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

4,300
Open access books available

117,000
International authors and editors

130M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

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

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

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter

Escherichia coli: A Versatile Platform for Recombinant Protein Expression

Wan-Keung Raymond Wong, Ka-Lun Ng, Thiyagarajan Sivakumar, Xiu-Hua Hu, Hao Wang and Lai-Cheuk Nelson Lai

Abstract

Among the living organisms, Escherichia coli has been the most common choice employed for recombinant protein expression. In addition to its well-characterized genetics, E. coli is fast growing, relatively cheap, and easy to handle. These fine properties, in conjunction with the success achieved in transforming plasmid DNA into E. coli, as well as the advent of various genetic engineering techniques in the 1970s, have enabled E. coli to be considered as the most favorable host for genetic manipulations. The recent advances in better comprehension of regulatory controls of gene expression and the availability of various novel approaches, which include both intracellular, e.g., through intein-mediated expression and self-cleavages, and extracellular, e.g., through the use of secretion signals, to achieve successful expression of the target proteins in E. coli further support the view that E. coli is the most promising host choice for heterologous protein expression.

Keywords: Escherichia coli, E. coli, recombinant protein expression, heterologous protein, authentic structures, fusion protein, affinity tags, secretion, excretion, inteins

1. Introduction

The achievements in unveiling the structure of DNA, deciphering the genetic code, understanding gene expression and regulation, and discovering extrachromosomal DNA (plasmid), restriction endonucleases and DNA ligases in the 1950s and 1960s laid the groundwork for the construction of the first chimeric (recombinant) DNA molecule [1]. In 1973, Cohen and Boyer reported their success in creating the first Escherichia coli transformant into which a recombinant plasmid molecule was introduced [2].

The possibility of inserting foreign DNA into E. coli has not only allowed the development of a vast number of molecular biology techniques for genetic manipulations, e.g., construction and characterization of cDNA libraries, DNA splicing and amplification, hybridization and sequencing, site-specific mutagenesis, research and applications of bacteriophages and DNA modifying enzymes, studies of regulation of gene expression, etc., but also the exploitation of E. coli for use as a surrogate host for the expression of heterologous proteins. Despite the presence of restriction
systems in *E. coli* [3, 4], foreign DNA from a wide variety of sources, ranging from a simple virus to a complex organism such as the human being, have been shown to be stably maintained in it. Therefore, although *E. coli* may not be employed as the final host cell for target protein expression, due to the difficulties of DNA manipulations, which include selection, characterization, and amplification of recombinant DNA constructs directly performed in the final host cell, *E. coli* is routinely recruited to fulfill the tasks. In this regard, *E. coli* acts as an indispensable stepping stone facilitating molecular genetic studies undertaken in eukaryotic and other bacterial host cells.

The fine properties including fast growth rate (with a doubling time of less than 30 min), ease of manipulation, relatively low cost of cultivation/protein expression, and the well-characterized genetics of *E. coli* are the key reasons why this bacterium has been commonly involved in the wide range of studies mentioned above. Thus, *E. coli* has been crowned the most popular host of choice employed for heterologous protein expression.

Being a Gram-negative bacterium, *E. coli* possesses two layers of cell envelope, the inner/cytoplasmic membrane and the outer membrane, between which is the narrow periplasmic space (Figure 1). Except for specific strains, e.g., a colicin-producing strain, *E. coli* does not produce proteins beyond the outer membrane in the living environment. Therefore, in the early days, *E. coli* was essentially employed for cytoplasmic/intracellular expression of recombinant proteins. However, the advancement of recombinant DNA technology in the 1970s–1980s not only enabled *E. coli* to serve as an efficient host system for protein expression but also facilitated *E. coli* to be transformed into a highly versatile expression platform, with which heterologous proteins may also be allowed to be produced in the culture supernatant. In addition, despite being expressed in the cytoplasm, target proteins resulting from a recently developed novel process, the intein-mediated expression approach [5, 6], have been shown to possess primary structures precisely the same as expected [7]. This new recombinant DNA approach may prove to be a practical

![Figure 1](image)

**Figure 1.** Schematic representation of the subcellular compartments of *E. coli* available for expression of heterologous proteins. These compartments such as cytoplasm, periplasm, and culture medium are shown. For a secretory protein, it is initially formed as a preprotein ( ), which consists of a signal peptide ( ). The preprotein is directed to the SecYEG pathway. Before reaching the secretion channel ( ), the preprotein is unfolded by the SecB protein ( ), followed by interacting with the SecA protein ( ), which helps hydrolyze ATP to support the active transport of the preprotein. The signal peptide is then cleaved off from the preprotein subsequent to the passage of the latter through the SecYEG pathway.
process for the recombinant production of medically valuable proteins that are preferred to share the authentic structures with their native counterparts.

