



Cite this article: Jia B, Jeon CO. 2016 High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biol.* **6**: 160196. <http://dx.doi.org/10.1098/rsob.160196>

Received: 30 June 2016
Accepted: 3 August 2016

Subject Area:
biotechnology

Keywords:
high-throughput, recombinant protein expression, *Escherichia coli*, 5'UTR and N-terminal codons, fusion tag, membrane protein

Authors for correspondence:

Baolei Jia
e-mail: baoleijia@cau.ac.kr
Che Ok Jeon
e-mail: cojeon@cau.ac.kr

High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives

Baolei Jia and Che Ok Jeon

Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea

id BJ, 0000-0001-6434-5604; COJ, 0000-0003-1665-2399

The ease of genetic manipulation, low cost, rapid growth and number of previous studies have made *Escherichia coli* one of the most widely used micro-organism species for producing recombinant proteins. In this post-genomic era, challenges remain to rapidly express and purify large numbers of proteins for academic and commercial purposes in a high-throughput manner. In this review, we describe several state-of-the-art approaches that are suitable for the cloning, expression and purification, conducted in parallel, of numerous molecules, and we discuss recent progress related to soluble protein expression, mRNA folding, fusion tags, post-translational modification and production of membrane proteins. Moreover, we address the ongoing efforts to overcome various challenges faced in protein expression in *E. coli*, which could lead to an improvement of the current system from trial and error to a predictable and rational design.

1. Introduction

High-throughput studies can be defined as research that allows thousands of concurrent measurements of biological molecules to be obtained and thus makes large-scale repetition feasible. This technology originated in the early 1990s when the first automated DNA sequencers were developed and human genome sequencing was initiated [1]. In the post-genomic era, the use of high-throughput techniques has increased dramatically in terms of measuring DNA, RNA, proteins, lipids and metabolites [2], and these techniques have been successfully applied to answer diverse biological questions related to cancer biology, ecology, cell biology and systems biology [3].

Protein expression and purification play a central role in biochemistry. Recombinant proteins can be expressed using prokaryotic systems (*Escherichia coli* and *Bacillus subtilis*), eukaryotic systems (yeast, insect cells and mammalian cells) or *in vitro* systems. The *E. coli* system is the first-choice host for the initial screening of recombinant protein expression, because these cells can be readily manipulated, are cultured inexpensively and grow rapidly [4,5]. In recent years, numerous new strains, vectors and tags have been developed to overcome the limitations of this system, which include codon bias, inclusion body formation, toxicity, protein inactivity, mRNA instability and lack of post-translational modification [4].

The *E. coli* expression system has been widely examined, but protein expression and purification performed using this system are labour-intensive and time-consuming. Thus, a parallel and high-throughput approach must be employed in protein expression and purification, which has been the bottleneck in studies of protein function, structure and application in the post-genomic era [6]. As high-throughput methods of protein production were proposed at the beginning of this century [7], the techniques have become widely available [8–11], and recombinant proteins in inclusion body forms have even been expressed and purified in a parallel approach [12]. We also developed our own systems for purifying proteins from archaea in parallel [13–15]. Because

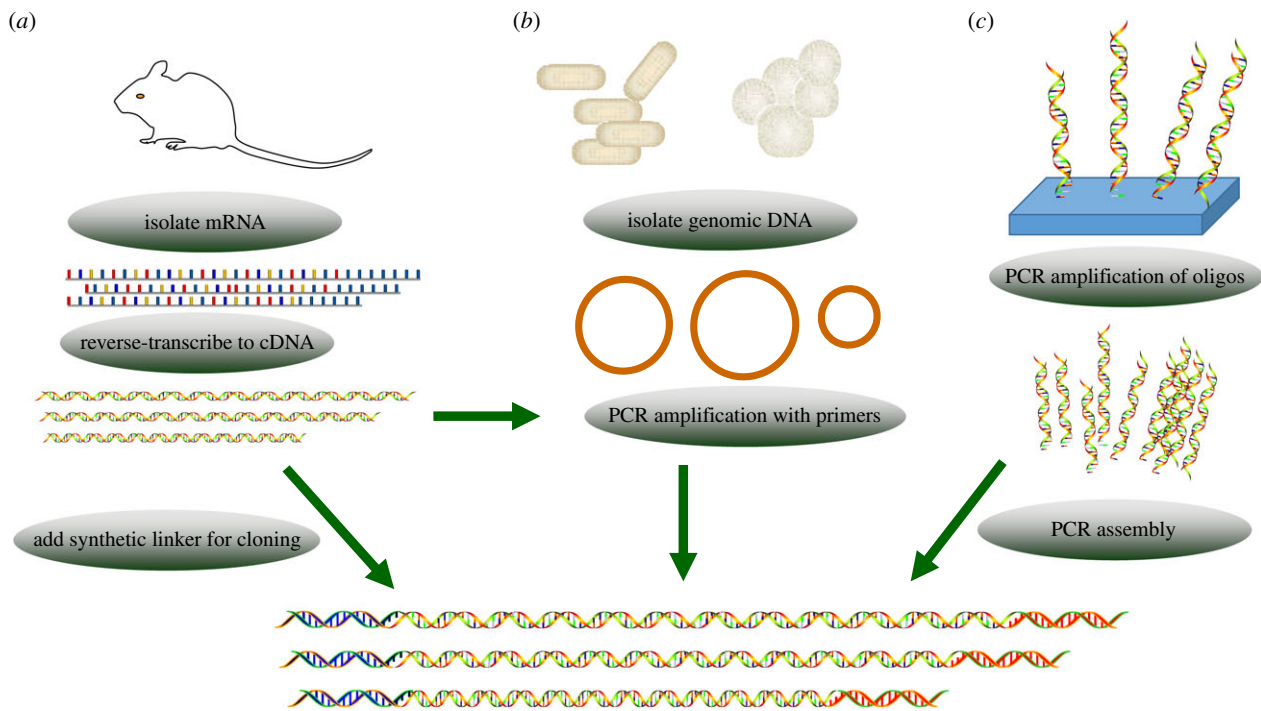


Figure 1. Three strategies for preparing target genes. (a) Target genes can be obtained from a cDNA library after reverse transcription. (b) PCR can be used to amplify genes from a cDNA library or genomic DNA. (c) Array-based gene synthesis through the assembly of short oligos can be used to produce customized genes.

numerous advances in these methods have been made over the past few years, in this review, we discuss the advantages and disadvantages of the current methods—specifically, those targeting gene cloning, vector construction, fusion tags and host strains.

