colony-stimulating factor (G-CSF) at 4.2 g/l in the periplasm of *E. coli*. Another study reported high-level periplasmic expression of insulin-like growth factor I (IGF-1), with a yield of 4.3 g/l, by using the LamB signal peptide [28]. It must be noted that besides signal peptides, the efficiency of periplasmic secretion also depends on the *E. coli* strain, promoter strength, and growth temperature. In some circumstances, coexpression of periplasmic chaperones, including disulfide bond oxidase (DsbA), isomerase (DsbC), and peptidylprolyl isomerase, could enhance the production of heterologous proteins in *E. coli* [34]. Reilly and Yansura [58] reported that overexpression of DsbA and DsbC improved the efficiency of the assembly of the light chain and the heavy chain of a full-length antibody in the periplasm, and increased the production from 0.1 to 1.05 g/l [58]. Another study reported that the yield of anti-CD20 scFv was increased, along with improved antigen binding affinity, when it was coexpressed with the periplasmic chaperone Skp [42]. Lee *et al*. [37] reported the development of an efficient expression system for full-length IgG in *E. coli*. Their study demonstrated that modification of the 5′UTR sequence and coexpression of DsbC foldase resulted in very high expression levels of heavy and light chains, and assembly of IgG was also dramatically improved in the periplasm. Under highly optimized condition, fully assembled and functionally active IgG was produced, as high as 362 mg/ml [37]. These studies clearly suggested that through proper engineering of the *E. coli* host, it would be possible to produce therapeutic monoclonal antibodies in *E. coli* in a very cost-effective and timely manner in the near future.

Post-Translational Modifications

E. coli is a favorite microorganism of biotechnologists for the large-scale production of therapeutic proteins [60]. However, the absence of post-translational modification processes in *E. coli* limits its use for the production of recombinant biopharmaceuticals. Various post-translational modifications, including glycosylation and phosphorylation, which are critical for functional activity, do not take place in *E. coli* due to its lack of such cellular machinery [25, 82]. N-Linked glycosylation of proteins is one of the most important post-translational modifications in eukaryotes. Wacker *et al*. [81] identified a novel N-linked glycosylation pathway in the bacterium *Campylobacter jejuni* and also showed the successful transfer of a functionally active N-glycosylation pathway into *E. coli* [81]. *Campylobacter jejuni* harbors *pgl* gene clusters, which are involved in the synthesis of various glycoproteins. By successfully transferring the *pgl* pathway into *E. coli*, various glycosylated proteins were produced in *E. coli*. Although the bacterial N-glycan structure is not similar to that seen in eukaryotes, the molecular engineering of the glycosylation pathway of *C. jejuni* into *E. coli* has paved the way for expressing glycosylated proteins in *E. coli* [77, 51]. Valderrama *et al*. [77] successfully engineered the eukaryotic glycosylation pathway in *E. coli*. Four eukaryotic glycosyltransferases, which include yeast uridine diphosphate-*N*-acetylglucosamine transferases Alg13 and Alg14, mannosyl-transferases Alg1 and Alg2, and bacterial oligosaccharyltransferase PglB from *C. jejuni*, were coexpressed in *E. coli* for synthesizing glycans, which were then successfully transferred to asparagine residues in the target eukaryotic protein [77]. This approach can also be utilized to develop glycoconjugate vaccines against several bacterial pathogens, which could be a more cost-effective and convenient alternative method to presently employed chemical-based methods of vaccine production. Currently, a glycoconjugate vaccine against *Shigella dysenteriae* O1, developed using this technology, has successfully cleared the phase I clinical trials. Initial efficacy and safety studies demonstrated that the glycoconjugate vaccine was safe and also elicited a strong immune response. This novel approach for glycoconjugate vaccine production using the engineered N-linked glycosylation system of *Campylobacter jejuni* can be exploited to produce vaccines against both gram-positive and gram-negative pathogens [8, 18, 21, 44, 83].