Proteins expressed in eukaryotic cells are subjected to post-translational modifications (PTM), of which many of them do not appear to occur in \textit{E. coli} cells. Despite this deficiency, \textit{E. coli} is still the most common choice employed for the expression of eukaryotic proteins. For example, about 30\% of the medically valuable proteins produced using recombinant DNA approaches are expressed employing \textit{E. coli} as the host [8]. The finding of successful expression of eukaryotic proteins in \textit{E. coli} suggests that many target proteins may not be posttranslationally modified, and even some of them are, PTM may not have a direct effect on functional activities. The observation also supports the view that \textit{E. coli} will continuously play an indispensable role in heterologous protein expression. In choosing the most appropriate tactic for the expression of a heterologous protein in \textit{E. coli}, it is important that we understand both the target protein and the available methods of choice well. \textit{E. coli} is recognized as being a “versatile” host from the perspective that it may facilitate heterologous protein expression in all three of the subcellular compartments including: (i) cytoplasm, (ii) periplasm, and (iii) culture medium (Figure 1). In this communication, we discuss how those compartments may be employed to express foreign proteins that share widely different biochemical properties, under the condition that the presence of PTM is not a prerequisite.

2. Expression of recombinant proteins in the cytoplasm of \textit{E. coli}

The breakthrough achievements in the construction of recombinant DNA molecules [1] and the transformation of chimeric DNA constructs into \textit{E. coli} [2] in the early 1970s have paved the way for rapid advances in the development of recombinant DNA approaches to the expression of a wide collection of useful/valuable proteins of various origins. Due to the aforementioned fine properties of \textit{E. coli}, this Gram-negative bacterium has been extensively studied and exploited for use as a host to facilitate expression of heterologous proteins (Table 1).

2.1 Fusion protein approach

A common strategy in the expression of heterologous proteins is to fuse the target protein with a fusion partner, of which a familiar example is the enzyme \(\beta\)-galactosidase (\(\beta\)-Gal) expressed intracellularly in \textit{E. coli}. Being well-characterized in terms of its structure and regulation of expression [27, 28], in the early days, \(\beta\)-Gal was one of the few \textit{E. coli} products to be employed as a reporter protein, for which convenient detection assays [29] were available. Fusing the short mammalian somatostatin (Som) comprising only 14 amino acids (aa) to \(\beta\)-Gal, in 1977, Itakura et al. demonstrated (Figure 2), for the first time, successful expression of bioactive recombinant somatostatin in \textit{E. coli} [9]. In the work, Som was fused to \(\beta\)-Gal through the application of oligonucleotide assembly. Thus, expression of the two proteins, which was under the regulatory controls of the Lac operon, would result initially in a \(\beta\)-Gal-Som precursor. The \(\beta\)-Gal component played two important roles in the fusion: first, it offered a facile screening assay for the selection of potentially positive clones expressing \(\beta\)-Gal-Som; second, it served as a guardian protecting Som from being attacked by proteolytic degradation from the N-terminus.

Since Som is a short polypeptide consisting of only 14 aa residues [30], which does not consist of Met as a member, in engineering the aforementioned \(\beta\)-Gal-Som fusion, a Met residue was intentionally inserted precisely between \(\beta\)-Gal and Som, thus resulting in a \(\beta\)-Gal-Met-Som precursor in the work [9]. In vitro cleavage with
the chemical, CNBr, which specifically attacks Met, resulted in the separation of the fusion to yield authentic Som comprising 14 aa as the final product, which was subsequently shown to be bioactive (Figure 2) [9].

A similar approach was also applied to express bioactive human insulin in *E. coli* in 1979 (Figure 2), once again, taking advantage of β-Gal as the fusion partner and the absence of a Met residue in the polypeptide [10].

However, the above two examples appear to be the exception rather than the rule. Despite the application of β-Gal to serving as a fusion partner in many other cases of recombinant protein expression, due to the presence of one or more Met residues in the target proteins, the intriguing tactic of employing CNBr to cleave the fusion precursors to free the desired final protein to be impractical for routine use.