2. High-throughput preparation of target genes

Historically, collections of genes to be expressed have been directly cloned from cDNA libraries as a pool into specific vectors (figure 1a) [16]. This method was used by Büsow *et al.*, who constructed a human fetal brain cDNA expression library in *E. coli* in 2000 [17]. The library contained a total of 193 536 clones, but only 37 830 (19.6%) clones expressed proteins. Further investigation revealed that some of the genes were not in the correct reading frame or contained partial coding sequences. Subsequently, a novel human cDNA expression library enabling the selection of open reading frames based on histidine prototrophy was developed in yeast [16]; in this library, approximately 60% of the clones were in the correct open reading frame. However, there are two limitations to the application of expression libraries, most notably in mammalian cells. First, the presence of untranslated regions at both ends of clones makes it challenging to attach fusion tags to either end of the proteins of interest. Second, although the process is laborious, genes of interest must be frequently fished out of a library for use in experiments [18].

Polymerase chain reaction (PCR) is the most widely used technique for obtaining target genes and is invariably the first step in any effort to express recombinant proteins (figure 1b). After genes of interest have been selected, a batch of primers can be designed based on the coding sequences using online tools such as PRIMERCE [19] and HTP-OLIGODESIGNER [20]. High-throughput PCR and PCR-product purification are now

mature technologies that can be completed using automated laboratory workstations [21]. However, problems such as the absence of a band (faint band) in gels, non-specific bands and primer-dimers may occur after PCR and slow the experimental process. These problems can be overcome by adjusting PCR parameters such as annealing temperature and primer concentration or by using the cloning methods discussed below.

Another approach used for obtaining target genes is *de novo* synthesis of DNA (figure 1c). Solid-phase (on-column) DNA synthesis involving chemical methods has been traditionally used, but the difficulty of synthesis increases with DNA length. Moreover, the synthesis can cost approximately \$0.15 per base, and considerably more for high-throughput synthesis. New array-based methods for synthesizing long DNA sequences with increased accuracy have been developed [22–24] and are expected substantially lower the synthesis cost [25]. The main advantage of *de novo* gene synthesis is that researchers can freely design genes of interest without limitations imposed by the use of natural templates [26]. Moreover, the use of codon-optimized genes can ensure reliable expression, increased protein yield and protein solubility [27]. With further developments in the technique, the applicability of *de novo* DNA synthesis to high-throughput assays is expected to increase.

3. High-throughput gene-cloning systems

After obtaining target genes, the next step is high-throughput construction of expression vectors. Various cloning methods have been developed to make the process simple, time-efficient and cost-effective (figure 2). Based on the underlying principle, the methods can be classified as restriction enzyme (RE)-based cloning, recombination-based cloning, and annealing-based or ligation-independent cloning (LIC). The advantages and limitations of these methods have been discussed in previous reviews [18,28,29]. In recent years, vast improvements in these methods have been made. Here, we concentrate on

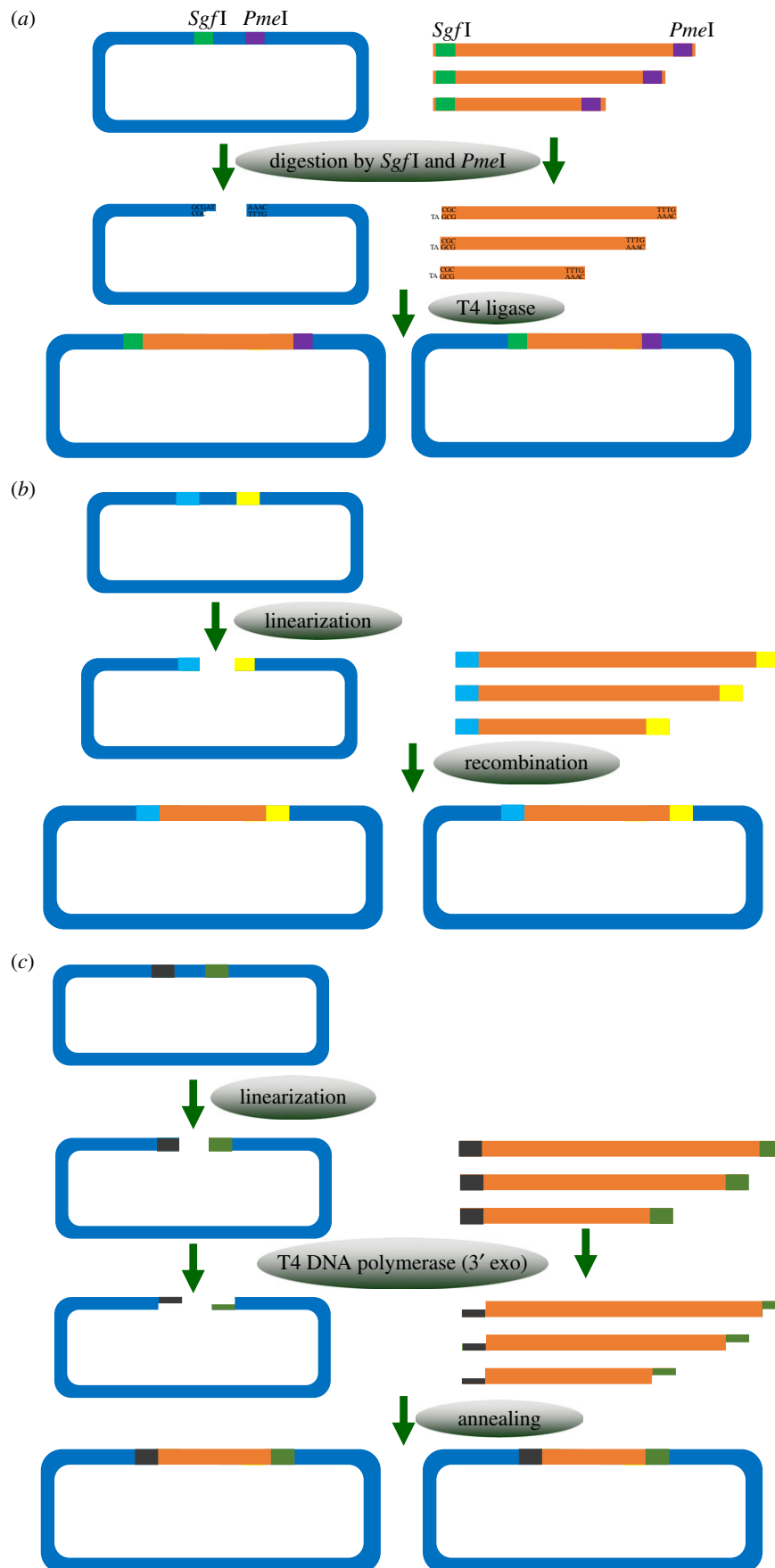


Figure 2. Schematic diagrams and principles of the construction of recombinant expression vectors. Target genes featuring two adapters are obtained from PCR or gene synthesis. (a) Construction of expression vectors using restriction enzymes and ligases. The vector and target genes harbouring restriction sites are digested using two rare-cutting enzymes, *SgfI* and *PmeI*. The linearized expression vector and inserts are ligated using T4 ligase to create the construct. (b) Construction of expression clones using recombination-based methods. The target genes are flanked by 15–25 bp recombination sites. Recombinase-mediated recombination between the homologous sites present in the insert and vector generates the final vector. (c) Construction of expression clones using LIC methods. Linearized vectors and target genes containing complementary 5'-tails are digested using enzymes possessing exonuclease activity in order to increase the proportion of recessed ends. The overhangs can anneal and are ligated *in vivo* after transformation into *E. coli*.

the most basic principles and the latest innovations in the existing methods.

