MINI-REVIEW

Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements

Chung-Jr Huang · Henry Lin · Xiaoming Yang

Abstract Nearly 30% of currently approved recombinant therapeutic proteins are produced in *Escherichia coli*. Due to its well-characterized genetics, rapid growth and high-yield production, *E. coli* has been a preferred choice and a workhorse for expression of non-glycosylated proteins in the biotech industry. There is a wealth of knowledge and comprehensive tools for *E. coli* systems, such as expression vectors, production strains, protein folding and fermentation technologies, that are well tailored for industrial applications. Advancement of the systems continues to meet the current industry needs, which are best illustrated by the recent drug approval of *E. coli* produced antibody fragments and Fc-fusion proteins by the FDA. Even more, recent progress in expression of complex proteins such as full-length aglycosylated antibodies, novel strain engineering, bacterial N-glycosylation and cell-free systems further suggests that complex proteins and humanized glycoproteins may be produced in *E. coli* in large quantities. This review summarizes the current technology used for commercial production of recombinant therapeutics in *E. coli* and recent advances that can potentially expand the use of this system toward more sophisticated protein therapeutics.

Keywords *Escherichia coli* · Recombinant therapeutics production · Aglycosylated antibody · *E. coli* N-linked glycosylation · Cell-free systems

C.-J. Huang (✉) · H. Lin · X. Yang (✉)
Cell Sciences & Technology, AMGEN Inc,
Thousand Oaks, CA, USA
e-mail: chuangjr@amgen.com

X. Yang
e-mail: xyang@amgen.com

Introduction

Proteins have been considered important therapeutic entities since the early 1900s when major resources were only available from plants and animals. With the advent of recombinant DNA technology in the 1970s, however, it was discovered that recombinant protein therapeutics could be produced by *E. coli* in a robust and economic manner. In the early 1980s, the FDA approved the first *E. coli*-produced recombinant insulin for diabetes treatment, opening the door and creating a model for developments of other recombinant therapeutics. Since then, in addition to *E. coli*, different expression hosts such as yeast, filamentous fungi, insect cells and mammalian cells have become available for producing different or more complex recombinant therapeutics such as monoclonal antibodies (mAbs).

Today, more than 151 unique recombinant therapeutics have been approved by the FDA and/or by the European Medicines Agency for different clinical indications. One-third of these approved protein therapeutics are produced in *E. coli*, indicating that it is a major workhorse for recombinant therapeutic production (Table 1) [15, 25, 29]. Compared to other recombinant microorganisms, *E. coli* remains the most attractive because of its well-characterized genetics, versatile cloning tools and expression systems, and the fact that it has been successfully used to express vastly different proteins. It is also advantageous to use *E. coli* for industrial scale production because of its rapid growth, low-cost media, ease of scale-up and capability to produce therapeutics with high yield and quality.

However, there are also some limitations to using *E. coli* as an expression host. These include the inability to perform certain posttranslational modifications (such as glycosylation) and insufficiencies in proteolytic protein maturation and disulfide bond formation [48]. Such

Table 1 Recent approved protein therapeutics using *E. coli* as an expression host

Generic name/protein (brand name)	Indications	Approved date, place and company
Ranibizumab (Lucentis)	Wet type age-related macular degeneration	2006 US, 2007 EU, Genentech
Somatropin (Accretropin)	Growth hormone deficiency; Turner syndrome	2008 US, Cangene
Certolizumab pegol (Cimzia)	Crohn's disease	2008 US, 2009 EU, UCB[a]
PEG interferon alfa-2b (PegIntron)	Chronic hepatitis C infection	2008 US, Schering-Plough
Romiplostim (Nplate)	Chronic immune thrombocytopenic purpura	2008 US, Amgen
Interferon beta 1b (Extavia)	Multiple sclerosis	2008 EU, 2009 US, Novartis
Pegloticase (Krystexxa)	Chronic gout	2010 US, Savient Pharms

Data were collected from http://www.fda.gov and http://www.ema.europa.eu; [a] Union Chimique Belge

drawbacks prohibit *E. coli* from expressing some complex and important therapeutics, such as mAbs, where both correct folding and glycosylation play crucial roles in their biological activities. Nevertheless, *E. coli* continues to be used to produce other important recombinant therapeutics. In this review, we will discuss important considerations for producing protein therapeutics in *E. coli*, with an emphasis on commercial production. Recent advancements in protein expression in *E. coli*, such as complex protein production, bacterial N-linked glycosylation, novel strain engineering and creation of *E. coli* cell-free systems, will also be covered.

Expression considerations for recombinant therapeutics in *E. coli*

Host

Escherichia coli K12 and its derivatives are the main strains used in recombinant therapeutic production in the biotech industry. A big advantage using *E. coli* K12 was given by the National Institutes of Health when it made this strain the standard and provided guidelines for safety. In addition, large-scale industrial production with *E. coli* requires approval by the local Biosafety Authority, which may be reluctant to approve other *E. coli* strains without the same safety level as *E. coli* K12. Common K12 derivatives used in the biotech industry include *E. coli* RV308 and W3110 [14, 28].

To further improve recombinant protein production, several studies have genetically modified K12 strains to reduce acetate accumulation during the cell growth. Acetate is an undesirable by-product formed during aerobic fermentation, and its accumulation at a high concentration has negative effects on cell growth and recombinant protein production [19]. Reducing glucose consumption and redirecting carbon flow from acetate formation pathway are the two major strategies used to engineer low acetate production strains. For example, by deletion of the *pstHI* operon in *E. coli* GJT 001, Wong and coworkers increased protein

production yield by 25-fold in a batch bioreactor. This high productivity was attributed to low acetate accumulation, although cell growth was compromised [124]. The *E. coli* B strain was first designed and engineered by Studier and Moffatt, and has been a model strain for studying phage sensitivity, restriction-modification systems, bacterial evolution and recombinant protein expression in laboratories as well as in the biotech industry [100]. Since its original use, the B strain has been modified and engineered to create additional strains, such as BL21, C41 and C43, that have different capabilities to express varying types of proteins. B strain and its derivatives have several advantages such as low acetate accumulation under high glucose concentrations, specific protease deficiencies and high outer membrane permeability, making them desirable hosts for expression of protein therapeutics [82, 107, 112].

Some *E. coli* strains have also been designed to tailor for special needs. Recently, Caparon and coworkers have shown that deletion of fours genes, *dppA*, *oppA*, *malE* and *ompT*, of the *E. coli* strain BC50 can significantly reduce the level of host cell contaminants in recombinant Apolipoprotein A 1 Milano production, with no adverse effects on fermentation productivity [10]. To improve the secretion of periplasmic recombinant proteins, an *E. coli* strain with compromised outer membrane structure has been engineered to increase the extracellular secretion of antibody fragments [72]. Details of examples of novel strain engineering for improved therapeutic production will be discussed in the latter sections of this review.

Vector design

Optimal gene transcription is normally a function of both gene dosage (plasmid) and promoter functionality. The productivity of recombinant protein is known to be affected by plasmid copy number and its structural and segregational stability. Choosing the optimal plasmid copy number is critical. Too low of a copy number will result in a low mRNA pool, as well as low protein productivity. A high copy number generally leads to high productivity; however,

Table 2 Commonly used promoters for protein therapeutics production in *E. coli* [38, 85, 111, 112]

Promoters	Induction methods	Key features
lac	IPTG	Relative low-level expression; titratable; leaky
tac, *trc*	IPTG	High-level expression; titratable; leaky
T7	IPTG	Very high-level expression; titratable; leaky
Phage λ pR, pL	Temperature shift	High-level expression; tight control
PhoA	Phosphate depletion	High-level expression; tight control; media limitation
p*lux*	Homoserine lactone	High-level expression; tight control; inexpensive inducer

it also tends to impose metabolic burdens on cells. The plasmid copy number depends largely on the replication of origin, which dictates either flexible or rigid control over a plasmid. Both high copy number plasmids (e.g., pUC, 500–700 copies) and medium copy number plasmids (e.g., pBR322, 15–20 copies) have been used for therapeutic production in *E. coli* [14, 38].

For large-scale therapeutic production, the use of high copy number plasmids is not desirable. First, in order to maintain high copy number plasmids in cells, selection markers such as antibiotics are usually required and included in the growth media. However, the FDA discourages the use of antibiotics in clinical manufacturing, as they cause allergic reactions in some patients and raise concerns over the development of drug-resistant pathogens. Second, high copy number plasmids have been shown to possess higher segregational instability, especially in the absence of antibiotics [7, 27]. This leads to the overgrowth of plasmid-free cells, resulting in a significant loss of protein productivity. Lastly, production and maintenance of high copy number plasmids in cells require tremendous energy, which will inevitably reduce cell growth and protein synthesis. Therefore, high copy number plasmids are normally used for the expression of recombinant therapeutic genes for gene therapy [11].

An ideal expression system is critical for high-level therapeutic production in *E. coli* to allow tightly regulated and efficient transcription. Choosing an appropriate vector system is largely dictated by the strength and the control of its promoter. To ensure high-level expression, a promoter with certain characteristics must be incorporated into the plasmid [38]. For example, the promoter has to be strong to allow the recombinant protein production to account for 10–30% or more of total cellular protein. It should also be tightly regulated with limited basal expression in the non-induced state. Leaky expression can cause metabolic burdens on the cells during the growth period by diverting the carbon and energy source to premature protein formation. This situation can be detrimental when the expressed protein is highly toxic. In addition, some promoters must be used within specific *E. coli* strains to achieve optimal protein expression. Other important considerations are that the induction method should be simple and cost-effective, and,

in most cases, the induction must be independent of the media components.