### 2.2 Direct expression

If a heterologous protein is insusceptible to proteolytic degradation by *E. coli* proteases, perhaps a simple method to achieve expression of the protein in *E. coli* is to clone its gene determinant downstream from a regulatory region comprising both the promoter and RBS sequences carried on a suitable expression vector. Examples

<table>
<thead>
<tr>
<th>Approach</th>
<th>Examples</th>
<th>Section concerned</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion expression</td>
<td>Hormone somatostatin</td>
<td>2.1</td>
<td>[9, 10]</td>
</tr>
<tr>
<td></td>
<td>Human insulin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct expression</td>
<td>Human growth hormone</td>
<td>2.2</td>
<td>[11–15]</td>
</tr>
<tr>
<td></td>
<td>Human fibroblast interferon</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human leukocyte interferon (LeIF A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human leukocyte interferon (LeIF B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse granulocyte-macrophage colony stimulating factor (mGM-CSF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tagged protein expression</td>
<td>Human insulin-like growth factor 1 (IGF-1)</td>
<td>2.3</td>
<td>[16–18]</td>
</tr>
<tr>
<td></td>
<td>Human parathyroid hormone (hPTH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human fibroblast growth factor 21 (hFGF21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intein-mediated</td>
<td>Cre recombinase</td>
<td>3</td>
<td>[5, 6, 19, 20]</td>
</tr>
<tr>
<td></td>
<td>Cecropin</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human basic fibroblast growth factor (hFGF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretion/excretion</td>
<td>Human epidermal growth factor (Secretion)</td>
<td>4</td>
<td>[21–26]</td>
</tr>
<tr>
<td></td>
<td>Hirudin</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human growth hormone (hGH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human granulocyte-macrophage colony stimulating factor (Hgm-CSF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human epidermal growth factor (Excretion)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Various approaches for heterologous proteins expression in *E. coli*. 

The above two examples appear to be the exception rather than the rule. Despite the application of β-Gal to serving as a fusion partner in many other cases of recombinant protein expression, due to the presence of one or more Met residues in the target proteins, the intriguing tactic of employing CNBr to cleave the fusion precursors to free the desired final protein to be impractical for routine use.

### 2.2 Direct expression

If a heterologous protein is insusceptible to proteolytic degradation by *E. coli* proteases, perhaps a simple method to achieve expression of the protein in *E. coli* is to clone its gene determinant downstream from a regulatory region comprising both the promoter and RBS sequences carried on a suitable expression vector. Examples
Escherichia coli: A Versatile Platform for Recombinant Protein Expression
DOI: http://dx.doi.org/10.5772/intechopen.82276

including human growth hormone [11], human hemoglobin [31], interleukin [32], etc., have been expressed in E. coli using this approach (Figure 2).

The alignment between the target gene and its expression regulatory elements could be conveniently achieved using site-directed mutagenesis. However, the translation initiator, N-formyl-methionine (fMet), which is present in proteins formed in bacteria, may cause adverse effects on the bioactivity and stability of the target protein [33]. The efficiency of removal of fMet in the cell is incomplete and is highly dependent on the adjacent two residues next to the initiator [34, 35].

Various strategies have been described to remove fMet from heterologous proteins including the use of both in vivo and in vitro approaches [34, 36, 37]. However, none of the available protocols is able to result in a homogeneous product that is free of fMet [34, 36]. The target protein, being contaminated by the presence of the undesirable fMet-bearing variant, may exhibit increased immunogenicity [33] and reduced levels of stability and bioactivity [34], which might have a correlation with fMet which has been speculated to serve as a degradation signal [38].

2.3 Applications of affinity tags

A major goal in recombinant DNA expression is to achieve efficient production of a target protein on a large scale. Common strategies including the use of: (1) plasmids with increased copy numbers such as the ones employing runaway replicons [39, 40]; (2) strong transcriptional control signals including PL, Tac, and T7 promoters [41, 42]; (3) efficient ribosome binding sites such as the Shine-Dalgarno sequence [43]; (4) inducible promoters which may be activated by heat shocking [44], light induction [45] or chemicals, e.g., isopropyl
The Universe of Escherichia coli

β-D-1-thiogalactopyranoside (IPTG); (5) a codon-optimized gene sequence [46, 47]; and (6) an efficient plasmid maintenance system. These various methods have been commonly applied, either individually or in conjunction with a fusion approach, to achieving efficient expression of target proteins in *E. coli*.

Although high yields of products may result from the application of above mentioned expression approaches, oftentimes, the products present themselves as insoluble inclusion bodies or aggregates. Unfortunately, these inclusion bodies are composed of denatured and misfolded proteins, which are functionally inactive [48, 49]. Due to the rearrangements of disulfide bridges in the aggregates, despite going through the processes of denaturation and renaturation, the target proteins are unlikely able to regain their functional activities [48].