3.1. Restriction enzyme-based cloning

RE-based cloning performed with DNA ligation has been used for four decades, but it was previously considered to be unsuitable for high-throughput methods because appropriate and compatible REs must be selected for each cloning procedure [7]. The method has received increased attention since 2006, when *SgfI* and *PmeI*, the two most rare-cutting REs in the human DNA, were used and the Flexi Cloning system was developed by Promega (Madison, WI, USA). The combination of *SgfI* and *PmeI* has been suggested to allow the cloning of more than 95% of genes of selected model organisms (figure 2a) [30]. The experimental procedure is similar to that use for conventional RE-based cloning: target genes are amplified using primers containing adapter sequences and then digested by two enzymes. The vector is also digested, releasing highly toxic *barnase* gene for lethal selection, which can be used as a marker against the parental vector. Subsequently, the target gene and vector are ligated and transformed into competent cells. Nagase *et al.* [31] used the Flexi Cloning system to produce proteins from 1929 open reading frame clones of human genes, demonstrating that this system can be successfully used in a high-throughput manner.

The Golden Gate method [32] relies on the RE *BsaI*. This method involves restriction digestion and ligation cycling in one tube, which can greatly increase efficiency. One potential limitation of this method may be the occasional presence of one or several internal *BsaI* site(s) in the gene of interest. An improvement has been made using *SapI* with a rarer cut site than that of *BsaI* [33]. Another method termed methylation-assisted tailorable ends rational (MASTER) uses the endonuclease *MspJI*, which specifically recognizes methylated 4-base-pair (bp) sites. Because this modification avoids cuts on corresponding sites within the fragments amplified by PCR, the MASTER method is more suitable for high-throughput cloning [34]. However, it requires expensive methylated primers and PCR amplification of regions, which may introduce errors in longer regions [35].

RE-based cloning methods may hold greater promise than the original methods, because they will be considerably easier to set up for researchers who continue to use traditional digestion–ligation protocols. With the modification of Flexi Cloning and Golden Gate cloning, RE-based cloning methods are expected to emerge as simple, efficient, universal and cost-effective methods for protein production.

3.2. Recombination-based cloning

Recombination-based cloning became widely used following the introduction of three cloning systems: Gateway (Thermo Fisher Scientific, Waltham, MA, USA), Echo Cloning (Thermo Fisher Scientific) and Creator (Clontech, Mountain View, CA, USA). Other commercial kits have also been developed, such as Cold Fusion from System Biosciences (Palo Alto, CA, USA) and CloneEZ from GenScript (Piscataway, NJ, USA). In these systems, a site-specific recombinase is employed to construct the required recombinant vector without using any REs and ligases (figure 2b). Gateway may be the most popular recombination-based cloning technology for high-throughput approaches and has been used since the late 1990s. The

Gateway cloning system exploits the site-specific recombination system used by bacteriophage λ to shuttle sequences between plasmids bearing flanking-compatible recombination attachment (*att*) sites. Once captured as an entry clone, a DNA fragment can be recombined into a variety of destination vectors, resulting in expression clones that can be used in specific applications. The recombination reactions are driven by two enzyme blends known by their commercial names: BP Clonase and LR Clonase [36]. One of the main advantages of the Gateway method is that once an entry clone has been made, the gene of interest can be easily subcloned into a wide variety of destination vectors using the LR reaction.

However, the general use of recombination methods has been limited by high costs and restrictions in the sequence or hosts [37]. Zhang *et al.* [38] created the Seamless Ligation Cloning Extract (SLiCE) method to assemble DNA fragments into vectors in a single *in vitro* recombination reaction using cell extracts from a modified DH10B *E. coli* strain expressing an optimized λ prophage Red recombination system. Motohashi [39] further modified the method by using several common *RecA*[−] *E. coli* laboratory strains such as DH5 α , JM109, DH10B, XL10-gold and Mach1 T1 with careful harvesting (at late log phase) and lysis (at 4°C). Moreover, the cell extracts can be prepared in a simple buffer containing Triton X-100 rather than the expensive commercial lytic reagent [40,41]. The homemade SLiCE from the laboratory strain JM109 can be used in place of the commercial kit at a cost of approximately \$0.003 per reaction [41]. The SLiCE-cloning protocol is a simple, convenient and ultra-low-cost method for performing high-throughput cloning.

3.3. Ligation-independent cloning

LIC, developed 26 years ago [42], enables directional cloning of any insert after the generation of DNA fragments containing single-stranded complementary ends. The lack of requirement for REs, ligases or recombinases makes LIC inexpensive and easily adaptable to high-throughput performance. However, LIC still requires enzymes such as T4 DNA polymerase and T5 exonuclease, depending on the protocols used, to generate single-stranded complementary ends in target genes and vector sequences (figure 2c). Several effective and convenient methods based on the LIC principle have been developed, including Gibson Assembly from NEB (Ipswich, MA, USA) [43], In-Fusion from Clontech [44], polymerase incomplete primer extension cloning [45], sequence and LIC [46], and overlap extension cloning [47,48]. The Gibson Assembly method [39] uses T5 exonuclease to remove portions of the 5' ends to generate single-stranded complementary overhangs, which are joined together covalently by fusion DNA polymerase and Taq DNA ligase. In a one-step isothermal *in vitro* reaction at 50°C, the fragments can be assembled into a single circular DNA molecule. Since its introduction 7 years ago, the Gibson Assembly method has become a preferred cloning method. Gibson Assembly allows the insertion of one or more DNA fragments into virtually any position of the linearized vector and does not rely on the presence of restriction sites within a particular sequence to be synthesized or cloned. Advantages of using Gibson Assembly in high-throughput cloning include speed, efficiency, scarless assembly with vector and versatility [49].