Multiple promoters have been successfully used to produce different recombinant proteins in the past (Table 2). Among them, *lac* and its derivatives, *tac* and *trc*, are commonly used both in basic research and industry [29, 38]. The synthetic *tac* and *trc* promoters are stronger than *lac*, and all of them are induced by isopropyl-β-D-thiogalacto-pyranoside (IPTG). Since IPTG is used to derepress the *lac* repressor, the expression level of recombinant protein is titratable by varying the IPTG concentration in the media. IPTG, however, is expensive and toxic to some *E. coli* strains. To circumvent these drawbacks, a thermosensitive *lac* repressor mutant is available to induce protein expression by shifting temperature instead of using IPTG [3].

T7 promoter from T7 bacteriophage is another powerful system that is widely used for recombinant protein expression. It is specifically recognized by T7 bacteriophage RNA polymerase, which can elongate DNA chains five times faster than *E. coli* RNA polymerase. Typically, the chromosome of the host strain contains a prophage (λDE3) encoding T7 bacteriophag RNA polymerase that is under the control of a *lac* promoter derivative, L8-UV5. This new version of the *lac* promoter is less dependent on the cyclic AMP level and is less sensitive to glucose repression. Therefore, the T7 system is generally considered a stronger protein expression system than the *lac* system. It can lead to very high recombinant protein expression, and production of up to 50% of the total cell protein has been reported, mostly in the form of inclusion bodies [7, 94]. Nevertheless, basal expression is a commonly found problem in the T7 system. Despite the fact that co-expression of T7 lysozyme using pLysS or pLysE plasmids can improve the control over the T7 promoter, expression of T7 lysozyme can also increase cell stress and reduce recombinant protein yield [107]. Although the T7 system can lead to a very high level of protein expression, it is not commonly used in the industry.

A thermo-regulated promoter system, such as bacteriophage lambda pL and/or pR promoters, has also been used for production of many therapeutic proteins. Inclusion of the temperature-sensitive cI857 repressor cassette in the plasmid allows control of the pL and pR promoter by

changing the growth temperature. The cI857 repressor blocks transcription at 28–37°C and is inactivated at 42°C. Therefore, recombinant gene expression is induced when temperature is shifted to 42°C. Notably, many human therapeutics, such as interferon-γ, insulin, tumor necrosis factor and granulocyte colony-stimulating factor, have been successfully produced by this system with high yield [111]. This system is especially suitable for large-scale industrial production because it is highly productive, easy to operate and scale up, and, importantly, it minimizes culture contamination. On the other hand, high temperature induction may have other drawbacks, such as reduced cell growth and increased cell stresses (heat shock response and SOS response) in addition to the stress caused by the recombinant protein expression. Together, these stresses may compromise protein yield and quality. The lambda pL and/or pR promoter can also become constitutive at lower temperatures, making this system ideal for production of soluble or proteolytically susceptible proteins [70].

Nutritionally inducible promoters, such as *phoA* and *trp*, are also commercially available and used in recombinant therapeutic production. The *phoA* promoter has been used to successfully express full-length antibodies and antibody fragments in *E. coli* [97]. In these cases, media composition should be carefully designed in both batch media and the feed to ensure phosphate depletion and protein induction. Normally, the induction will occur over a broader range of time compared to other promoters [83]. Another robust and tightly regulated promoter is the *lux* system. By taking the quorum sensing elements from *Vibrio fisheri*, gene expression is controlled by the *luxI* promoter, and transcription is activated by addition of the autoinducer, Acyl homoserine lactone (AHL). AHL is typically used at 1/1,000 of the concentration of IPTG in other systems, making it a very cost-effective system at an industrial scale [108].

Protein translation

Protein translation in *E. coli* can be divided into four phases: initiation, elongation, termination and ribosome cycling. In most cases, translation initiation is a rate-limiting step for protein biosynthesis. The efficiency is determined by the sequence and structure of the translation initiation region (TIR) at the 5′ end of each mRNA [87, 90]. The TIR is composed of four different sequences: (1) the Shine-Dalgarno (SD) sequence, (2) the start codon, (3) the spacer between the SD and the start codon, and (4) translational enhancers located at the upstream of the SD and/or downstream of the start codon. Modulation of TIR sequences has been shown to improve or control recombinant protein expression in *E. coli* [1, 98, 123]. Vimberg et al. [119] demonstrated that a six-nucleotide SD (AGGAGG) sequence is more efficient than other shorter or

longer SD sequences in expression of recombinant green fluorescent protein (rGFP). In this study, incorporation of an A/U-rich upstream enhancer further improved rGFP expression by 13-fold. Modifying the TIR region to avoid possible mRNA secondary structures that reduce the accessibility of the SD and/or the start codon is also critical to achieve optimal protein expression. For example, a single base mutation at the SD region could decrease the expression of RNA bacteriophage MS2 coat protein by 500-fold [103]. Exposing the AUG start codon from a base-paired mRNA structure significantly improved the translation efficiency of IL-10 in *E. coli* and resulted in a 10-fold increase in IL-10 production compared to the wild-type genes [133]. Recently, a biophysical model of translation initiation was developed that can design synthetic ribosome binding sites targeting different translation initiation rates (http://salis.psu.edu/software/). This technology enables rational control and fine-tuning in recombinant protein expression [87]. This precision is especially important for recombinant protein secretion as a high translation rate may overwhelm the secretory apparatus, leading to low production yields. In fact, for enhanced recombinant protein secretion in *E. coli*, optimizing, as opposed to maximizing, the translation level usually leads to high-level secretion of recombinant therapeutics [97, 98].

The stability (half-life) of mRNA also affects the protein expression rate in *E. coli*. Degradation of mRNA in *E. coli* is mediated by several different RNases, including endonucleases (RNase E, K and III) and 3′ exonucleases (RNase II and polynucleotide phosphorylase). The enzymatic activities of these RNases also depend upon the growth conditions [12, 38]. For high-level recombinant protein expression, two common strategies have been used to improve mRNA stability: the introduction of protective elements at two ends of the mRNA and the inhibition of RNase activities by strain engineering and by manipulating growth conditions [6]. It has been shown that adding the 5′ UTR of *ompA* can prolong the half-life of a number of heterologous mRNAs in *E. coli* by incorporating a stem loop structure at or near the 5′ terminus of the mRNA [67]. Fusing the transcription terminator of the *penP* gene from *Bacillus thuringiensis* to the cDNA of human IL-2 significantly enhanced the mRNA stability and improved IL-2 production in *E. coli* [67]. Lopez and colleagues showed that the C-terminal region of the *rne* gene, which encodes RNase E, is required for mRNA degradation. They found that mutation of *E. coli rne* can increase mRNA stability and improve recombinant protein yields [63]. Strains containing this mutation are currently commercially available.

In addition to the TIR sequence and mRNA stability, species-specific variation in codon usage (codon bias) between *E. coli* and other organisms can also affect protein translation. Expression of heterologous genes containing

rare codons can lead to growth arrest, premature translational termination and increased frameshifts, deletions and misincorporations in the recombinant proteins, especially if the rare codons are clustered together [30, 31]. One method to overcome codon bias is to synthetically optimize the gene sequence with codons preferably used by *E. coli*. Several design algorithms are available to optimize protein expression in *E. coli* and other hosts [79, 118]. For example, the expression of a soluble scFv (against hepatitis B antigen) in the periplasm was increased more than 100-fold in a codon-optimized sequence compared to the original gene [109]. Supplementing cognate tRNA for rare codons is another effective strategy to remedy codon bias. The *E. coli* Rosetta (DE3) strain contains a plasmid (pRARE) encoding tRNA for codons that are commonly used by eukaryotes but rarely used in *E. coli*. Using this strain to express different human recombinant proteins, Tegel et al. [106] showed that protein yields were increased for 35 of the 68 tested proteins. This strain is especially more efficient at expressing proteins that are difficult to express in *E. coli* BL21 (DE3). The use of the pRARE plasmid is a practical method to improve the yield of heterologous proteins in *E. coli* and has been demonstrated in several studies [30, 33, 37].

Cytoplasmic expression

Inclusion bodies

Production of functional proteins in *E. coli* requires a delicate balance between DNA transcription, protein translation and protein folding. High-level expression of recombinant protein in *E. coli* can produce over 30% of total cellular protein, where folding chaperones and modulators are highly titrated. In addition, correct folding of many proteins requires disulfide bond formation and/or glycosylation, which are absent in the *E. coli* cytoplasm. Therefore, expression of many human therapeutic proteins in *E. coli* generates high amounts of unfolded and misfolded proteins in the cytoplasm, where they tend to aggregate and form inclusion bodies (IBs).

Despite the fact that IBs are biologically inactive, many commercial and developmental therapeutics, such as interferons, interleukins and Fc-fusion proteins, are produced as IBs because of the multiple advantages of these protein aggregates [29]. High-yield production and versatility to express different proteins are two advantages associated with IB formation [65]. IBs are also stable protein aggregates and are resistant to protease activities in vivo. In addition, proteomic analysis showed that IBs are relatively homogeneous in composition, and, in some cases, the recombinant protein can account for more than 90% of the total imbedded polypeptides [115]. During the downstream processing, IBs can be easily isolated after cell disruption, and the resultant IB paste can be stored frozen for several months, providing manufacturing flexibility. Together, these characteristics allow IBs to be produced at high yield, as well as isolated and purified with simple and minimal efforts. However, this method also has its downside. Especially, the refolding of IBs to active protein represents a challenge, because efficient and high-yield refolding requires considerable optimization for each target protein. Resolubilization of IBs using chaotropic agents may also affect the integrity of the refolded proteins [86]. Even so, acceptable recovery usually can be achieved at large industrial scale by using established strategies [18, 60].

Soluble proteins

Production of soluble and bioactive protein without the requirement of refolding can also be achieved in the *E. coli* cytoplasm. Two strategies, reduced growth temperature and improved folding environment, are commonly used to enhance the solubility of recombinant protein without changing its protein sequence. Lowering the growth temperature decreases the rate of protein synthesis and prevents accumulation of folding intermediates in the cytoplasm. It also decreases protein aggregation by reducing both inter- and intra-molecular hydrophobic interactions that facilitate IB formation [21]. This method has been shown to be effective in improving the solubility of a number of different proteins including interferon α-2, Fab fragments and human growth hormone (GH) [39, 114].