Fusion of a target protein to an affinity tag presents a viable approach to not only the purification of the final product, but also the preservation of the product as a soluble protein. It has been shown that protein tags such as maltose-binding protein [50], glutathione S-transferase [51], small ubiquitin modifying protein [52, 53], and thioredoxin [54] might help improve the solubility of fusion products formed between the tags and target proteins (Figure 2). Given that a fusion product is expressed as a soluble and properly folded intermediate, and that it is readily purified using affinity chromatography and proteolytically processed at a recognition site engineered between the tag and target proteins, the frailty of this fusion approach is how the affinity tag may be removed from the target protein on condition that the latter till possesses the peptide sequence as stipulated. Thus, this approach may not be able to meet the stringent demand from therapeutic proteins of which any discrepancy found in their primary structures may result in undesirable side effects such as increased immunogenicity [33], reduced levels of stability and bioactivity [34, 55, 56], and worse still, greater tendency to promote malignancy. It is believed that target proteins bearing the authentic structures are as safe as their native counterparts in performing biological functions [57].

3. Intein-mediated expression of heterologous proteins

3.1 Inteins as fusion partners

Since the first intein, or protein intron, was discovered in the late 1980s [58], over 600 putative intein genes have been discovered [59]. Being able to undergo autocatalytic cleavages of themselves from sequences flanking their two termini, the N- and C-exteins, the application of inteins to the development of *E. coli* expression platforms has revolutionized the production of recombinant proteins in two different facets. First, fusion proteins formed between inteins and target proteins may undergo auto-cleavage activities in the cytoplasm of *E. coli* [5, 6, 60–63]. Second, despite taking place intracellularly, the detached target proteins possess the requisite structures, e.g., the authentic N-terminal sequences which are the same as those of their native counterparts [5, 6, 60–63].

3.2 Autocatalytic cleavages of intein-target fusion proteins: through an in vitro method

In the early days of exploiting the application of inteins to protein expression, fusion precursors formed among three components, comprising an N-terminal protein tag, a common example being a chitin-binding domain (CBD; [64]), a central intein, and a C-terminal target protein (CBD-I-TP), were frequently expressed as biologically inactive inclusion bodies in the cytoplasm of *E. coli* [65–67]. Subsequent to denaturation
and renaturation of the protein aggregates [67], the renatured precursors comprising CBD and the target proteins, e.g., Cre recombinase, α-1-antitrypsin, human epidermal growth factor [67] (Figure 2), collected in a chitin column was cleaved [64] by modulating the environmental conditions to release the target proteins [67] (Figure 3).

Being expressed as inclusion bodies, as discussed in Section 2.3, it is unlikely that the renatured CBD-I-TP molecules would all be bound to the chitin matrix or be correctly refolded. Therefore, the above described intein-mediated expression process working in conjunction with an in vitro autocatalytic cleavage protocol is expected to result in a substantial loss of bioactive target proteins.

3.3 Autocatalytic cleavages of intein-target fusion proteins: in vivo

Despite the inducibility of self-cleavages of inteins by modulating the environmental conditions [68], the exact mechanisms regarding how the induction works is not clear. Recent findings have shown that the ability of an intein element in fusion proteins to undergo self-cleavages appears to be dependent upon the presence of a pair of “well-matched” heterologous “exteins.” If this condition is fulfilled, autocatalytic cleavages might take place at the two terminal junctions where the intein is fused with the two exteins (Figure 3). It was demonstrated that when human epidermal growth factor (EGF) and basic fibroblast growth factor (bFGF) were precisely fused at the N- and C-termini, respectively, of the Sce VMA intein, auto-cleavage processing occurred [5] (Figure 2). Both EGF and bFGF were retrieved and shown to share not only authentic structures, but also potent bioactivities with their native counterparts [5].

Figure 3.
Processing of precursors formed between inteins and target proteins. The processing may be achieved by one of the following two methods: (A) induced in vitro cleavage; (B) in vivo cleavage. To attain a soluble precursor protein for processing in (A), the insoluble precursor protein express in the host cells is first required to be denatured and renatured. The solubilized precursor is then adsorbed onto an affinity column containing, e.g., chitin resin; changes of the conditions in the column may result in self-cleavage and hence the detachment of the target protein. In (B), the self-cleavage reaction takes place automatically in the host without extra input of effort. In the figure, hatched box ( ) denotes affinity tag, white box ( ) denotes intein, dotted box ( ) denotes target protein and affinity resin is denoted by . The correctly folded affinity tag, intein, and target protein are denoted as , , and , respectively. However, the in vitro cleavage process shown in (A) may also result in incorrectly folded proteins.
Moreover, since EGF was fused to the OmpA signal peptide (OmpA) in the abovementioned work (Figure 2), the EGF-VMA-bFGF fusion was also shown to be secretory and both EGF and bFGF were finally detected to be present in the culture medium of their E. coli host [5]. Interestingly, when EGF was absent in the fusion, thus leaving the formation of OmpA-VMA-bFGF, and when the positions of EGF and bFGF in the fusion were switched, thus leading to the expression of OmpA-bFGF-VMA-EGF, neither of the two precursors resulted in successful self-cleavages to yield authentic bFGF as the final product [5]. The results support the idea that not only the presence of a matched pair of exteins, but also their relative position in the fusion is important in effecting autocatalytic cleavages of the extein from their intein fusion partner.