Future Perspectives

Although there are several different expression systems such as yeast, mammalian cell lines, transgenic animals, and plants that are currently being used for the production of recombinant proteins, new technological advancements are continuously being made to improve the *E. coli* expression system. Production of heterologous proteins in *E. coli* is always preferred because of the ease of genetic manipulations, well-characterized genome, accessibility of versatile plasmid vector, availability of different kinds of host strains, cost-effectiveness, and very high expression levels as compared with other expression systems. However, there are certain limitations that might hinder the efficient use of the *E. coli* system to overexpress heterologous genes, such as biased codon usage, protein solubility, mRNA stability, and secondary structure. The translational errors due to the presence of rare codons in heterologous proteins may result in amino acid substitutions or frame-shift mutations, which ultimately leads to undesired products. Hence, the codon usage of the recombinant protein plays a critical role in defining its expression levels. Expression of therapeutic proteins in *E. coli* can be enhanced by replacing rare codons in the genes with more favorable major codons. Similarly, the coexpression of genes coding for tRNAs for rare codons could increase the expression levels of therapeutic proteins in *E. coli.* In addition, periplasmic secretion of heterologous proteins offers several advantages, including proper folding, solubility, ease in purification, and higher yield of proteins. Antibody fragments that have been approved for therapeutic use were produced in the periplasm of *E. coli,* suggesting that this approach is commercially viable. Various recent studies have shown that *E. coli* strains can be modified specifically for each therapeutic protein to achieve high product yields as well as high-quality products.

Acknowledgments

This work was supported by the NSTIP strategic technologies program in the Kingdom of Saudi Arabia (Project No. 10-BIO1257-03). The authors also acknowledge financial assistance from the Science & Technology Unit, Deanship of Scientific Research, and Deanship of Graduate Studies, King Abdulaziz University, Jeddah, KSA.

References

1. Arya R, Sabir JSM, Bora RS, Saini KS. 2015. Optimization of culture parameters and novel strategies to improve protein solubility. In Insoluble Proteins: Methods and Protocols. Methods in Molecular Biology series, edited by Dr. *Elena Garcia Fruitos*, Vol. 1258, *Chapter 3*. Pages 45-63; Springer Science and Business Media, New York, USA.
2. Betiku E. 2006. Molecular chaperones involved in heterologous protein folding in *Escherichia coli*. *Biotechnol. Mol. Biol.* **1:** 66-75.
3. Brinkman U, Mattes RE, Buckel P. 1989. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene* **85:** 109-114.
4. Calderone TL, Stevens RD, Oas TG. 1996. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *J. Mol. Biol.* **262:** 407-412.
5. Carrio MM, Villaverde A. 2003. Role of molecular chaperones in inclusion body formation. *FEBS Lett.* **537:** 215-221.
6. Chen R. 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol. Adv.* **30:** 1102-1107.
7. Choi JH, Lee SY. 2004. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **64:** 625-635.
8. Cuccui J, Wren B. 2015. Hijacking bacterial glycosylation for the production of glycoconjugates, from vaccines to humanised glycoproteins. *J. Pharm. Pharmacol.* **67:** 338-350.
9. Cui SS, Lin XZ, Shen JH. 2011. Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948. *Protein Expr. Purif.* **77:** 166-172.
10. de Marco A. 2007. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. *Nat. Protoc.* **2:** 2632-2639.
11. De Smit MH, van Duin, J. 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **87:** 7668-7672.
12. Dieci G, Bottarelli L, Ballabeni A, Ottonello S. 2000. tRNA assisted overproduction of eukaryotic ribosomal proteins. *Protein Expr. Purif.* **18:** 346-354.
13. El-Baky NA, Redwan EM. 2015. Therapeutic alpha-interferons protein: structure, production, and biosimilar. *Prep. Biochem. Biotechnol.* **45:** 109-127.
14. Etchegaray JP, Inouye M. 1999. Translational enhancement by an element example of molecular misreading in Alzheimer's disease. *Trends Neurosci.* **21:** 331-335.
15. Fahnert B, Lilie H, Neubauer P. 2004. Inclusion bodies: formation and utilization. *Adv. Biochem. Eng. Biotechnol.* **89:** 93-142.
16. Feng Y, Xu Q, Yang T, Sun E, Li J, Shi D, Wu D. 2014. A novel self-cleavage system for production of soluble recombinant protein in *Escherichia coli*. *Protein Expr. Purif.* **99:** 64-69.
17. Ferrer-Miralles N, Villaverde A. 2013. Bacterial cell factories for recombinant protein production; expanding the catalogue. *Microb. Cell Fact.* **12:** 113.
18. Fisher AC, Haitjema CH, Guarino C, Celik E, Endicott CE,