Cytoplasmic folding modulators, such as folding chaperones (e.g., DnaK and GroEL), holding chaperones (e.g., IbpA and B) and disaggregating chaperone (ClpB), play important but distinctive roles in maintaining correct protein folding. Coexpression of these chaperones has been reported to improve the solubility of different recombinant proteins [53]. For example, with coexpression of GroEL/GroES, 65% of total produced anti-B-type natriuretic peptide single chain antibody (scFv) was soluble, which is 2.4-fold higher than in the strain without coexpressed chaperones [66]. However, chaperone coexpression does not guarantee improved protein solubility. It could also cause an extra-metabolic burden to the cells, resulting in low protein productivity. Different *E. coli* strains have also been developed to improve protein folding and solubility in the cytoplasm. Disruption of two genes, *txrB* (thioredoxin reductase) and *gor* (glutaredoxin reductase), in *E. coli* improved the folding of recombinant tissue plasminogen activator by facilitating disulfide bond formation and isomerization in the cytoplasm. Overexpression of the periplasmic disulfide bond isomerase (DsbC) in the cytoplasm of the same strain further enhanced disulfide bond formation [8]. Despite all of the progress made, soluble expression of

recombinant therapeutics in the cytoplasm still can lead to protein misfolding, relatively low protein yield, laborious downstream processing and susceptibility to proteolytic activities [29]. Therefore, if a soluble protein is desired, soluble periplasmic expression is the preferred route in industry.

Periplasmic and extracellular secretion

Periplasmic secretion

Secretion of recombinant proteins into the periplasm of *E. coli* offers an attractive approach to produce complex and/or large protein therapeutics. Disulfide bonds are introduced by the Dsb protein family in the periplasm, creating an oxidizing environment to facilitate proper folding of recombinant proteins. The cleavage of the signal peptide following the translocation is more likely to yield an authentic N-terminus of the expressed protein. In addition, compared to the cytoplasm, the periplasm has a lower concentration of host cell proteins and lower proteolytic activities, so purification can be easily implemented [67]. Therefore, a variety of different recombinant proteins is produced by periplasmic secretion, including commercialized Fab fragments (Leucentis® and Cimza®) and some clinical therapeutics, such as full-length aglycosylated antibodies and scFvs [45, 74].

Recombinant proteins are typically exported to the periplasm via one of three pathways that utilize the type II secretion system of *E. coli*: Sec-dependent, SRP-dependent and twin-arginine translocation (TAT) [71]. The targeting of recombinant protein to each pathway is normally determined by the secretion signal fused to the N-terminus of the protein. Different Sec-dependent secretion signals have been successfully used for recombinant therapeutic secretion, including OmpA, LamB, PhoA, STII, endoxylanase from *Bacillus* sp. and PelB from *Erwinia carotovora* [16]. For example, high-level periplasmic expression of granulocyte colony-stimulating factor (G-CSF) at 4.2 g/l has been reported by using the endoxylanase signal peptide, and the secreted insulin-like growth factor I (IGF-1) reached 4.3 g/l when utilizing the LamB signal peptide [43, 129]. In addition to the signal peptide, the efficiency of protein secretion also depends on the host strain, promoter strength, cultivation temperature and type of protein to be secreted. Therefore, in most cases, a trial-and-error approach is required to optimize protein secretion.

Despite many successful studies, high-level secretion usually overwhelms the translocation machinery, as well as folding capacity in the periplasm, leading to IB formation and misfolding of recombinant proteins. In these cases, coexpression of periplasmic chaperones such as disulfide bond oxidase (DsbA), isomerase (DsbC) and peptidylprolyl isomerase (Skp, FkpA and SurA) could improve the accumulation of the functional proteins [53]. Overexpressing DsbA and DsbC has been shown to increase the assembly efficiency of the light chain and the heavy chain of a full-length antibody in the periplasm from 15 to 95%, as well as to improve the production titer from 0.1 to 1.05 g/l [83]. Improved yield coupled with enhanced antigen binding affinity of anti-CD20 scFv was achieved when it was coexpressed with Skp [69]. Another feasible strategy is to enhance the solubility of a recombinant protein before its transport. This can be accomplished by coexpression of cytoplasmic folding chaperones (DnaK and DnaJ) to prevent aggregation, or export through the TAT system, which only recognizes folded proteins as substrates [17, 78].

Extracellular secretion

Extracellular secretion of recombinant proteins not only has the advantages inherited from periplasmic expression, it also displays additional attractive features in bioprocessing. Low native protein concentration, decreased protease activity and far less endotoxin in the extracellular space ensure easy purification and better protein quality. Secretion of recombinant proteins out of the periplasm also improves the production capacity and alleviates some stresses derived from overexpression of recombinant proteins. Several approaches have been explored to facilitate recombinant protein secretion in *E. coli* in the past, which include a secretory fusion partner from *E. coli* type I and V secretion systems. Hemolysin toxin (HlyA) can be directly secreted from cytoplasm to the extracellular milieu using the hemolysin transport system, which forms a protein channel across the inner and outer membranes [71]. By fusing the HlyA secretion signal at the C-terminus, recombinant IL-6 and scFv were successfully secreted into the culture medium at concentrations of 2 and 7 mg/l, respectively [16, 24]. Autotransporters, such as AIDA-1, have also been used for the secretion of recombinant proteins used in vaccines, such as cholera toxin and OspA, an outer membrane lipoprotein A of *Borrelia burgdorferi* [44]. To identify more secretory fusion partners, Qian and colleagues analyzed the secretome of *E. coli* BL21 and found 22 potential proteins, of which OsmY, an osmotically inducible protein, was the most promising. OsmY-Leptin was excreted into the culture medium at the concentration of 0.25 g/l, which was the highest titer among the screened proteins, in a fermentor culture [80]. This approach, however, required removal of the fusion protein in the latter purification process.

Another strategy is to modify the outer membrane structure to promote non-specific release of periplasmic proteins. Deletion of lpp, an outer membrane lipoprotein, significantly increased the outer membrane permeability, while showing limited growth defects [75]. It has been

demonstrated recently that an lpp mutant can efficiently secrete antibody fragments at yields over 2.2 g/l in 10-l fermentors [72, 130]. The reduced guanosine tetraphosphate formation in *relA* and *spoT* mutants also affected membrane flexibility, and increased secretion of different recombinant proteins in these strains has been reported [96]. In addition to strain engineering, other parameters such as medium composition and cell growth rate also have impacts on protein secretion. For example, supplementation of 2% glycine and 1% Triton X-100 was suggested to change the morphology and integrity of the cell membrane, and this addition increased the excretion of the scFv-TNF-α fusion protein from 0.3 to 50 mg/l [125]. Different growth rates were also found to influence different degrees of periplasmic leakage due to changes in fatty acid composition in the phospholipids or decreases in the amount of outer membrane proteins [95]. In one study, the partitioning of a Fab fragment between the periplasm and supernatant was dramatically affected by the glycerol feeding rate, and the concentration of secreted Fab can reach over 1 g/l [5]. Over the past 2 decades, significant progress has been made in understanding extracellular protein secretion in *E. coli* and its applications in therapeutics production. While it represents an attractive opportunity for the industry, however, efficient extracellular secretion of recombinant proteins across *E. coli* outer membranes remains a challenge and often produces relatively low yields when compared to IB formation and periplasmic secretion.

Fermentation processes

To produce recombinant proteins in large quantities, fermentation technology is generally applied to increase cell density and protein productivity. Fermentation provides control over key chemical, physical and biological parameters that affect cell growth, as well as recombinant therapeutic protein production. These include, but are not limited to, temperature, dissolved oxygen (DO) level, pH and nutrient supply. A robust industrial fermentation process would also need to consider the composition and cost of the media, feeding strategies and scale-up process. If possible, the processes should also be designed to meet FDA guidelines, such as implementation of Process Analytical Technology (PAT) and Quality by Design (QbD), for process characterization and validation. In this section, important components for industrial fermentation are discussed.

Media

The media used to cultivate *E. coli* usually require several essential components, such as a carbon source, nitrogen source, essential salts, minerals and some growth factors, in order to reach an optimal cell density [93]. In general, three types of media, chemically defined (CD), semi-defined and complex medium, are used to support bacteria growth. CD medium is composed of chemicals of known identities and concentrations, while complex medium contains ingredients of natural origin, such as yeast extract and protein hydrolysate, in which the composition is not completely known. Semi-defined medium is constituted of mostly defined components with very few complex components. To have high cell growth as well as high recombinant protein production, the concentration of each component in these three different types of media must be carefully formulated to contain not only all necessary components, but also the optimal concentrations to avoid growth inhibition [57]. Currently, semi-defined and complex media are popularly used in industry because they offer flexibility and enable both high cell density and protein yields in most production processes.

A plethora of different types of therapeutics, including growth factors, antibody fragments and Fc-fusion proteins, have been produced using either semi-defined or complex media [14, 28, 126]. The use of protein hydrolysate and yeast extract in the semi-defined and complex media can also significantly reduce the cost of raw material when compared to CD medium. In some studies, these components also help cells to utilize acetic acid during carbon limitation and enhance recombinant protein production, particularly during high cell density fermentation [110]. However, fermentation using complex ingredients can show inconsistent performance that affects product yield and quality because of the lot-to-lot variation associated with poorly defined components. This issue is highly undesirable for protein therapeutics because processes are considered an integral part of product definition. In this case, using CD medium for commercial fermentation becomes a practical alternative. Despite CD medium being generally known for slow growth and/or low productivity, recent studies have shown that growth in CD medium can reach a similar growth profile and protein titer as its complex medium counterpart (unpublished data). Zhang and Greasham have also illustrated several advantages of using CD medium for commercial therapeutics manufacturing. These benefits include enhanced process consistency, improved process control and simplified protein purification, which will ultimately provide a high degree of assurance of process reproducibility and product quality [132].