Another noteworthy observation from the above work is the soluble nature of the fusion precursor, EGF-VMA-bFGF. This unusual condition, which contrasts markedly with the results of insoluble aggregates reported previously [64], has facilitated auto-cleavages of the fusion precursor to undergo self-cleavages directly in the cytoplasm, thereby avoiding the involvement of a time-consuming and ineffective process of denaturation and renaturation, followed by the extra time and effort spent on implementing the in vitro cleavage operation (Figure 3).

The in vivo autocatalytic processing approach introduced above may also be extended for use in the co-expression of other target proteins [60] (Figure 4). Moreover, through a combined protocol of gene amplification and refined fed-batch fermentation, the EGF-VMA-bFGF fusion has been upgraded to result in an expression of EGF-VMA-bFGF-VMA-bFGF as the precursor in E. coli [6]. Despite 92% bigger in size than EGF-VMA-bFGF, which was shown to have a mass of 73 kDa [5], EGF-VMA-bFGF-VMA-bFGF was found to be expressed as a soluble protein, which was still able to undergo autocatalytic cleavages to result in authentic and bioactive

![Figure 4](image_url)

*Figure 4.* Analysis of the CBD-DnaB-IFN precursor by polyacrylamide gel electrophoresis. Cell lysate samples were prepared from E. coli transformants harboring plasmid pTWIN-CBD-DnaB-IFN grown in 2×YT medium before and after 4 h of IPTG induction. Cell pellets were lysed by a chemical lysis protocol as described previously [5]. Lanes containing soluble proteins retrieved from the clear lysate (CL) and insoluble proteins retrieved from cell debris (Deb) were resolved as shown. The products CBD-DnaB (■) and IFN (△) resulting from autocatalytic cleavages of CBD-DnaB-IFN are denoted.
bFGF as the final product [6, 61–63]. In addition, fermentative production of EGF-VMA-bFGF-VMA-bFGF resulted in a dramatic improvement in the yield bFGF, amounting to 610 mg L\(^{-1}\) of cell culture [6], which was over 2.4 times higher than that resulting from the processing of EGF-VMA-bFGF expressed previously [5].

4. Expression of heterologous proteins across the inner membrane of E. coli

The approach of secretory expression of heterologous proteins stemmed from the work of W. Gilbert’s group, which employed the N-terminal 23 amino acid leader sequence of the E. coli penicillinase [69], to direct secretion of eukaryotic proteins, using rat proinsulin as the model protein, to the periplasmic space of E. coli in the late 1970s and early 1980s [69–72]. The secreted proinsulin was not only shown to possess an authentic structure, with the cleavage of the signal peptide done precisely [71], but also shown to be more stable than its cytoplasmic counterparts fused to defective signal sequences [72]. Over the next few years, different eukaryotic proteins, e.g., EGF [21], human interferon-α [73], hirudin [22], human growth hormone [23], and human granulocyte-macrophage colony stimulating factor [24], were also successfully expressed though secretion using various bacterial signal peptides in E. coli (Figure 2).

Meanwhile, E. coli mutants that were able to leak endogenous enzymes from the periplasm were isolated [74, 75]. The results suggested that heterologous proteins might also leak from the periplasm to the culture medium in E. coli. As expected, a few years later, heterologous proteins including bacterial endoglucanases [76, 77], a penicillinase of an alkalphilic Bacillus [78], as well as human proteins, such as β-endorphin [79], EGF, parathyroid hormone, and interleukin-6 [80], were expressed as extracellular products using either wildtype or leaky E. coli strains as hosts.