The LIC method has been successfully used for high-throughput cloning of genes: 130 genes encoding glycoside

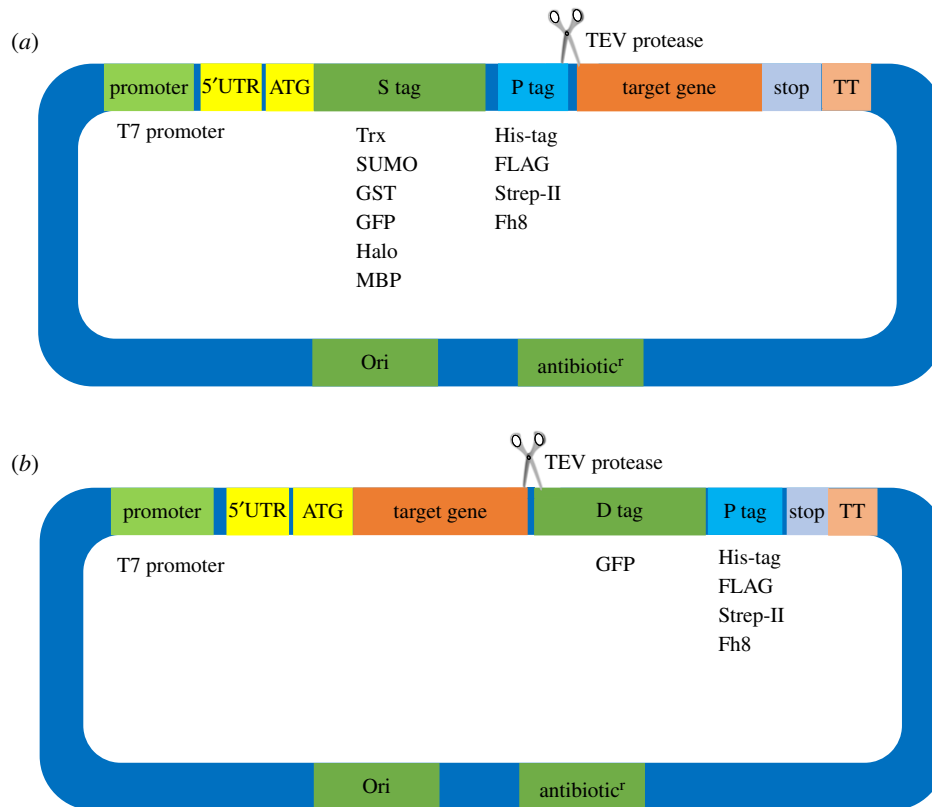


Figure 3. Basic expression vectors for high-throughput expression in *E. coli* of (a) cytoplasmic proteins and (b) membrane proteins. The T7 promoter is used to control expression of the protein in *E. coli*. The high-throughput assay requires tandem affinity tags, larger tag for protein expression initiation, protein solubility and soluble detection, and smaller tag for purification. TEV protease can be used to remove the tags. The tags for membrane proteins are located at the C-terminus for protein targeting, and GFP is a favourable choice for use as an indicator of protein folding. D tag, detection tag; P tag, purification tag; S tag, solubility and translation initiation tag; TT, transcriptional terminator; 5'UTR, 5' untranslated region.

hydrolases from 13 different organisms were cloned in parallel using LIC and subjected to protein expression screening in *E. coli* [50]. The method also allowed the automated assembly of more than 600 genes encoding transcription activator-like effector nucleases from *Xanthomonas* species in a single day [51]. Moreover, a three-person team cloned 2125 genes from *Pyrococcus furiosus* in three weeks and obtained at least 80% positive clones in a 96-well-plate cloning format using a modified λ -exonuclease-based LIC method [52].

4. Expression vectors for high-throughput protein expression

An *E. coli* expression vector possesses the same features found in any vector, such as a selection marker (e.g. antibiotic resistance), origin of replication, transcriptional promoter, 5' untranslated region (5'UTR) and translation initiation site (figure 3). Another critical feature of these expression vectors is the presence of a fusion tag(s) that is transcribed in-frame with the target gene in contrast to the aforementioned elements. Among these various elements, the promoters, 5'UTR, N-terminal codons and fusion tags most strongly affect transcription, protein yields, solubility and purification.

4.1. Promoters

An effective promoter for heterologous protein expression in *E. coli* has four key characteristics: first, the promoter is sufficiently strong to allow the accumulation of recombinant

protein to greater than or equal to 10–30% of the total cellular proteins; second, it exhibits minimal basal transcriptional activity, and thus unwanted transcription is avoided before induction; third, the promoter enables simple and inexpensive induction; and fourth, promoter activity can be precisely tuned.

The Arabinose promoter and hybrid promoters (*trc* and *tac* promoters) are widely used in protein expression. The Arabinose promoter exhibits the lowest basal transcriptional activity, but the efficiency of repression is gene-dependent and the repression level does not always reach zero [53,54]. By contrast, hybrid promoters exhibit leaky expression, and thus these promoters can be problematic for protein expression [55].

The Arabinose promoter and hybrid promoters are considered to be strong promoters, but are not as strong as the T7 promoter [56]. The pET expression system featuring the T7 promoter is by far the most widely used system for heterogeneous expression in *E. coli* [57]. T7 promoter activity is strong, and a recombinant protein can accumulate to up to 50% of total cellular proteins [58]. T7 expression hosts such as DE3 strains contain a chromosomal copy of the T7 phage RNA polymerase gene under control of the *lac* promoter derivative *lacUV5*. When isopropyl β -D-1-thiogalactopyranoside (IPTG) is added, LacI binding to the *lac* operator is inhibited, allowing for the expression of T7 polymerase, which transcribes the target gene and leads to recombinant protein production (figures 3 and 4) [59]. Recombinant protein expression can be controlled by coexpressing T7 lysozyme, which inhibits transcription by T7 RNA polymerase [60]. Moreover, previous studies have demonstrated that mutations in the *lacUV5* promoter can