Optimization of medium components for enhanced protein therapeutic production is also a common practice in industry. One simple way to accomplish this goal is to modify the published media recipes, when high cell growth and protein expression have been demonstrated. An alternative strategy, which is intensively used in industry, is to use statistical analysis, including design of experiments (DOE),

to evaluate effects of different components and their interactions on important responses, such as cell growth, protein titer and quality. Since traditional medium optimization is labor-intensive, application of DOE can significantly reduce the number of experiments, efficiently evaluate the component effects and, in some cases, predict optimal medium composition [122]. Sometimes, identification of essential medium components can be carried out in an *E. coli* chemostat culture by addition of different components at varied concentrations. Based on the corresponding cellular responses, the importance of each component can be experimentally determined [128].

Batch fermentation

Batch fermentation is an easy way to culture cells to reach high cell density in a very short time. However, due to its low productivity compared to fed-batch culture, it is usually used to acquire a small amount of protein for the purposes of protein characterization or toxicology study. If a therapeutic protein is to be used in a low-dose therapy or is an orphan drug, a batch process is a good choice to ensure efficient and reliable production [126]. In most cases, however, it is not a commonly used process for industrial therapeutic manufacturing.

Fed-batch fermentation and feeding strategies

Unlike batch fermentation, a fed-batch protocol can achieve high cell density culture (HCDC) by continuously providing required nutrients to sustain and control cell growth to reach high protein productivity. After the batch phase, which is usually indicated by sharp DO and pH increases, highly concentrated nutrients are fed into the bioreactor. In most cases, feeds are split into two categories, the nitrogen source and the carbon source, which usually contains other necessary minerals and trace elements. Because *E. coli* generates several by-products, such as acetate, that have negative effects on cell growth and protein production, an optimal feeding procedure must be applied to control the concentration of these by-products and maintain proper cell growth and protein production [19]. Therefore, several different feeding strategies have been developed for HCDC, which include pH–stat control, DO-stat control, and feeding based on a specific growth rate and on the limitation of an essential substrate such as glucose.

In pH–stat control, a feed is activated when the pH of the medium increases, indicating nutrient depletion and cell death. Kim and coworkers have used pH–stat control to produce human glucagon-like peptide 1, resulting in a yield of 11.3 g/l [50]. It has also been used for the production of cancer vaccines, such as NY-ESO-1, Malan-A and SSX2, and showed high productivities [33, 64]. This strategy, however, reflects the starvation of the cells instead of cell growth. It tends to be a conservative feeding method and results in a growth rate below the threshold for acetate accumulation [19]. In DO-stat control, the feeding rate is controlled by the DO level in the medium. This strategy provides flexibility in manipulating the cell growth rate of HCDC by increasing aeration and oxygen supply to the bioreactor. Production of human antibody fragment (Fab′) in *E. coli* using DO-stat control has been reported to reach over 60 mg/l in a 300-l pilot scale bioreactor [84]. Chan and coworkers also used a similar feeding strategy to produce a bivalent antibody fragment in *E. coli*. Interestingly, in this study, by lowering DO level near to zero during the fermentation, the production titer can reach over 2 g/l [14].

Exponential feeding by controlling the cell specific growth rate (μ) can also provide desirable metabolic regulation and promote high protein production in HCDC. To keep μ at a pre-determined level, a feed-forward exponential feeding strategy is normally used, where the amount of nutrient required for cells to reach the desired μ is calculated in advance, and the feed is added accordingly during the process [54, 127]. Automatic control of the μ in the fed-batch fermentation has also been developed, where on-line data such as oxygen utilization rate and culture weight are used to calculate the feeding rate [58]. Ideally, high μ translates into high productivity and less process time. By controlling μ at a maximum attainable value of 0.55 h^{-1} at the beginning of the fed-batch phase and 0.4 h^{-1} during the induction phase, Babaeipour and colleagues showed that the productivity of interferon-γ was about eight-fold higher than μ controlled at a fixed value of 0.12 h^{-1} [4]. However, for a sustainable HCDC, μ should be maintained at less than the threshold growth rate where accumulation of glucose and acetate become inhibitory for cell growth and protein production. Direct feedback feeding is also possible by measuring the concentration of growth-limiting substrates such as glucose or the metabolic state of the cells to automatically adjust the feed rate [42, 51]. However, these techniques are rarely applied in industry.

Fermentation scale-up

For commercial protein therapeutics production, the fermentation usually starts in a laboratory-scale bioreactor (e.g., 5–30 l) to identify suitable growth and protein expression conditions. The process then transfers to pilot level (e.g., 200–600 l) to establish optimal operating parameters and finally to manufacturing scale (e.g., over 2,000 l) to reach high productivity. The scale-up process for any therapeutic protein should aim for high productivity with consistency in the protein quality and specific yield. However, as the scale increases, important biological, chemical and physical parameters affecting cell growth, as well as protein

expression, will also change. This makes the scale-up process a challenging task. The common problems associated with scale-up originate from poor mixing, which increases circulation time and creates stagnant regions. This leads to imbalanced and zonal distribution of oxygen, nutrients, pH, heat and metabolites inside the bioreactor [20]. For example, in a normal large vessel, a vertical DO gradient (from bottom to top) is usually observed if the bulk mixing rate is slower than the mass transfer rate and a gradient can also occur with the feeding substrate, such as glucose (from top to bottom). In this case, cells at the top of the bioreactor are simultaneously exposed to high glucose concentration and oxygen limitation, resulting in metabolic overflow and mixed acid fermentation. Since mixing of the cells is still in progress, the same cells may shift to different locations of the bioreactor with different environmental stresses. Passing through the different stress zones induces metabolic shifts in cells, which will ultimately reduce cell growth and have negative impacts on both productivity and quality [89]. Therefore, several strategies have been used as principles to scale up *E. coli* fermentation to minimize the differences between scales by keeping one or more parameters constant from laboratory to manufacturing bioreactors. These parameters include power input per liquid volume (P/V), oxygen transfer rate (OTR), oxygen mass transfer coefficient (k_La), impeller tip speed, mixing time and impeller Reynold's number (N_{Re}) [47].

Traditionally, the constant P/V has been shown to be a successful scale-up criterion for industrial fungal and mammalian cell fermentation, but may be limited to recombinant *E. coli* culture because of its high energy requirement [89, 113]. Eslam and coworkers have suggested that scaling up based on k_La is the most appropriate approach for microorganisms, such as *E. coli* growing under aerobic conditions, and other studies have shown that it is indeed the most applied physical scale-up variable [36, 89]. The k_La is directly related to a bioreactor's configuration and can be modulated by manipulating the bioreactor's agitation speed, impeller design and air/O_2 flow rate. In a large-scale and high-performance *E. coli* fermentation where power input and mixing are not an issue, a similar OTR and heat transfer rate will normally be used to ensure high productivity [126]. Scale-up based on a constant mixing time or tip speed usually has a higher success rate when the scale-up factor is small (4–40 l, scale-up factor 10) [13, 89].

In practice, however, the important scale-up parameters for each product may be different because of process differences and the operation limitations imposed by manufacturing facilities. Therefore, a detailed and comprehensive process characterization must be carried out in advance to identify critical parameters influencing protein yield and quality. Those parameters then should be kept constant

during the scale-up processes. Several approaches have been utilized to characterize fermentation processes. These include real time on-line measurement of cellular responses, such as biomass and viability, and the concentrations of different substrates and metabolites. This information allows scientists and engineers to pinpoint complex interactions between cells and their environment, as well as their impacts on protein productivity and quality [89]. Another strategy is to create a laboratory-scale bioreactor that mimics the performance of its large scale counterpart, in terms of oxygen transfer, mixing and medium sterilization, and to use this scale-down model to evaluate important process parameters on a specific product [81]. Recently, computational fluid dynamics techniques have also been used to simulate and predict the mixing quality and the substrate gradient zones in a large-scale bioreactor [22]. The knowledge gained is then used to optimize the fermentation processes or to design a better bioreactor to improve the robustness of the process in a scale-up practice. Other challenges one might encounter in large-scale upstream manufacturing include availability of appropriately sized seed train bioreactors and feed tanks, medium sterilization and condensation, poor heat transfer capacity, variability caused by evaporation and holding stability of feed media solution [126].

Recent advancements in recombinant therapeutics production

Antibody fragments and full-length aglycosylated antibodies

Several recent technological advances have made *E. coli* a more appealing host to produce recombinant therapeutics (Table 3). These include the ability to produce different antibody fragments in high yields. It is estimated that at least 54 antibody fragments have entered clinical studies, with most of them being Fabs and scFvs [74]. Like full-length glycosylated antibodies, antibody fragments can also exhibit prominent clinical benefits for different oncological and immunological diseases. Although these fragments have disadvantages, such as short half-life in serum and inability to induce antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) when compared to full-length glycosylated antibodies, they show similar binding specificity and can penetrate tissues and tumors with higher efficiency. In addition, production of these small and aglycosylated fragments can be implemented in microbial systems rather than a costly mammalian cell culture [34, 121]. In fact, two recently approved Fab fragments, Lucentis® and Cimza®, are produced in *E. coli* (Table 1), indicating that *E. coli* has

Table 3 Recent advancements in *E. coli* and their significance for industrial applications

Advancements	Significance	References
Antibody fragments and full-length aglycosylated antibodies	Complex therapeutic expression; novel therapeutic properties	[45, 59, 74, 83, 111]
Strain engineering	Improved protein yield and quality; simplified purification	[10, 14, 92]
E. coli N-linked glycosylation	N-linked glycosylated protein synthesis	[62, 77, 91, 120]
Fusion proteins and PEGylation	Enhanced pharmacokinetics; novel therapeutic design	[41, 101, 104]
E. coli cell-free systems	Novel method for protein synthesis at the industrial scale	[76, 102, 131]

emerged as an ideal and robust host to express antibody fragments.