Reading CA, *et al*. 2011. Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli. Appl. Environ. Microbiol.* **77:** 871-881.
19. Folwarczna J, Moravec T, Plchova H, Hoffmeisterova H, Cerovska N. 2012. 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 Expr. Purif.* **85:** 152-157.
20. Guzzo J. 2012. Biotechnical applications of small heat shock proteins from bacteria. *Int. J. Biochem. Cell Biol.* **44:** 1698-1705.
21. Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thöny-Meyer L. 2010. Production of glycoprotein vaccines in *Escherichia coli. Microb. Cell Fact*. **11:** 9: 61.
22. Iost I, Dreyfus M. 1994. mRNAs can be stabilized by DEAD-box proteins. *Nature* **372:** 193-196.
23. Iost I, Bizebard T, Dreyfus M. 2013. Functions of DEAD-box proteins in bacteria: current knowledge and pending questions. *Biochim. Biophys. Acta* **1829:** 866-877.
24. Jensen EB, Carlsen S. 1990. Production of recombinant human growth hormone in *Escherichia coli*: expression of different precursors and physiological effects of glucose, acetate and salts. *Biotechnol. Bioeng*. **36:** 1-11.
25. Jenkins N. 2007. Modifications of therapeutic proteins: challenges and prospects. *Cytotechnology* **53:** 121-125.
26. Jeong W, Shin HC. 1998. Supply of the ArgU gene product allows high-level expression of recombinant human interferon-alpha-2a in *Escherichia coli*. *Biotechnol. Lett.* **20:** 19-22.
27. Jhamb K, Sahoo DK. 2012. Production of soluble recombinant proteins in *Escherichia coli:* effects of process conditions and chaperone co-expression on cell growth and production of xylanase. *Bioresour. Technol.* **123:** 135-143.
28. Joly JC, Leung WS, Swartz JR. 1998. Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation. *Proc. Natl. Acad. Sci. USA* **95:** 2773-2777.
29. Jung ST, Kang TH, Kelton W, Georgiou G. 2011. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. *Curr. Opin. Biotechnol.* **22:** 858-867.
30. Kane JF, Violand BN, Curran DF, Staten NR, Duffin KL, Bogosian G. 1992. Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. *Nucleic Acids Res.* **20:** 6707-6712.
31. Kane JF. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6:** 494-500.
32. Khattar SK, Kundu PK, Gulati P, Singh V, Bughani U, Bajpaim M, Saini KS 2007. Optimization and enhanced soluble production of biologically active recombinant human p38 mitogen-activated-protein kinase (MAPK) in *Escherichia coli*. *Protein Peptide Lett.* **14:** 756-760.
33. Kim R, Sandler SJ, Goldman S, Yokota H, Clark AJ, Kim SH. 1998. Overexpression of archaeal proteins in *Escherichia coli*. *Biotechnol. Lett.* **20:** 207-210.
34. Kolaj O, Spada S, Robin S, Wall JG. 2009. Use of folding modulators to improve heterologous protein production in *Escherichia coli. Microb. Cell Fact.* **8:** 9.
35. Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **69:** 101-123.
36. Lebendiker M, Danieli T. 2014. Production of prone-to-aggregate proteins. *FEBS Lett.* **588:** 236-246.
37. Lee YJ, Lee DH, Jeong KJ. 2014. Enhanced production of human full-length immunoglobulin G1 in the periplasm of *Escherichia coli. Appl. Microbiol. Biotechnol.* **98:** 1237-1246.
38. Levy R, Weiss R, Chen G, Iverson BL, Georgiou G. 2001. Production of correctly folded Fab antibody fragment in the cytoplasm of *Escherichia coli* trxB gor mutants *via* the coexpression of molecular chaperones. *Protein Expr. Purif.* **23:** 338-347.
39. Linder P, Daugeron M-C. 2000. Are DEAD-box proteins becoming respectable helicases? *Nat. Struct. Biol.* **7:** 97-99.
40. Maeng BH, Nam DH, Kim YH. 2011. 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 J. Microbiol. Biotechnol.* **27:** 1391-1398.
41. Makrides SC. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60:** 512-538.
42. Mavrangelos C, Thiel M, Adamson PJ, Millard DJ, Nobbs S, Zola H, Nicholson IC. 2001. Increased yield and activity of soluble single-chain antibody fragments by combining high-level expression and the Skp periplasmic chaperonin. *Protein Expr. Purif.* **23:** 289-295.
43. McNulty DE, Claffee BA, Huddleston MJ, Kane JF. 2003. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli. Protein Expr. Purif.* **27:** 365-374.
44. Merritt JH, Ollis AA, Fisher AC, DeLisa MP. 2013. Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnol. Bioeng.* **110:** 1550-1564.
45. Mohammed Y, El-Baky NA, Redwan NA, Redwan EM. 2012. Expression of human interferon-α8 synthetic gene under P(BAD) promoter. *Biochemistry (Mosc.)* **77:** 1210-1219.
46. Mohammed Y, El-Bakym NA, Redwan EM. 2012. Expression, purification, and characterization of recombinant human consensus interferon-alpha in *Escherichia coli* under λP(L) promoter. *Prep. Biochem. Biotechnol.* **42:** 426-447.
47. Nausch H, Huckauf J, Koslowski R, Meyer U, Broer I, Mikschofsky H. 2013. Recombinant production of human interleukin 6 in *Escherichia coli. PLoS One* **8:** e54933.
48. Nelson AL, Reichert JM. 2009. Development trends for therapeutic antibody fragments. *Nat. Biotechnol.* **27:** 331-337.
49. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF. 1987. Stabilization of translationally active mRNA by