Not all proteins, e.g., the cytoplasmic enzyme—β-galactosidase, may be possibly expressed as secreted or excreted products in E. coli, despite their fusions to secretory proteins [81, 82] or directly to the signal peptides of these proteins [83]. Intracellular proteins do not appear to possess a molecular structure that is compatible with the SecYEG pathway, the major translocation machinery located in the inner membrane for protein transport [70, 82, 84–86]. On the other hand, when a naturally secreted target protein, e.g., EGF, is fused to a signal peptide, it may end up as a mature protein in either the periplasm [21] or the culture medium [80] of E. coli cells, depending essentially upon the efficiencies of expression and secretion of the protein (see below).

4.1 Secretory expression of target proteins in E. coli

In E. coli, several protein export systems, including the SecYFG (a trimeric complex comprising three polypeptides: SecY, SecE and SecG), Tat (twin-arginine translocation), and SRP (signal recognition particle) pathways which are embedded in the inner or cytoplasmic membrane, are responsible for the transport of proteins from the cytoplasm to the periplasm [86–89] (Figure 1). Among them, the SecYEG translocon is a general, conserved, and essential pathway which is found in both prokaryotic and eukaryotic cells [85, 86]. Being the major protein transport system, over 90% of the translocated proteins are secreted through the SecYEG pathway [87, 88] (Figure 1).

To enable proteins to be secreted using the SecYEG translocon, they are required to be expressed first as preproteins, which are fused at their N-termini with a short (commonly less than 24 amino acids) signal peptide [87]. In the cytoplasm, a
preprotein is maintained in an extended (export-competent) state by interacting with the SecB chaperone. Subsequent to an interaction formed between the signal and the SecA ATPase, the preprotein–complex then associates with the SecYEG pathway. With repeated pushes of SecA, the preprotein is secreted through the translocon in an ATP-dependent manner, followed by removal of the signal peptide by signal peptidase before the mature protein is released to the periplasm [87, 89, 90].

A wide range of heterologous proteins including degradative enzymes [91, 92], human hormones, and growth factors [5, 25, 26] have been successfully expressed as secretory proteins in *E. coli*. Many of the secreted proteins were not only shown to be bioactive, but also confirmed by sequence determination to possess the correct structures, supporting that the signal peptides fused at the N-termini of the preproteins had been removed correctly during the process of secretion [5, 25]. These advantages, together with lower levels of protease complexity and activity [72], and the relatively more oxidative environment that may help proper folding and disulfide-linkage formation [93, 94], enable the periplasmic space to be considered as a reasonable and appropriate destination for the expression of recombinant secreted proteins. Later on, with the help of various genetic and/or biochemical manipulations [95], or even merely through improving the levels of the secreted proteins concerned, which was referred as the “self-driven approach” [95], interestingly, the target proteins might then be allowed to leak out to the culture medium, a process termed excretion, which is essentially caused by non-specific leakage of periplasmic proteins (see below).

4.2 Excretory production of target proteins in *E. coli*

In the mid-1980s, researchers from different groups discovered that heterologous proteins expressed and secreted to the periplasm of *E. coli* might also be further excreted to the culture supernatant [22, 79]. For example, the development of sensitive screening assays, e.g., the Congo red plate assay (Figure 5), helped to confirm that the detection of a recombinant endoglucanase (Eng) encoded by the *cenA* gene of a Gram-positive bacterium, *Cellulomonas fimi* [77, 91, 96], in the culture supernatant of its *E. coli* host was due to a new phenomenon, excretion (extracellular production), rather than from cell lysis.

Efficient expression regulatory elements such as the strong promoters including tac, pl, and T7 [97, 98], the consensus ribosome binding site [99], the coding sequence for the potent OmpA signal [100], an effective inducible system, e.g., the *lac* operator/repressor system for transcriptional regulation [101, 102], etc., which are carried on a stable and high-copy number vector, e.g., pUC18 [103], have been made available to improve not only secretory, but also excretory expression of a wide variety of proteins in *E. coli*. The achievement of this research milestone was well exemplified by the development of an efficient protocol for extracellular production of EGF [104]. In early attempts to express EGF as a secretory protein in *E. coli*, the relatively weak *phoA* promoter was employed to perform transcription of the *egf* gene and the less efficient PhoA signal peptide to direct EGF for secretion. Despite the demonstration of EGF secretion, the EGF detected in the periplasm was only at a considerably low level of 2.4 mg L⁻¹ [21]. However, when the *tac* promoter and the *ompA* leader sequence were employed to facilitate EGF expression and secretion, respectively, it was shown that the level of excreted, but not secreted, EGF was markedly improved in *E. coli* cells [80] (Figure 6). Moreover, further improvements in EGF expression resulted in a dramatic increase in the yield of excreted EGF [26, 95, 104, 105]. Similar trends were observed in increasing the levels of excretory production of other heterologous proteins, e.g., *C. fimi* cellulases Eng [91, 106, 107] and exoglucanase, as well as bFGF [6].
4.3 Difficulties in implementing an effective excretory process and potential solutions