Expression of functional antibody fragments is normally achieved by targeting both light chains and heavy chains or scFvs into the periplasm, where proper protein folding and disulfide bond formation can occur. This idea was first implemented by Skerra and Pluckthun, but a very low titer was achieved [32]. Since then, different approaches have been adopted to improve Fab fragment production in *E. coli*. These methods include optimization of the expression ratio of the heavy chain and light chain, co-expression of different folding chaperons and fermentation optimization [2]. Chen and coworkers have shown that an anti-CD18 F(ab′)₂ can be produced at 2 g/l by increasing the stability of the light chain in the periplasm [14]. Recently, a yield of over 10 g/l has also been reported for different scFvs [59]. Cytoplasmic expression has also been explored as an alternative for high-level antibody fragment expression in *E. coli*. Although both the light chain and the heavy chain can be expressed separately as IBs, the refolding process was not efficient and not economically competitive [32]. Expression of antibody fragments in an *E. coli* strain with an oxidizing cytoplasm also showed some promising results; however, this approach has not been adopted by the industry yet [116]. Periplasmic secretion is still the major route used by the industry for production of functional antibody fragments.

In addition to antibody fragments, therapeutic applications of aglycosylated full-length antibodies have gained much attention in the past decade. Unlike glycosylated antibodies, which can activate the innate immune system for targeted cell death (ADCC/CDC), aglycosylated antibodies can be used in other clinical applications, such as antigen blocking, and receptor agonist and antagonist roles, where engagement of the immune system is not required and may even cause unwanted side effects [56]. In this case, *E. coli* is an ideal expression host for aglycosylated antibodies, as it possesses several bioprocessing advantages, as well as enables the production of fully aglycosylated antibodies. Simmons and coworkers reported the first efficient production method for an aglycosylated antibody, anti-tissue factor IgG1, in the *E. coli* periplasm. They obtained a yield of

approximately 0.15 g/l using a bicistronic plasmid expressing heavy chains and light chains in a favorable ratio [97]. The produced antibody retained a similar half-life to its glycosylated counterpart and exhibited specific binding to the antigen and neonatal receptor (FcRn), but not to the FcγRI and C1q effectors. Coexpression of DsbA and DsbC further improved the yield to 1 g/l in an optimized HCDC culture [83]. Currently, several aglycosylated antibodies are being evaluated in clinical trials for different indications. One of them is MetMAb, a one-armed anti-cMet antibody containing a single Fab and aglycosylated Fc, which is the first therapeutic aglycosylated antibody manufactured in *E. coli*.

Recently, aglycosylated antibodies have also been engineered to have similar functionalities to glycosylated antibodies, especially the ability to bind different Fcγ receptors. Using the yeast surface display system, Sazinsky and coworkers identified an aglycosylated Fc variant (S298G/T299A) showing comparable binding affinities to FcγRIIa and FcγRIIb as glycosylated antibodies [88]. The binding affinity to FcγRI was also partially restored in this selected variant. A similar idea but different approach also led to the isolation of the E382V/M428I mutant in the CH3 region, which conferred binding affinity to FcγRI at a nearly identical level as that observed for glycosylated antibodies [46]. Remarkably, when Jung and coworkers introduced these double mutations into the anti-Her2 antibody trastuzumab, the mutated variant (trastuzumab-Fc5, which is produced in *E. coli*), elicited a significantly higher dendritic cell-mediated ADCC response than the glycosylated trastuzumab. Based on the differences observed in protein structures between aglycosylated and glycosylated antibodies, this group also hypothesized that higher conformational flexibility found in aglycosylated antibodies provides more degrees of freedom to identify variants with selective binding affinity towards different Fc receptors that may show novel therapeutic efficacies, as demonstrated in trastuzumab-Fc5. Therefore, recent breakthroughs in high-yield production of aglycosylated antibodies in *E. coli* and in engineering them for increased binding affinity with Fc receptors indicate that aglycosylated antibodies have become an important family of protein therapeutics produced in *E. coli*.

Strain engineering for improved product yield and quality

It is estimated that more than 3% of the enzymatic activities in *E. coli* are proteolytic at any given time [68]. Recombinant proteins produced in *E. coli*, both as IBs and soluble proteins, are susceptible to different levels of proteolysis, which is unfavorable for industrial production. Degradation creates protein fragments that can decrease product yield, affect protein quality and increase the production cost. One of the strategies used to improve recombinant protein stability has been deletion of crucial cytoplasmic or periplasmic proteases of the expressing strains. Lon and ClpP are two major ATP-dependent proteases located in the cytoplasm, and deletion of both genes with ClpYQ has been shown to improve the stability of recombinant human pro-urokinase by five-fold [49]. IB formation is known to protect recombinant protein from degradation. However, a recent study indicated that protein aggregates can also undergo direct proteolytic attack. In this report, the absence of Lon and ClpP also reduced IB disintegration up to 40% [117]. Even though the protein stability improved significantly in these studies, the ClpP-deficient strain exhibited a reduced growth rate and higher cell lysis in HCDC [85]. *E. coli* B and its derivatives are also deficient in Lon and OmpT, an outer membrane protease, and improved protein stability has also been observed in B strains compared with the K12 strains. In periplasm, DegP, Prc and protease III (ptr) are three major proteases. Chen and colleagues demonstrated that production of anti-CD18 F(ab′)$_2$ in a protease double mutant (*DegP* and *prc*) could significantly decrease the generation of truncated light-chain species compared to the wild-type strain. When this double mutant was characterized in the HCDC, a dramatic increase in cell lysis was also observed, and most of the produced F(ab′)$_2$ was located in the supernatant with low yield. Later, a third mutation in *spr*, a Prc suppressor, was included to compensate for the growth defects of Prc-deficient strain. This triple mutant increased the protein titer to 2.4 g/l, which is sevenfold higher than the double mutant [14]. Together, these studies indicate that protease inactivation has profound effects on recombinant protein stability, and therefore deletion of these proteases can significantly improve protein productivity.

Escherichia coli strains can also be designed specifically for each protein therapeutic to guarantee process robustness, high product yield and quality. One example was demonstrated by Caparon and coworkers who used proteomic techniques to identify several difficult-to-remove host cell proteins (HCPs), OppA, DppA and MalE, in the bioprocess producing ApoA-1M, a potential therapeutic for coronary heart disease. Genetic deletion of these three problematic HCPs and OmpT protease created a quadruple knockout strain, GB004, that exhibited similar growth char-

acteristics to the parental strain. The new strain enabled conventional process optimization to be carried out, increased the ApoA-1M titer from 3.2 to 5 g/l and, most importantly, significantly decreased the HCP levels in the final purified protein [10].

Recently, with the advent of synthetic biology, *E. coli* strains with genome size reductions up to 14% were created by deleting of large unnecessary DNA segments of *E. coli* K12 genome. These strains were evaluated for different beneficial properties, including recombinant protein production (chloramphenicol acetyltransferase) during HCDC, and similar growth and protein yield were observed when compared to the parental strain [92]. By using this technology, a Clean Genome® *E. coli* strain was engineered as a novel biological factory providing enhanced genetic stability and high protein yield. Using a similar design concept, rational and systematic manipulation of the *E. coli* genome is expected to create a new method for developing more robust *E. coli* strains for recombinant therapeutic production. Therefore, strain engineering has continued to be a practical and effective method to enhance the production yield and quality of recombinant therapeutics.

E. coli N-linked glycosylation

The first prokaryotic N-linked protein glycosylation was uncovered in *Campylobacter jejuni*, which contains a gene cluster, *pgl*, involved in the biosynthesis of a number of different immunogenic glycoproteins [105]. By transferring the *pgl* pathway into *E. coli*, Wacker and colleagues successfully produced glycosylated proteins in these cells, opening the door for using *E. coli* to produce complex and glycosylated human therapeutics [120]. Compared to eukaryotes, this *E. coli* glycosylation system possesses similar but distinctive characteristics in its glycosylation machinery. For example, the oligosaccharyltransferase in *C. jejuni*, PglB, is the major enzyme to carry out glycan transfer; however, for eukaryotes, a multimeric protein complex is required. PglB also displays flexible substrate specificity, for both native and non-native proteins, and is promiscuous with regard to glycan structures that are transferred to proteins [23, 26]. The primary consensus sequence for N-glycosylation by PglB is extended to D/E-Y-N-X-S/T (Y, X ≠P), and the glycosylation efficiency increases when this domain is in the flexible and solvent-exposed region of a folded protein [55]. In addition, the *pgl* operon synthesizes a heptasaccharide glycan, which is completely different from eukaryotic N-glycans.

Since there are several intrinsic differences in N-glycosylation between this *E. coli* system and humans, overcoming these disparities is important for the development of the first humanized *E. coli* glycoprotein. One step forward in this direction was demonstrated by Lizak and coworkers,

who introduced bacterial glycosylation sites into the linker region of an scFv, and the expressed scFv was found to be glycosylated with *C. jejuni* N-glycans. The glycosylated scFv showed improved protein stability and solubility [62]. A two-step method to produce eukaryotic N-glycans in *E. coli* was also developed. In this study, *E. coli* was glyco-engineered by deleting the genes responsible for the synthesis and transfer of bacillosamine, an immunogenic non-human saccharide. Upon coexpressing WecA, this strain was able to transfer GlcNAc-1-phosphate to UND-P to produce the $(GalNAc)_5GlcNAc_2$ glycan and form the GlcNAc-Asn (human-like) linkage. After subsequent in vitro glycan trimming and enzymatic transglycosylation, a eukaryotic-like ($Man_3GlcNAc_2$) glycoprotein was produced [91]. In this study, the $Man_3GlcNAc_2$ structure was successfully introduced to the bacterial protein AcrA as well as two eukaryotic proteins, human IgG-Fc (CH2 domain) and scFv F8, a potential therapeutic antibody fragment. However, the in vivo glycosylation efficiency was low and varied (5–40%) among these tested acceptor proteins. This could be attributed to the undesirable metabolic state in *E. coli* that generates an insufficient amount of N-glycans for protein glycosylation. To improve the in vivo glycosylation efficiency of AcrA, Pandhal and coworkers further developed a comprehensive approach to forward engineer *E. coli* to enhance AcrA glycosylation efficiency by threefold through overexpression of isocitrate lyase [77].