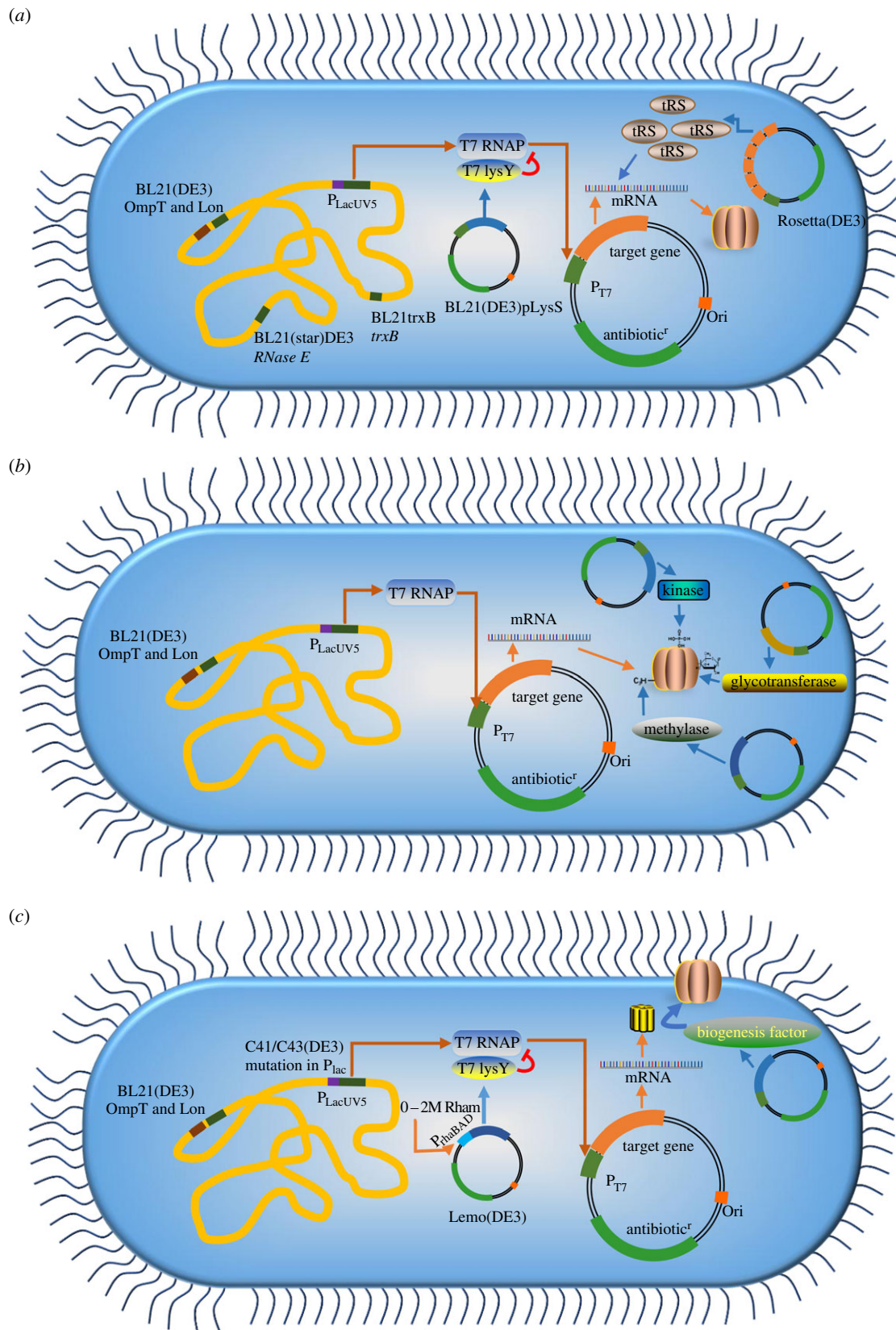


Figure 4. *Escherichia coli* strains for protein expression. (a) *Escherichia coli* strains widely used in recombinant protein production. In the expression vector, the target gene is under control of the T7 promoter. In the *E. coli* genome, the gene encoding T7 RNA polymerase is under control of the *lacUV5*. The strain BL21(DE3) is deficient in OmpT and Lon proteases. BL21STAR(DE3) is mutated in RNase E, reducing mRNA degradation. BL21trxB promotes the formation of disulfide bonds. In BL21pLysS(DE3), T7 lysozyme is expressed, and the enzyme inactivates any T7 RNA polymerase that may be produced without induction. Rosetta strains are designed to improve the expression of proteins encoded by genes containing rare codons used in *E. coli*. (b) Strategy for expressing a protein with post-translational modification in *E. coli*. Genes encoding kinases, glycosyltransferases, methylases, ligases or other modifying enzymes are coexpressed in order to produce post-translationally modified proteins. (c) Overview of *E. coli* strains used in membrane protein production. Walker strains (C41(DE3) and C43(DE3)) are commonly used to overcome the toxicity of membrane proteins. In Lemo21(DE3), expression can be tuned by adding different concentrations of L-rhamnose to the culture. Coexpression of membrane protein biogenesis factors may also facilitate the localization of target proteins. lysY, lysozyme; RNAP, RNA polymerase; tRS, tRNA synthetase.

govern the expression of T7 RNA polymerase and lower basal transcription [61]. Tunable expression can be achieved by varying the level of lysozyme produced under the control of the exceptionally well-titratable rhamnose promoter [62]. These advantages make the T7 promoter an attractive choice for the high-throughput production of recombinant proteins.

4.2. 5'UTR and N-terminal codons

Gene expression in *E. coli* is influenced by the efficiency of translation, particularly by the initiation step [63]. Both the 5'UTR upstream from the initiation codon and 5' coding region of a gene transcript are closely related to translation initiation and protein expression [64]. Structural features of the 5'UTR play an important role in controlling translation efficiency, as protein expression is initiated by binding of the ribosome to the Shine–Dalgarno (SD) sequence in the 5'UTR. For example, nucleotide changes to the 5'UTR causing differential formation of mRNA secondary structures can affect protein production levels by up to 600-fold [65]. The spacing and nucleotide sequences between the SD sequence and initiation triplet also have a marked effect on translation efficiency and protein production [66,67]. Optimization of the nucleotide sequences at the junction between the pET vector and coding sequence may enhance protein production [68]. Sequence variants in the region modulate protein expression by as much as 1000-fold; low GC content and relaxed mRNA stability in this region are key, but are not the only factors affecting high expression [68].

Furthermore, the 5' coding region can also influence translational initiation and gene expression, as the ribosome occupies approximately 15–25 nucleotides on either side of the initiation codon [69,70]. In bacteria, selection pressure favours codons that reduce mRNA folding around the translation start, regardless of whether these codons are frequent or rare [71]. However, rare codons are enriched at the N-terminus of natural genes in most organisms [72,73]. Rare codons at the beginning of genes, which are frequently A/T-rich in the third position in *E. coli*, further correlate with decreased mRNA folding. Using rare codons rather than common codons at the 5' coding region increases protein expression in *E. coli* by approximately 14-fold (median fourfold) [72]. A recent study further confirmed that the first 18 nucleotides in the coding sequence strongly influence expression based on a study of the expression of 6348 genes from diverse phylogenetic sources. In this region, A and G increase and reduce the probability of high expression, respectively, whereas C and U have intermediate effects [74]. A model based on these experiments indicated that the influential mRNA-folding effects are restricted to the initial approximately 16 codons and that five genes designed by maximizing the folding energy (minimizing folding stability) in the 5' coding region showed uniformly high expression [74].

To decrease the propensity by the mRNA around the ribosome binding site to form secondary structures, optimization of the AT-content of N-terminal codons has been demonstrated to be a useful strategy, which was used to promote the over-expression of several proteins from bacteria [75], plants [76] and mammals [77] in *E. coli*. Moreover, computational tools have been developed to estimate protein expression and design optimal sequences, such as ExEnSo (Expression Enhancer Software) [78], RBS CALCULATOR [79], RBS DESIGNER [80], UTR DESIGNER [81] and EMOPEC [82]. All calculators were

designed for use with *E. coli* and have been shown to give good approximations of protein expression levels [83].