prokaryotic REP sequences. *Cell* **48:** 297-310.
50. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli. Appl. Environ. Microbiol.* **64:** 1694-1699.
51. Ollis AA, Zhang S, Fisher AC, DeLisa MP. 2014. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* **10:** 816-822.
52. Overton TW. 2014 Recombinant protein production in bacterial hosts. *Drug Discov. Today* **19:** 590-601.
53. Ow DSW, Lim DYX, Nissom PM, Camattari A, Wong VVT. 2010. Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production. *Microb. Cell Fact.* **9:** 22.
54. Papaneophytou CP, Kontopidis G. 2014. Statistical approaches to maximize recombinant protein expression in *Escherichia coli*: a general review. *Protein Expr. Purif.* **94:** 22-32.
55. Park YS, Seo SW, Hwang S, Chu HS, Ahn JH, Kim TW, *et al*. 2007. Design of 5′-untranslated region variants for tunable expression in *Escherichia coli. Biochem. Biophys. Res. Commun.* **356:** 136-141.
56. Parker J. 1989. Errors and alternatives in reading the universal genetics code. *Microbiol. Rev.* **53:** 273-298.
57. Poole ES, Brown CM, Tate WP. 1995. The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli*. *EMBO J.* **14:** 151-158.
58. Reilly DE, Yansura DG. 2010. Production of monoclonal antibodies in *E. coli*, pp295-308. *In* Shire SJ, Gombotz W, Bechtold-Peters K, Andya J. (eds). *Current Trends in Monoclonal Antibodies Development and Manufacturing*. Springer, New York.
59. Redwan EM. 2006. Optimal gene sequence for optimal protein expression in *Escherichia coli*: principle requirements. *Arab. J. Biotechnol.* **11:** 493-510.
60. Redwan EM. 2007. Cumulative updating of approved biopharmaceuticals. *Hum. Antibodies* **16:** 137-158.
61. Redwan EM, Matar SM, El-Aziz GA, Serour EA. 2008. Synthesis of the human insulin gene: protein expression, scaling up and bioactivity. *Prep. Biochem. Biotechnol*. **38:** 24-39.
62. Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L. 1992. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **6:** 1219-1229.
63. Rodriguez V, Asenjo JA, Andrews BA. 2014. Design and implementation of a high yield production system for recombinant expression of peptides. *Microb. Cell Fact.* **13:** 65.
64. Ronez F, Arbault P, Guzzo J. 2012. 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. *Biotechnol. Lett.* **34:** 935-939.
65. Sahdev S, Khattar SK, Saini KS. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307:** 249-264.
66. Sanchez JC, Padron G, Santana H, Herrera L. 1998. Elimination of an HuIFN alpha 2b readthrough species, produced in *Escherichia coli,* by replacing its natural translation stop signal. *J. Biotechnol.* **63:** 179-186.
67. Seetharam R, Heeren RA, Wong EY, Braford SR, Klein BK, Aykent S, *et al*. 1988. Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. *Biochem. Biophys. Res. Commun.* **155:** 518-523.
68. Seo SW, Yang JS, Cho HS, Yang J, Kim SC, Park JM, *et al*. 2014. Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Sci. Rep*. **4:** 4515.
69. Sørensen HP, Laursen BS, Mortensen KK, Sperling-Petersen HU. 2002. Bacterial translation initiation — mechanism and regulation. *Recent Res. Dev. Biophys. Biochem.* **2:** 243-270.
70. Sørensen HP, Sperling-Petersen HU, Mortensen KK. 2003. A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr. Purif.* **32:** 252-259.
71. Sørensen HP, Sperling-Petersen HU, Mortensen KK. 2003. Dialysis strategies for protein refolding: preparative streptavidin production. *Protein Expr. Purif.* **31:** 149-154.
72. Sorensen HP, Mortensen KK. 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact*. **4:** 1.
73. Sprengart ML, Porter AG. 1997. Functional importance of RNA interactions in selection of translation initiation codons. *Mol. Microbiol.* **24:** 19-28.
74. Stenstrom C, Jin H, Major L, Tate W, Isaksson LA. 2001. Codon bias at the 3′-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene* **263:** 273-284.
75. Tegel H, Tourle S, Ottosson J, Persson A. 2010. Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta (DE3). *Protein Expr. Purif.* **69:** 159-167.
76. Terra VS, Mills DC, Yates LE, Abouelhadid S, Cuccui J, Wren BW. 2012. Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design. *J. Med. Microbiol.* **61:** 919-926.
77. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, *et al*. 2012. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.* **8:** 434-436.
78. Valente CA, Prazeres DMF, Cabral JMS, Monteriro GA. 2004. Translation feature of human alpha 2b interferon production in *Escherichia coli. Appl. Environ. Microbiol.* **70:** 5033-5036.
79. Vasina JA, Baneyx F. 1997. Expression of aggregation prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli cspA* and *tac* promoters systems.