Recently, expression of active human sialyltransferase was also achieved in *E. coli*, making possible the addition of sialic acid to *E. coli* glycoproteins [99]. Therefore, in the past few years, significant achievements toward the first humanized glycoprotein in *E. coli* have been demonstrated. However, a critical step to put the above-mentioned components together in one single *E. coli* is still missing. At this point, production of therapeutic glycoprotein in *E. coli* has shown promising results, and it is possible that *E. coli* may be used for industrial production of therapeutic glycoproteins in the near future.

Protein modification: fusion proteins and PEGylation

In general, small therapeutic proteins and peptides can be manufactured in high yield using microbial systems. However, these proteins tend to have short in vivo half-lives and fast renal clearance, resulting in the need for frequent dosing to maintain the desired drug level. To improve the pharmacokinetics of these proteins, efforts have focused on genetic engineering of the protein sequences and on increasing the size of the proteins by adding large molecules such as fusion proteins and polyethylene glycol (PEG). The latter strategy (increasing protein size) presents a great opportunity to expand *E. coli*-derived therapeutics, as illustrated by recently approved Cimza® and Nplate®.

The Fc antibody fragment has been used in several FDA-approved protein therapeutics including Enbrel® (TNFR2-Fc), Orencia® (CTLA4-Fc) and Nplate® (TPO-R binding peptide-Fc). Nplate is a hybrid protein consisting of a small peptide moiety and Fc fragment. It is the only commercial Fc fusion therapeutic produced in *E. coli* that is expressed as IBs and refolded to regain its biological function. The therapeutic design and production platform of Fc fusion proteins are promising, especially when the engagement of glycosylated Fc with Fc receptors seems unnecessary in some Fc fusion therapeutics [41]. Another clinically used fusion construct is human serum albumin (HSA), a widely distributed inert serum protein with a long half-life. Several HSA-fused protein therapeutics, such as albINF-γ, alb-insulin and alb-GH, exhibit prolonged serum half-life with unaltered biological activities [101]. Most of the HSA fusion therapeutics are produced as recombinant secreted proteins in *S. cerevisiae* and *Pichia pastoris*. *E. coli* is generally considered an unsuitable host to express HSA because of the protein's complex molecular structure (17 disulfide linkages) [35].

PEGylation is the most successful and commonly used chemical modification approach to improve the pharmacokinetics of protein therapeutics. Although PEGylation also increases the production cost and the heterogeneity of the final product, this technology has become increasingly popular in recent years. Several PEG-modified proteins, including PEGylated G-CSF, human growth hormone, interferon-α and anti-TNFα Fab fragments, are available on the market, and most are produced in *E. coli* [104]. Therefore, newly developed fusion proteins and protein modification technologies have offset some of the drawbacks of *E. coli*-derived therapeutics and expanded the use of this host for more diversified therapeutic applications.

Cell-free systems for protein therapeutic synthesis

Cell-free systems are attractive alternatives to producing recombinant therapeutics in vivo. Compared to traditional protein production in living cells, cell-free systems offer several advantages for the development and production of therapeutics. Due to the simplicity and easy manipulation of this system, in vitro protein synthesis and its functional analysis can be carried out in few hours, making it a powerful tool for high-throughput protein screening and genome-wide protein production [73]. Because there is no need to maintain cell growth, the whole system can be optimized for the production of a single protein. The optimization includes transcription and translation modulation and incorporation of non-natural amino acids into the protein. In addition, reduced amounts of endotoxin and cell debris make subsequent protein purification processes easy to implement. Two different types of cell-free systems are

now commercially available. One is the PURE system, which is composed of individually purified components of the *E. coli* translational machinery. The other is a cell-extract based system, where different cell extracts, including *E. coli*, wheat germ, rabbit reticulocyte and insect cells, are used to supply the protein translational apparatus [76].

In the past decade, cell-free systems using *E. coli* extract have made several breakthroughs, creating a reliable platform for industrial therapeutic production. In 2004, Jewett and Swartz engineered the "Cytomin" system that activates bacterial oxidative phosphorylation pathway and uses pyruvate as the energy source to allow the energy generation inside the system to be cost-effective [40]. This system also prolonged the protein synthesis reaction by reducing by-product accumulation and stabilizing the pH of the reaction. The quality of *E. coli* extracts also has been improved in several ways, including optimization of extract preparation methods and engineering of an *E. coli* KC6 strain that maintains stable pools of amino acids during the protein synthesis [9, 61]. The KC6 strain was then further engineered in favor of synthesis of disulfide bonded proteins by creating an optimal thiol redox potential [52]. At this point, the system has demonstrated a variety of applications, including production of protein therapeutics such as murine granulocyte macrophage colony stimulating factor (GM-CSF), scFvs and IGF-I. For commercial therapeutic production, however, scaling-up a cell-free system was a more significant challenge [102]. Recently, Zawada and coworkers showed that efficient, predictable and scalable protein production can be achieved in this system by optimization of important parameters, including extract incubation conditions, DNA sequence and redox environment for disulfide bond formation. The process used in this study produced recombinant GM-CSF at a concentration of 700 mg/l within 10 h in a 100-l standard bioreactor [131]. This result is a big breakthrough for the cell-free system since it confirms linear scalability from 200 µl to 100 l with similar titers. It also demonstrates that this technology is approaching its utility for industrial production.

Conclusion

Almost 30 years after the first recombinant insulin was approved by the FDA, *E. coli* is still widely used by the industry for recombinant therapeutics production. Versatile genetic tools are now available for tightly regulated and high-level protein expression in *E. coli*, where protein induction remains simple and cost-effective. While production of recombinant proteins in IBs and their refolding have become common practices in industry, periplasmic secretion provides advantages, such as proper folding and yield of proteins with authentic N-terminus. Two recently

approved antibody fragments were produced in the periplasm, indicating that this technology is now commercially competitive. In addition, improved extracellular protein secretion now can be achieved in leaky strains. These developments allow the industry to leverage the benefits in each *E. coli* expression method and provide flexibility in protein expression. High cell density fermentation in *E. coli* can further enhance the overall productivity of recombinant therapeutics. Various tools, including different feeding strategies and scale-up protocols, have been established to ensure high-yield protein production at a manufacturing scale. However, challenges remain for the continued use of *E. coli* for protein production, because of its assumed inability to express some complex and glycosylated proteins. Recently, several advancements in *E. coli* protein production have challenged these assumptions and made *E. coli* an even more robust expression host. A full-length aglycosylated antibody can now be expressed in *E. coli* with high yield (∼1 g/l), and production of different Fab fragments was also achieved in high titers in protease-deficient *E. coli* mutants. These studies demonstrate that *E. coli* can express complex proteins as efficiently as mammalian cells. The N-glycosylation pathway has been engineered in *E. coli*, and the progress towards the first humanized glycoprotein produced from *E. coli* looks promising. Moreover, addition of fusion partners to improve the pharmacokinetics of *E. coli*-produced therapeutic proteins, as well as the use of cell-free systems to generate therapeutics, further expands the capability of *E. coli* to express different protein therapeutics. With all these advancements and advancements yet to come, *E. coli* will definitely remain a workhorse for the recombinant therapeutic production in the industry.

References

1. Alibolandi M, Mirzahoseini H, Abad MAK, Azami movahed M (2010) High level expression of human basic fibroblast growth factor in *Escherichia coli*: evaluating the effect of the GC content and rare codons within the first 13 codons. Afr J Biotechnol 9(16):2456–2462

2. Andersen DC, Reilly DE (2004) Production technologies for monoclonal antibodies and their fragments. Curr Opin Biotechnol 15(5):456–462

3. Andrews B, Adari H, Hannig G, Lahue E, Gosselin M, Martin S, Ahmed A, Ford PJ, Hayman EG, Makrides SC (1996) A tightly regulated high level expression vector that utilizes a thermosensitive *lac* repressor: production of the human T cell receptor V beta 5.3 in *Escherichia coli*. Gene 182(1–2):101–109

4. Babaeipour V, Shojaosadati SA, Robatjazi SM, Khalilzadeh R, Maghsoudi N (2007) Over-production of human interferon-gamma by HCDC of recombinant *Escherichia coli*. Process Biochem 42(1):112–117

5. Backlund E, Reeks D, Markland K, Weir N, Bowering L, Larsson G (2008) Fedbatch design for periplasmic product retention in *Escherichia coli*. J Biotechnol 135(4):358–365

6. Balbas P (2001) Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. Mol Biotechnol 19(3):251–267

7. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol 10(5):411–421

8. Bessette PH, Aslund F, Beckwith J, Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc Natl Acad Sci USA 96(24):13703–13708

9. Calhoun KA, Swartz JR (2006) Total amino acid stabilization during cell-free protein synthesis reactions. J Biotechnol 123(2):193–203

10. Caparon MH, Rust KJ, Hunter AK, McLaughlin JK, Thomas KE, Herberg JT, Shell RE, Lanter PB, Bishop BF, Dufield RL, Wang X, Ho SV (2010) Integrated solution to purification challenges in the manufacture of a soluble recombinant protein in *E. coli*. Biotechnol Bioeng 105(2):239–249

11. Carnes AE, Hodgson CP, Williams JA (2006) Inducible *Escherichia coli* fermentation for increased plasmid DNA production. Biotechnol Appl Biochem 45(3):155–166