4.3. Fusion tags

A prerequisite for high-throughput purification is the addition of a fusion tag at the N- or C-terminus of recombinant proteins. An optimal fusion tag must fulfil these criteria: the tag must enable (i) easy detection of protein expression, (ii) high protein expression and solubility, and (iii) easy isolation of highly pure proteins from *E. coli*. The tags used in early studies were all large proteins, such as Protein A (280 amino acids (aa)) and LacZ (1024 aa) [84,85]. A wide range of tags have been developed [85–87], and the general features of the commonly used tags are listed in table 1. Because the strategies used for expressing cytoplasmic and membrane proteins in *E. coli* differ considerably, we discuss the tags used for these proteins individually below.

4.3.1. Fusion tags for cytoplasmic proteins

Fusion tags are invariably introduced at the N-terminus of cytoplasmic proteins, which can provide a reliable context for efficient translation initiation (figure 3*a* and table 1). [86]. The polyhistidine affinity tag, also known as the 6×His-tag, His6 tag and/or hexa-histidine tag, typically consists of six consecutive histidine residues that can bind to several types of immobilized ions (such as nickel, cobalt and copper) [88]. Recombinant galactose dehydrogenase fused with a His-tag was the first protein purified using immobilized metal affinity chromatography [94]. The His-tag is one of most ubiquitously used purification tags, and highly pure protein (more than 80%) can be obtained in a single chromatographic step from *E. coli* together with high expression. The FLAG tag (8 aa) [89] and Strep-II tag (8 aa) [90] are also small tags, but the purification costs may be higher compared with the His-tag. The benefit of adding small fusion tags with minimal charge is that the effects of the tags on recombinant protein structure, activity and characteristics are minimized; however, the recombinant proteins may readily form inclusion bodies [87].

Because the soluble expression and the expression certain non-expressed targets in *E. coli* represent a major bottleneck in protein production, studies continue to develop additional fusion tags for enhancing protein solubility and expression. Large fusion tags positively influence protein solubility and expression efficiency. Thioredoxin (Trx), small ubiquitin-like modifier (SUMO), glutathione S-transferase (GST), green fluorescent protein (GFP), HaloTag and maltose binding protein (MBP), which range in size from 100 to 495 aa, have been widely reported to increase protein expression and solubility [87,91,95–98]. However, the immunogenicity of the tags and their effect on the structure and function of recombinant proteins are major limitations compared with the use of small fusion tags. Another limitation of many of these fusion tags is that they do not function equally well with all target proteins [98]. Recently, an Fh8 tag system (Hitag) with small size (8 kDa) was reported as a robust fusion partner that enables both soluble protein production and the purification of several proteins rapidly and cost-effectively [99].

To overcome the problems associated with different tags, tandem affinity purification (TAP), which involves the use of two affinity tags attached to a target protein, is now commonly used in recombinant protein production. TAP offers an effective

Table 1. Main characteristics of commonly used fusion tags for high-throughput protein production.

tag	length/size (kDa)	matrix/elution	typical uses	comments	references
His-tag	2–10, typically 6 (0.84)	divalent metal ion (Ni, Co, Cu, Zn)/imidazole or low pH	purification and detection	most common purification tag; denaturing purification possible; rarely affects the structure or function of fusion proteins; an anti-His antibody can be used for detection	[88]
FLAG	8 (1.0)	FLAG antibody/low pH, EDTA or FLAG peptide	purification and detection	small size and high solubility; the presence of an internal enterokinase cleavage site; very expensive resins with limited re-use cycles	[89]
Strep-II	8 (1.1)	Strep-Tactin/biotin or desthiobiotin	purification and detection	short, biologically inert and proteolytically stable; does not interfere with membrane translocation or protein folding	[90]
Fh8	69 (8.0)	Ca ²⁺ -dependent hydrophobic interaction/ EDTA	purification, increased solubility and expression	relatively low molecular weight; with the combined features of enhancing protein solubility and purification	[86]
Trx	109 (11.7)	phenylarsine oxide/thiol containing reducing agents	purification and increased solubility	one of the best N-terminal protein fusions to promote soluble expression; purification must be conducted in absence of thiol containing reducing agents until elution step; large tag or elution conditions may affect properties of fusion protein	[91]
SUMO	100 (12.0)	an affinity tag must be added (typically His-tag)	increased solubility and expression	has all the advantages of Trx; SUMO protease efficiently cleaves the tag; enhances membrane proteins expression	[92]
GST	211 (26.0)	glutathione/reduced glutathione	purification, detection and increased expression and solubility	very common purification tag; one-step purification of relatively pure protein; denaturing purification impossible	[85]
GFP	238 (26.9)		detection, increased solubility and expression	native detection protein solubility and expression without antibody, particularly for membrane proteins	[93]
HaloTag	312 (34.0)	Chloroalkane/HaloTag buffer and TEV protease	purification, increased solubility and expression	allow for <i>in vivo</i> labelling; functions quickly and results in a highly pure, tag-free protein; cleavage of the tag may result in aggregation of proteins	[85]
MBP	396 (42.0)	cross-linked amylose/maltose	purification, detection, increased expression and solubility	can alleviate toxicity of fusion proteins; the target protein is prone to aggregation after removing tag; the large tag size may affect fusion protein properties and cause immunogenicity	[87]

and highly specific method for purifying target proteins. After two successive affinity chromatography purifications, the target protein is sufficiently pure for biochemical research. For example, the use of a tandem (His)₆-calmodulin fusion tag, which combines metal affinity chromatography and hydrophobic interaction chromatography, resulted in the production of eGFP and human p53 that were more than 97% pure after the (His)₆-calmodulin-tag was cleaved at a thrombin recognition site [100]. Because this technique has been widely exploited, various tags based on other types of TAP have been developed. However, the traditional His-tag, FLAG tag and Strep-II tag remain favourable candidates for use as TAP-tag components [101].

4.3.2. Fusion tags for membrane proteins

Investigation of the structure and function of membrane proteins is challenging because of the difficulties associated with purifying large amounts of these proteins. One difficulty is that membrane proteins must insert into the cytoplasmic membrane and fold properly. To obtain membrane proteins in the folded form, both fusion tags and *E. coli* strains must be designed to be optimal for the membrane protein production process. In membrane proteins, the first hydrophobic transmembrane segment provides the required signal for membrane targeting and insertion [102]; thus, fusion tags are routinely attached to the C-terminus rather than the N-terminus of a target membrane protein, and then the tags are used to monitor the localization, quantity, quality and purification of the membrane protein (figure 3b).