The findings described above support the view that excretion is a promising approach for recombinant production of heterologous proteins in *E. coli*. However, it has been shown that not all naturally secreted proteins may be expressed using excretion, despite using efficient transcriptional and translational controls, as well as effective secretion signal. For example, using the same regulatory elements which enabled a high level (325 mg L$^{-1}$) of excretion of the 53-amino acid (aa) EGF peptide in *E. coli* [26, 95, 104, 105], attempts to produce authentic bFGF (146 residues) by excretion in *E. coli* were unsuccessful [5]. One might wonder whether the discrepancy between the results of EGF and bFGF excretion was due to the marked difference between their molecular sizes. However, cellulases such as Eng and Exg, which possess large mature forms comprising 418 aa [91] and 443 aa [108], respectively, have been shown to be efficiently produced by excretion in *E. coli* [92, 107, 109]. Therefore, in addition to the molecular size of a heterologous protein which might have some effect on the efficiency of excretory production of the protein (see below), it appears that other factor(s), which may be associated to either the protein itself or the host (or both), play a crucial role in determining whether a protein may be expressed as a secretory/excretory product or not.

A major hurdle for excretory production of heterologous products is the dramatic cell death during enhanced expression of the preproteins—the fusion precursors formed between signal peptides and target proteins. A model designated “Saturated Translocation” was proposed to explain the phenomenon of cell lethality resulting from hyper-expression of the preproteins [110]. According to the model, when a preprotein exceeded a tolerable level, it would saturate the capacity of the SecYEG pathway and interfere with its normal function in exporting endogenous proteins. These functional disorders resulted finally in cell death [110]. The model also explained why heterologous proteins of different sizes, which were undergoing secretory expression, would trigger rapid cell death (Figure 7) if the presence of their preproteins had exceeded their individual allowable thresholds, the “Critical Values (CV).” A CV is defined as the largest quotient between an intracellular...
preprotein and its secreted mature counterpart that was tolerable by the host cells [92]. Deletion of the signal peptide from its mature partner, despite the possibility of incurring the formation of inclusion bodies [48], interestingly it would help avoid the onset of the deadly effect resulting from an efficiently expressed secretory protein [110]. The results clearly indicate that the bottleneck of secretory/excretory production of a heterologous protein is at the stage of secretion.

Different approaches have been attempted to attain or even improve the CV, and hence the maximum production of a secretory heterologous protein on a per cell basis. Since cell death results from hyper-expression, strategies of optimizing, rather than maximizing, protein expression, e.g., less efficient promoters [92, 107, 111] and start codon [92], as well as defined minimal media and sub-optimal cultivation conditions [26] have been employed and shown to provide beneficial effects. More encouragingly, excretory production of Exg was
markedly enhanced when the level of Phage shock protein A (PspA) was elevated in the same host [109]. In the presence of additional PspA, the CV of secretory Exg was shown to be markedly increased from 20/80 to 45/55 [109]. Presumably, PspA helped the host cells to maintain membrane integrity and an energized membrane [112–114], which was readily equipped to cope with the “stress,” the presence of secretory Pre-Exg, by efficiently transporting it through the SecYEG pathway [109].

5. Conclusions

Since the advent of recombinant DNA technology in the 1970s, E. coli has been the most favorable host choice for the expression of heterologous proteins. Strategies including both intracellular and secretory methods have been designed for the expression of proteins of interest. Despite possessing an outer membrane, a wide variety of naturally secreted proteins including hormones, factors, and degradative enzymes have also been shown to be produced as extracellular products in E. coli. In undertaking both intracellular and secretory/excretory protein expression, a fusion approach is commonly adopted. Affinity tag proteins including β-galactosidase, glutathione S-transferase, and 6xHis-tag have been employed to form fusion precursors with desired proteins. To enable separation between the tag and target proteins, a protease cleavage site is required to be placed between the two proteins. However, on the one hand, it may be difficult to achieve the exact processing result through proteolytic cleavage. On the other hand, it is cost-ineffective to implement proteolytic cleavage on a large scale. Fusions of target proteins with inteins and secretion signal peptides have presented a practical approach to protein cleavages in cells without relying on the use of external proteases. It has been well demonstrated that using both methods of protein fusion, target proteins possessing the exactly processed sequences are obtainable through autocatalytic or signal peptidase cleavages in vivo in E. coli.