12. Carrier TA, Keasling JD (1997) Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. Biotechnol Prog 13(6):699–708

13. Cascaval D, Galaction AI, Camarut S (2011) Scale-up of aerobic stirred bioreactors using the mixing time criteria 1. simulated broths. Chem Biochem Eng Q 25(1):43–54

14. Chen C, Snedecor B, Nishihara JC, Joly JC, McFarland N, Andersen DC, Battersby JE, Champion KM (2004) High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (*degP prc spr*) host strain. Biotechnol Bioeng 85(5):463–474

15. Choi JH, Keum KC, Lee SY (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem Eng Sci 61(3):876–885

16. Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl Microbiol Biotechnol 64(5):625–635

17. DeLisa MP, Tullman D, Georgiou G (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci USA 100(10):6115–6120

18. Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? Biotechnol J 5(6):547–559

19. Eiteman MA, Altman E (2006) Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotechnol 24(11):530–536

20. Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jurgen B, Kruger E, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Tragardh C, Fuchs L, Revstedt J, Friberg PC, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, van der Lans R, Luyben K, Vrabel P, Manelius A (2001) Physiological responses to mixing in large scale bioreactors. J Biotechnol 85(2):175–185

21. Fahnert B, Lilie H, Neubauer P (2004) Inclusion bodies: formation and utilisation. Adv Biochem Eng Biotechnol 89:93–142

22. Fang Z (2010) Applying computational fluid dynamics technology in bioprocesses. Biopharm Int 23(4):38

23. Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, Morris HR, Dell A, Valvano MA, Aebi M (2005) Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. Proc Natl Acad Sci USA 102(8):3016–3021

24. Fernandez LA, Sola I, Enjuanes L, de Lorenzo V (2000) Specific secretion of active single-chain Fv antibodies into the supernatants of *Escherichia coli* cultures by use of the hemolysin system. Appl Environ Microbiol 66(11):5024–5029

25. Ferrer-Miralles N, Domingo-Espín J, Corchero J, Vázquez E, Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8(1):17

26. Fisher AC, Haitjema CH, Guarino C, Celik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP (2011) Production of secretory and extracellular N-Linked glycoproteins in *Escherichia coli*. Appl Environ Microbiol 77(3):871–881

27. Friehs K (2004) Plasmid copy number and plasmid stability. Adv Biochem Eng Biotechnol 86:47–82

28. Furman TC, Epp J, Hsiung HM, Hoskins J, Long GL, Mendelsohn LG, Schoner B, Smith DP, Smith MC (1987) Recombinant human insulin-like growth factor-I expressed in *Escherichia coli*. Bio-Technology 5(10):1047–1051

29. Graumann K, Premstaller A (2006) Manufacturing of recombinant therapeutic proteins in microbial systems. Biotechnol J 1(2):164–186

30. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22(7):346–353

31. Gvritishvili AG, Leung KW, Tombran-Tink J (2010) Codon preference optimization increases heterologous PEDF expression. PLoS ONE 5(11):e15056

32. Harrison JS, Keshavarz-Moore E (1996) Production of antibody fragments in *Escherichia coli*. Ann NY Acad Sci 782:143–158

33. Huang CJ, Chen RH, Vannelli T, Lee F, Ritter E, Ritter G, Old LJ, Batt CA (2007) Expression and purification of the cancer antigen SSX2: a potential cancer vaccine. Protein Expr Purif 56(2):212–219

34. Huang CJ, Lowe AJ, Batt CA (2010) Recombinant immunotherapeutics: current state and perspectives regarding the feasibility and market. Appl Microbiol Biotechnol 87(2):401–410

35. Huang YS, Chen Z, Chen YQ, Ma GC, Shan JF, Liu W, Zhou LF (2008) Preparation and characterization of a novel exendin-4 human serum albumin fusion protein expressed in *Pichia pastoris*. J Pept Sci 14(5):588–595

36. Islam RS, Tisi D, Levy MS, Lye GJ (2008) Scale-up of *Escherichia coli* growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched *k*La. Biotechnol Bioeng 99(5):1128–1139

37. Ivanov AV, Korovina AN, Tunitskaya VL, Kostyuk DA, Rechinsky VO, Kukhanova MK, Kochetkov SN (2006) Development of the system ensuring a high-level expression of hepatitis C virus nonstructural NS5B and NS5A proteins. Protein Expr Purif 48(1):14–23

38. Jana S, Deb JK (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl Microbiol Biotechnol 67(3):289–298

39. 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):1–11

40. Jewett MC, Swartz JR (2004) Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng 86(1):19–26

41. Jiang XR, Song A, Bergelson S, Arroll T, Parekh B, May K, Chung S, Strouse R, Mire-Sluis A, Schenerman M (2011) Advances in the assessment and control of the effector functions of therapeutic antibodies. Nat Rev Drug Discov 10(2):101–110

42. Jobe AM, Herwig C, Surzyn M, Walker B, Marison I, von Stockar U (2003) Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. Biotechnol Bioeng 82(6):627–639

43. 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(6):2773–2777

44. Jong WSP, Saurí A, Luirink J (2010) Extracellular production of recombinant proteins using bacterial autotransporters. Curr Opin Biotechnol 21(5):646–652

45. Jung ST, Kang TH, Kelton W, Georgiou G (2011) Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. Curr Opin Biotechnol 22(6):858–867

46. Jung ST, Reddy ST, Kang TH, Borrok MJ, Sandlie I, Tucker PW, Georgiou G (2010) Aglycosylated IgG variants expressed in bacteria that selectively bind Fc gamma RI potentiate tumor cell killing by monocyte-dendritic cells. Proc Natl Acad Sci USA 107(2):604–609

47. Junker BH (2004) Scale-up methodologies for *Escherichia coli* and yeast fermentation processes. J Biosci Bioeng 97(6):347–364

48. Kamionka M (2011) Engineering of therapeutic proteins production in *Escherichia coli*. Curr Pharm Biotechnol 12(2):268–274

49. Kanemori M, Nishihara K, Yanagi H, Yura T (1997) Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of sigma32 and abnormal proteins in *Escherichia coli*. J Bacteriol 179(23):7219–7225

50. Kim SG, Shin SY, Park YC, Shin CS, Seo JH (2011) Production and solid-phase refolding of human glucagon-like peptide-1 using recombinant *Escherichia coli*. Protein Expr Purif 78(2):197–203

51. Kleman GL, Chalmers JJ, Luli GW, Strohl WR (1991) A predictive and feedback control algorithm maintains a constant glucose concentration in fed-batch fermentations. Appl Environ Microbiol 57(4):910–917

52. Knapp KG, Goerke AR, Swartz JR (2007) Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source. Biotechnol Bioeng 97(4):901–908

53. 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

54. Koo TY, Park TH (1999) Increased production of recombinant protein by *Escherichia coli* deficient in acetic acid formation. J Microbiol Biotechnol 9(6):789–793

55. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M (2006) Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25(9):1957–1966

56. Labrijn AF, Aalberse RC, Schuurman J (2008) When binding is enough: nonactivating antibody formats. Curr Opin Immunol 20(4):479–485

57. Lee SY (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol 14(3):98–105

58. Levisauskas D (2001) Inferential control of the specific growth rate in fed-batch cultivation processes. Biotechnol Lett 23(15):1189–1195

59. Liddell JM (2009) Production strategies for antibody fragment therapeutics. Biopharm Int Suppl 22(6):36–42

60. Lilie H, Schwarz E, Rudolph R (1998) Advances in refolding of proteins produced in *E. coli*. Curr Opin Biotechnol 5:497–501

61. Liu DV, Zawada JF, Swartz JR (2005) Streamlining *Escherichia coli* S30 extract preparation for economical cell-free protein synthesis. Biotechnol Prog 21(2):460–465

62. Lizak C, Fan YY, Weber TC, Aebi M (2011) N-Linked glycosylation of antibody fragments in *Escherichia coli*. Bioconjug Chem 22(3):488–496

63. Lopez PJ, Marchand I, Joyce SA, Dreyfus M (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. Mol Microbiol 33(1):188–199

64. Lowe AJ, Bardliving CL, Huang CJ, Teixeira LM, Damasceno LM, Anderson KA, Ritter G, Old LJ, Batt CA (2011) Expression and purification of cGMP grade NY-ESO-1 for clinical trials. Biotechnol Prog 27(2):435–441

65. Luo Y, Fan DD, Ma XX, Wang DW, Mi Y, Hua XF, Li WH (2005) Process control for production of human-like collagen in fed-batch culture of *Escherichia coli* BL 21. Chin J Chem Eng 13(2):276–279

66. Maeng BH, Nam DH, Kim YH (2011) Coexpression of molecular chaperones to enhance functional expression of anti-BNP scFv in the cytoplasm of *Escherichia coli* for the detection of B-type natriuretic peptide. World J Microbiol Biotechnol 27(6):1391–1398

67. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev 60(3):512–538

68. Maurizi MR (1992) Proteases and protein-degradation in *Escherichia coli*. Experientia 48(2):178–201

69. 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(2):289–295

70. Menart V, Jevsevar S, Vilar M, Trobis A, Pavko A (2003) Constitutive versus thermo inducible expression of heterologous proteins in *Escherichia coli* based on strong PR, PL promoters from phage lambda. Biotechnol Bioeng 83(2):181–190

71. Mergulhao FJ, Summers DK, Monteiro GA (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnol Adv 23(3):177–202

72. Mucke M, Ostendorp R, Leonhartsberger S (2009) *E. coli* secretion technologies enable production of high yields of active human antibody fragments. BioProcess Int 7(8):40–47

73. Mureev S, Kovtun O, Nguyen UT, Alexandrov K (2009) Species-independent translational leaders facilitate cell-free expression. Nat Biotechnol 27(8):747–752

74. Nelson AL, Reichert JM (2009) Development trends for therapeutic antibody fragments. Nat Biotechnol 27(4):331–337

75. Ni Y, Chen R (2009) Extracellular recombinant protein production from *Escherichia coli*. Biotechnol Lett 31(11):1661–1670