One commonly used approach is to fuse a membrane protein to GFP in order to track protein expression, partly because GFP becomes fluorescent only if the upstream target membrane protein integrates into the membrane (table 1) [93,103]. Moreover, GFP fluorescence can be used to rapidly, accurately and easily measure protein expression both in liquid cultures and standard SDS gels [104]. Furthermore, once protein expression has been optimized, the fluorescence from GFP can considerably accelerate detergent screening and purification [105]. However, GFP fusion proteins present certain notable disadvantages; for example, they generate false-positives and protein aggregation occurs after GFP cleavage. Thus, a fluorescent probe that interacts with small His-tag-fused membrane proteins was recently developed; using this probe, target proteins were detected sensitively to 0.02 mg l⁻¹ in crude lysates [106].

Whether a given recombinant membrane protein will become localized to the cell membrane or inclusion bodies cannot be predicted. Therefore, additional fusion partners have been developed to facilitate the targeting of membrane proteins to the lipid bilayer. The adenovirus-receptor immunoglobulin variable-type domains were successfully overexpressed as fusions with a set of short, non-globular, negatively charged peptides [107]. Mistic, a short and non-globular *B. subtilis* integral-membrane protein, has been used as a fusion tag for the high-level production of various membrane proteins in their native conformations, including several eukaryotic proteins that are toxic to *E. coli*. [108]. Leviatan *et al.* [109] reported that YaiN and YbeL, two short hydrophilic bacterial proteins, fused to the ends of membrane proteins may facilitate proper folding.

4.3.3. Detection of protein expression using fusion tags

Fusion tags can also be used in protein expression screening, which is essential for obtaining well-expressed and functional proteins. If a His-tag is attached to a target protein, an anti-

His antibody can be used to detect the expression and solubility of the recombinant protein in a 96-well format [110]. Proteins can also be labelled with GFP. Here, inclusion body formation leads to the misfolding of GFP and thus a loss of its fluorescence, but if the fusion protein is folded properly, GFP can be synthesized in a fluorescent form. Alternatively, a fluorescent amino acid derivative, BODIPY-FL-lysine, can be translationally incorporated into target proteins; these specifically labelled proteins in cell lysates can be detected using a fluorescence detector [96]. A previous study also reported the fusion of another coloured protein, photoactive yellow protein (or its miniaturized version), to a target protein. In this case, the addition of a precursor of the chromophore to the coexpressed photoactive yellow protein causes a yellow colour to appear; this colour development not only allows target protein expression to be monitored through visual inspection within a few seconds, but also enables protein concentration and purity to be quantified using a spectrometer within a few minutes [111].

4.3.4. Fusion tags and inclusion bodies

Inclusion body formation is a commonly encountered problem, and to promote the solubility of target proteins, high-molecular-weight N-terminal tags such as MBP and GST can be used [97,98]. The soluble expression of recalcitrant proteins can be also improved by designing variants with more favourable native-state energy. Up to five variants encoding from 9 to 67 mutations relative to wild-type can be designed by using the PROSS webserver. The tested variants show higher soluble expression and stability with no change in enzymatic function [112].

However, inclusion body formation does not mean that protein production has failed. The advantages of inclusion bodies are that they (i) produce proteins that are toxic to host cells, (ii) generally allow a high level of expression, and (iii) can be readily separated from bacterial cytoplasmic proteins through centrifugation. The most commonly used methods for refolding inclusion body proteins involve dialysis and on-column folding. Yuan *et al.* [113] reported the continuous-flow mode of a vortex fluid device that enabled parallel processing of protein refolding, and substantially shortened purification times, lowered costs and decreased structure waste streams associated with protein expression. High-throughput inclusion body purification can also be performed using a robotic micro-fuge: key mutants of RNA polymerase from *Sulfolobus shibatae* are predominantly expressed in an insoluble form, and hundreds of mutants can be automatically purified without the use of tags because inclusion bodies can be readily separated from soluble proteins through centrifugation [12].

4.3.5. Removal of fusion tags

Because many of the aforementioned tags are large polypeptides and may affect the structure and function of target proteins, tag removal is frequently necessary. In all expression vectors, a protease cleavage site is engineered between the tag and target protein. Several proteases can be selected to remove the tag, including SUMO protease, enteropeptidase, thrombin, factor Xa, PreScission and tobacco etch virus (TEV) protease. Among these, SUMO protease only cleaves SUMO tags [92], enteropeptidase and thrombin are incompatible with buffers containing reducing agents [114], factor Xa should not be

Table 2. Main characteristics of commonly used expression strains for high-throughput protein production.

strains	genotype	features	references
BL21(DE3)	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3)	the most common protein expression strain; leaky expression can lead to uninduced expression of potentially toxic proteins	[117]
BL21Star(DE3)	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm <i>rne131</i> (DE3)	mRNA levels and RNA stability are increased in the strain; thus, protein expression may be increased	[118]
Origami(DE3)	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm <i>trxB</i> <i>gor</i> (DE3)	the <i>trxB</i> and <i>gor</i> mutations enable cytoplasmic disulfide bond formation and can be combined with a fusion to Trx	[119]
BL21(DE3)pLysS	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3) [pLysS Cam ^r]	the pLysS plasmid produces T7 lysozyme to reduce basal level expression, which is suitable for expression of toxic genes	[120]
BL21-CodonPlus(DE3)-RIPL	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3) <i>endA</i> Hte [<i>argU</i> <i>proL</i> Cam ^r] [<i>argU</i> <i>ileY</i> <i>leuW</i> Strep/Spec ^r]	the CodonPlus strains provide additional copies of rare tRNA genes; the RIPL strain carries genes for Arg (AGA and AGG), Ile (AUA), Pro (CCC) and Leu (CUA)	[121]
Rosetta(DE3)	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3) [pRARE Cam ^r]	the Rosetta strains enhance the expression of proteins that contain codons rarely used in <i>E. coli</i> ; the Rosetta (DE3) strain carries genes for Arg (AGG, AGA and CGG), Ile (AUA), Leu (CUA), proline (CCC) and glycine (GGA)	[122]
C41(DE3)/C43(DE3)	selected mutants from BL21(DE3)	the strains harbour mutations in <i>lacUV5</i> promoter, which are effective for expressing toxic and membrane proteins	[56]
Lemo21(DE3)	F ⁻ OmpT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3) [pLemo Cam ^r]	the strain allows for tunable expression of difficult clones; for difficult soluble proteins, tuning the expression level may also result in more soluble, properly folded protein	[123]

used in the presence of chelating agents because it binds calcium ions [115], and PreScission leaves behind a Gly-Pro dipeptide on the N-terminus of the recombinant protein after digestion [116]. TEV protease is not inhibited by reducing agents, exhibits very high specificity, is inexpensive, and in most cases cleaves recombinant proteins in a manner that leaves the native protein intact [98,114]. Thus, TEV protease shows the greatest number of advantages as an endoprotease for removing affinity tags for high-throughput purposes.