Acknowledgements

This study was funded by RGC project: GRF16101515 and Research Contracts: 13142580CLIL07W011 and 1617219-0.
The Universe of *Escherichia coli*

### Conflict of interest

The authors declare that they have no conflict of interest.

### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid(s)</td>
</tr>
<tr>
<td>bFGF</td>
<td>basic fibroblast growth factor</td>
</tr>
<tr>
<td>β-gal</td>
<td>β-galactosidase</td>
</tr>
<tr>
<td>CBD</td>
<td>chitin-binding domain</td>
</tr>
<tr>
<td><em>C. fimi</em></td>
<td><em>Cellulomonas fimi</em></td>
</tr>
<tr>
<td>CNBr</td>
<td>cyanogen bromide</td>
</tr>
<tr>
<td>CV</td>
<td>critical value</td>
</tr>
<tr>
<td>DnaB</td>
<td>DnaB helicase</td>
</tr>
<tr>
<td>EGF</td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td>Eng</td>
<td>endoglucanase</td>
</tr>
<tr>
<td>Exg</td>
<td>exoglucanase</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>fMet</td>
<td>N-formyl-methionine</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>I</td>
<td>Intein</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo dalton(s)</td>
</tr>
<tr>
<td>MBP</td>
<td>maltose-binding protein</td>
</tr>
<tr>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>OmpA</td>
<td>outer membrane protein A</td>
</tr>
<tr>
<td>PhoA</td>
<td>alkaline phosphatase</td>
</tr>
<tr>
<td>Pre-</td>
<td>premature</td>
</tr>
<tr>
<td>PspA</td>
<td>phage shock protein A</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>Sce</td>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>SDA</td>
<td>self-driven approach</td>
</tr>
<tr>
<td>SecYEG</td>
<td>trimeric complex of SecY, SecE, and SecG</td>
</tr>
<tr>
<td>Som</td>
<td>somatostatin</td>
</tr>
<tr>
<td>Ssp</td>
<td><em>Synechocystis</em> species</td>
</tr>
<tr>
<td>TP</td>
<td>target protein</td>
</tr>
<tr>
<td>VMA</td>
<td>vacuolar ATPase subunit</td>
</tr>
</tbody>
</table>
Author details

Wan-Keung Raymond Wong*, Ka-Lun Ng, Thiyagarajan Sivakumar, Xiu-Hua Hu, Hao Wang and Lai-Cheuk Nelson Lai
Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

*Address all correspondence to: bcwkrw@ust.hk
References


[38] Piatkov KI, Vu TT, Hwang C-S, Varshavsky A. Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. Microbial Cell. 2015;2:376


[46] Hale RS, Thompson G. Codon optimization of the gene encoding a domain from human type 1 neurofibromin protein results in a threefold improvement in expression level in *Escherichia coli*.
Protein Expression and Purification. 1998;12:185-188


[61] Wong WKR. Means and method for expression of authentic human epidermal growth factor and/or basic fibroblast growth factor in cytoplasm and/or culture medium of *Escherichia coli*. 2017; US Patent No. 9,580,484 B2

epidermal growth factor and/or basic fibroblast growth factor in cytoplasm and/or culture medium of *Escherichia coli*. 2017; Hong Kong Short-Term Patent No. HK1219200

[63] Wong WKR, Kwong WY. Improved means and method for hyper-production of authentic human basic fibroblast growth factor in *Escherichia coli*. 2017; Hong Kong Short-Term Patent No. HK1228651


[76] Collmer A, Wilson DB. Cloning and expression of a Thermomonospora
Escherichia coli: A Versatile Platform for Recombinant Protein Expression
DOI: http://dx.doi.org/10.5772/intechopen.82276


[80] Wong WKR, Sutherland ML. Excretion of heterologous proteins from *E. coli*. 1993; US Patent No. 5223407


[82] Bieker KL, Silhavy TJ, PrlA is important for the translocation of exported proteins across the cytoplasmic membrane of *Escherichia coli*. Proceedings of the National Academy of Sciences. 1989;86:968-972


[92] Fu ZB, Ng KL, Lam CC, Leung KC, Yip WH, Wong WKR. A
two-stage refinement approach for the enhancement of excretory production of an exoglucanase from *Escherichia coli*. Protein Expression and Purification. 2006;48:205-214


[95] Wong WKR, Fu ZB, Wang Y, Ng KL, Chan KN. Engineering of efficient *Escherichia coli* excretion systems for the production of heterologous proteins for commercial applications. Recent Patents on Chemical Engineering. 2012;5:45-55