76. Ohashi H, Kanamori T, Shimizu Y, Ueda T (2010) A highly controllable reconstituted cell-free system- a breakthrough in protein synthesis research. Curr Pharm Biotechnol 11(3):267–271

77. Pandhal J, Ow SY, Noirel J, Wright PC (2010) Improving N-glycosylation efficiency in *Escherichia coli* using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. Biotechnol Bioeng 108:902–912

78. Perezperez J, Martinezcaja C, Barbero JL, Gutierrez J (1995) Dnak/Dnaj supplementation improves the periplasmic production of human granulocyte colony stimulating factor in *Escherichia coli*. Biochem Biophys Res Commun 210(2):524–529

79. Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35:W126–W131

80. Qian ZG, Xia XX, Choi JH, Lee SY (2008) Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in *Escherichia coli*. Biotechnol Bioeng 101(5):587–601

81. Rathore A, Krishnan R, Tozer S, Smiley D, Rausch S, Seely J (2005) Scaling down of biopharmaceutical unit operations—part 2: chromatography and filtration. Biopharm Int 18(4):58

82. Ray MVL, Vanduyne P, Bertelsen AH, Jacksonmatthews DE, Sturmer AM, Merkler DJ, Consalvo AP, Young SD, Gilligan JP, Shields PP (1993) Production of recombinant salmon calcitonin

by in vitro amidation of an *Escherichia coli* produced precursor peptide. Bio-Technology 11(1):64–70

83. Reilly DE, Yansura DG (2010) Production of monoclonal antibodies in *E. coli*. In: Shire SJ, Gombotz W, Bechtold-Peters K, Andya J (eds) Current trends in monoclonal antibodies development and manufacturing. Springer, New York, pp 295–308

84. Riscaldati E, Ciabini A, Baccante A, Moscatelli D, Errichetti M, Colagrande A, Cencioni S, Marcocci F, Di Cioccio V, Di Ciccio L, Allegretti M, Martin F, Maurizi G (2006) Set up and optimization of a fermentation protocol for the production of a human antibody fragment (Fab') express in *E. coli*. Pre-pilot and cGMP pilot scale studies. Microb Cell Fact 5(Suppl 1):P29

85. Rozkov A, Enfors SO (2004) Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. Adv Biochem Eng Biotechnol 89:163–195

86. 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(1–2):249–264

87. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27(10):946–950

88. Sazinsky SL, Ott RG, Silver NW, Tidor B, Ravetch JV, Wittrup KD (2008) Aglycosylated immunoglobulin G_1 variants productively engage activating Fc receptors. Proc Natl Acad Sci USA 105(51):20167–20172

89. Schmidt FR (2005) Optimization and scale up of industrial fermentation processes. Appl Microbiol Biotechnol 68(4):425–435

90. Schumann W, Ferreira LCS (2004) Production of recombinant proteins in *Escherichia coli*. Genet Mol Biol 27:442–453

91. Schwarz F, Huang W, Li CS, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX (2010) A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol 6(4):264–266

92. Sharma SS, Blattner FR, Harcum SW (2007) Recombinant protein production in an *Escherichia coli* reduced genome strain. Metab Eng 9(2):133–141

93. Shiloach J, Fass R (2005) Growing *E. coli* to high cell density–a historical perspective on method development. Biotechnol Adv 23(5):345–357

94. Shin CS, Hong MS, Bae CS, Lee J (1997) Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2]. Biotechnol Prog 13(3):249–257

95. Shokri A, Sanden AM, Larsson G (2002) Growth rate-dependent changes in *Escherichia coli* membrane structure and protein leakage. Appl Microbiol Biotechnol 58(3):386–392

96. Shokri A, Sanden AM, Larsson G (2003) Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. Appl Microbiol Biotechnol 60(6):654–664

97. Simmons LC, Reilly D, Klimowski L, Raju TS, Meng G, Sims P, Hong K, Shields RL, Damico LA, Rancatore P, Yansura DG (2002) Expression of full-length immunoglobulins in *Escherichia coli*: rapid and efficient production of aglycosylated antibodies. J Immunol Methods 263(1–2):133–147

98. Simmons LC, Yansura DG (1996) Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. Nat Biotechnol 14(5):629–634

99. Skretas G, Carroll L, DeFrees S, Schwartz MF, Johnson KF, Georgiou G (2009) Expression of active human sialyltransferase ST6GalNAcI in *Escherichia coli*. Microb Cell Fact 8:50

100. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA Polymerase to direct selective high level expression of cloned genes. J Mol Biol 189(1):113–130

101. Subramanian M, Zeuzem S, Neumann A, Nelson D, Bain V, Balan V, Corey A, Fiscella M (2007) Albinterferon alpha-2b

(ALB-IFN): Anti-HCV activity and pharmacokinetics/pharmacodynamics. J Interf Cytok Res 27(8):703

102. Swartz J (2006) Developing cell-free biology for industrial applications. J Ind Microbiol Biotechnol 33(7):476–485

103. Swartz JR (2001) Advances in *Escherichia coli* production of therapeutic proteins. Curr Opin Biotechnol 12(2):195–201

104. Szlachcic A, Zakrzewska M, Otlewski J (2011) Longer action means better drug: tuning up protein therapeutics. Biotechnol Adv 29(4):436–441

105. Szymanski CM, Yao RJ, Ewing CP, Trust TJ, Guerry P (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol 32(5):1022–1030

106. 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(2):159–167

107. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72(2):211–222

108. Thomas MD, van Tilburg A (2000) Overexpression of foreign proteins using the *Vibrio fischeri lux* control system. Methods Enzymol 305:315–329

109. Tiwari A, Sankhyan A, Khanna N, Sinha S (2010) Enhanced periplasmic expression of high affinity humanized scFv against Hepatitis B surface antigen by codon optimization. Protein Expr Purif 74(2):272–279

110. Tripathi NK, Sathyaseelan K, Jana AM, Rao PVL (2009) High yield production of heterologous proteins with *Escherichia coli*. Def Sci J 59(2):137–146

111. Valdez-Cruz NA, Caspeta L, Perez NO, Ramirez OT, Trujillo-Roldan MA (2010) Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. Microb Cell Fact 9:18

112. van de Walle M, Shiloach J (1998) Proposed mechanism of acetate accumulation in two recombinant *Escherichia coli* strains during high density fermentation. Biotechnol Bioeng 57(1):71–78

113. Varley J, Birch J (1999) Reactor design for large scale suspension animal cell culture. Cytotechnology 29(3):177–205

114. Vasina JA, Baneyx F (1997) Expression of aggregation prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli cspA* and *tac* promoter systems. Protein Expr Purif 9(2):211–218

115. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24(4):179–185

116. Venturi M, Seifert C, Hunte C (2002) High level production of functional antibody Fab fragments in an oxidizing bacterial cytoplasm. J Mol Biol 315(1):1–8

117. Vera A, Arís A, Carrió M, González-Montalbán N, Villaverde A (2005) Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. J Biotechnol 119(2):163–171

118. Villalobos A, Ness J, Gustafsson C, Minshull J, Govindarajan S (2006) Gene designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinform 7(1):285

119. Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. BMC Mol Biol 8:100

120. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298(5599):1790–1793

121. Weir ANC, Nesbitt A, Chapman AP, Popplewell AG, Antoniw P, Lawson ADG (2002) Formatting antibody fragments to mediate specific therapeutic functions. Biochem Soc Trans 30:512–516

122. Weuster-Botz D (2000) Experimental design for fermentation media development: statistical design or global random search? J Biosci Bioeng 90(5):473–483

123. Wilson BS, Kautzer CR, Antelman DE (1994) Increased protein expression through improved ribosome-binding sites obtained by library mutagenesis. Biotechniques 17(5):944–953

124. Wong MS, Wu S, Causey TB, Bennett GN, San KY (2008) Reduction of acetate accumulation in *Escherichia coli* cultures for increased recombinant protein production. Metab Eng 10(2):97–108

125. Yang J, Moyana T, Mackenzie S, Xia Q, Xiang J (1998) One hundred seventy-fold increase in excretion of an FV fragment-tumor necrosis factor alpha fusion protein (sFV/TNF-α) from *Escherichia coli* caused by the synergistic effects of glycine and Triton X-100. Appl Environ Microbiol 64(8):2869–2874

126. Yang X (2007) Large-scale microbial production technology for human therapeutic products. In: Langer ES (ed) Advances in large-scale biopharmaceutical manufacturing and scale-up production. ASM Press, Washington, pp 329–362

127. Yang XM (1992) Optimization of a cultivation process for recombinant protein production by *Escherichia coli*. J Biotechnol 23(3):271–289

128. Yee L, Blanch HW (1993) Defined media optimization for growth of recombinant *Escherichia coli* X90. Biotechnol Bioeng 41(2):221–230

129. Yim S, Jeong K, Chang H, Lee S (2001) High-level secretory production of human granulocyte-colony stimulating factor by fed-batch culture of recombinant *Escherichia coli*. Bioprocess Biosyst Eng 24(4):249–254

130. Yoon SH, Kim SK, Kim JF (2010) Secretory production of recombinant proteins in *Escherichia coli*. Recent Pat Biotechnol 4(1):23–29

131. Zawada JF, Yin G, Steiner AR, Yang J, Naresh A, Roy SM, Gold DS, Heinsohn HG, Murray CJ (2011) Microscale to manufacturing scale-up of cell-free cytokine production-a new approach for shortening protein production development timelines. Biotechnol Bioeng 108(7):1570–1578

132. Zhang J, Greasham R (1999) Chemically defined media for commercial fermentations. Appl Microbiol Biotechnol 51(4):407–421

133. Zhang W, Xiao W, Wei H, Zhang J, Tian Z (2006) mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in *Escherichia coli*. Biochem Biophys Res Commun 349(1):69–78