Protein Expr. Purif. **9:** 211-218.
80. Voulgaridou GP, Mantso T, Chlichlia K, Panayiotidis MI, Pappa A. 2013. Efficient *E. coli* expression strategies for production of soluble human crystallin ALDH3A1. *Plos One* **8:** e56582.
81. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, *et al*. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298:** 1790-1793.
82. Walsh G, Jefferis R. 2006. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **24:** 1241-1252.
83. Wetter M, Kowarik M, Steffen M, Carranza P, Corradin G, Wacker M. 2013. Engineering, conjugation, and immunogenicity assessment of *Escherichia coli* O121 O antigen for its potential use as a typhoid vaccine component. *Glycoconj. J.* **30:** 511-522.
84. Yan X, Hu S, Guan YX, Yao SJ. 2012. Coexpression of chaperonin GroEL/GroES markedly enhanced soluble and functional expression of recombinant human interferon-gamma in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **93:** 1065-1074.
85. Yarian C, Marszalek M, Sochacka E, Malkiewicz A, Guenther R, Miskiewicz A, Agris PF. 2000. Modified nucleoside dependent Watson-Crick and wobble codon binding by tRNALysUUU species. *Biochemistry* **39:** 13390-13395.
86. Yim S, Jeong K, Chang H, Lee S. 2001. High-level secretory production of human granulocytes-colony stimulating factor by fed-batch culture of recombinant *Escherichia coli*. *Bioprocess Biosyst. Eng.* **24:** 249-254.