5. *Escherichia coli* expression strains and cell culture

The choice of the strains used to express recombinant proteins also plays a major role in protein expression, solubility and yield. A few *E. coli* strains such as BL21 and its derivatives are widely used (figure 4). Different *E. coli* strains facilitate the expression of proteins containing disulfide bonds or those that are encoded by genes containing rare codons and proteins toxic to *E. coli*. Moreover, coexpression with some genes improves the expression of post-translationally modified proteins. To date, several *E. coli* strains that strongly improve membrane protein production have been engineered. The genotypes and characteristics of these strains are summarized in table 2.

5.1. Routine *Escherichia coli* strains

BL21 and its derivatives are routinely used for recombinant protein production in *E. coli* (figure 4a and table 2). These strains are deficient in the proteases Lon and OmpT, which can

increase protein stability. The strain BL21(DE3) contains a chromosomal copy of the T7 RNA polymerase gene for simple and efficient expression of genes under control of the T7 promoter [117]. BL21Star(DE3) contains a mutation in *rne*, the gene that encodes RNase E, and thus the use of BL21Star(DE3) increases mRNA stability and protein expression [118,124]. BL21trxB, a derivative of BL21(DE3), harbours a thioredoxin reductase (*trxB*) mutation, and the strain Origami(DE3) contains mutations in both *trxB* and the gene encoding glutathione reductase (*gor*), which markedly enhances disulfide bond formation in the cytoplasm [119]. BL21(DE3)pLysS contains a pLysS plasmid carrying the gene encoding T7 lysozyme; this strain is used to express proteins that are toxic to cells because T7 lysozyme lowers the leaky expression of target genes [120]. BL21-CodonPlus(DE3) strains provide additional copies of rare tRNA genes; for example, BL21-CodonPlus(DE3)-RIPL (contains the largest number of tRNA genes in the BL21-CodonPlus series) carries genes for Arg-, Ile-, Leu- and Pro- tRNAs [121]. The strains Rosetta and Rosetta (DE3) harbour the pRARE plasmid, in which the genes encoding aminoacyl-tRNA synthetases for Arg, Ile, Leu, Pro and Gly are coexpressed [122]. Both the BL21-CodonPlus(DE3) and Rosetta (DE3) strains efficiently promote the expression of genes harbouring rare codons at high frequencies.

5.2. Strategies for expressing proteins with post-translational modifications

The major limitation of using *E. coli* for protein expression is thought to be its lack of available machinery for

post-translational modifications. Coexpression of factors that promote post-translational modification appears to be a promising approach for solving this problem (figure 4b) [125]. Reversible protein phosphorylation is one of the most important and well-studied post-translational modifications. In *E. coli*, phosphorylation of a target molecule (a mouse or human protein) has been achieved by coexpression with human Jun N-terminal kinase 1 [126]. Protein glycosylation is another major post-translational modification that substantially affects protein stability, distribution and function. The discovery of *N*-linked protein glycosylation in *Campylobacter jejuni* and the functional transfer of this glycosylation system into *E. coli* enabled the production of recombinant glycoproteins in bacteria, although bacterial *N*-glycans structurally differ from their eukaryotic counterparts [127]. Glycoconjugated vaccines can be produced in *E. coli* using this strategy [128]. Furthermore, bacterial *N*-linked glycosylation occurs on scFv antibody fragments and improves the biophysical properties [129]. Ubiquitin is an 8 kDa polypeptide (76 aa) that can be appended to a lysine in target proteins. In *E. coli*, recombinant proteins can be ubiquitinated by co-overexpressing the target protein, ubiquitin and ubiquitin ligases [130]. Additionally, methylation, myristoylation and acetylation have been successfully performed in *E. coli* by coexpressing a methyltransferase, myristoyltransferase and acetylase, respectively [131–133]. Therefore, target proteins can be post-translationally modified in *E. coli* expression systems by coexpressing genes related to the modifications of interest.

5.3. *Escherichia coli* strains for expression of membrane proteins

Several recombinant membrane proteins exhibit toxicity upon induction in *E. coli*, and thus only low yields of the properly folded forms of these proteins are obtained [134]. Understanding the physiological response of *E. coli* to recombinant membrane proteins is crucial for identifying bottlenecks in expression and folding [135]. Most of the targeting and translocation of membrane proteins occur through a universally conserved signal-recognition particle (SRP)/secretory (Sec) pathway [136]. Ribosome nascent chain-SRP complexes contact the SRP receptor FtsY at the membrane and thus mediate the transfer of the nascent chain to the Sec translocon. Transfer of the complex into the Sec pore is driven by SecA and ATP hydrolysis. The SecDFYajC complex also plays a critical role in the biogenesis, translocation and folding of membrane proteins [137].

Saturation of the translocon pathway during membrane protein overexpression may cause the accumulation of cytoplasmic aggregates and broad perturbations in the proteome [134]. Two strategies for solving this problem have been employed: (i) tuning of transcription and translation rates and (ii) coexpression of biogenesis factors (figure 4c). The strains C41(DE3) and C43(DE3), which are also known as the Walker strains, are BL21(DE3) derivatives harbouring mutations in the *lacUV5* promoter, influencing the expression levels of T7 RNA polymerase (table 2). A mutation in the *lac* repressor LacI was also demonstrated to be crucial for favouring tolerance to membrane protein overexpression [138]. Subsequent production of comparatively lower amounts of target proteins in the Walker strains ensured that the Sec translocon was not saturated by the produced proteins [61,139].

Lemo21(DE3) is tunable for membrane protein overexpression, and the amount of membrane protein produced can be readily regulated by exploiting the Sec-translocon capacity of *E. coli* [123]. In Lemo21(DE3), the activity of T7 RNA polymerase can be precisely regulated by expressing T7 lysozyme under control of the *L*-rhamnose promoter and then modulating the target protein level by adding 0–2 mM *L*-rhamnose to the culture (table 2) [123].

A complementary approach to lowering protein expression involves increasing the amount of protein biogenesis machinery. Coexpression of the cytoplasmic DnaK/J chaperone system, which functions in protein targeting and folding, improved the production of the magnesium transporter CorA [140]. Moreover, coproduction of the protease FtsH, a membrane-bound quality-control factor, markedly enhanced the yields of G-protein coupled receptors [141]. However, most efforts employing this strategy have not been successful. For example, coexpression of membrane protein biogenesis factors (SRP/FtsY, SecA) and other factors with CorA or G-protein coupled receptors did not improve target protein production [140,141].

Previous studies have also used strategies involving either increasing the expression of factors that enhance membrane protein yields or deleting factors that limit protein production [142,143]. Our understanding of how membrane proteins are translocated and folded in *E. coli* is highly limited, and it appears that the optimal strain for membrane protein production is protein-specific [144]. Currently, C41(DE3), C43(DE3) and Lemo21(DE3) remain the first-choice strains for membrane protein expression.

5.4. Culture of *Escherichia coli*

Both culture media composition and culture conditions are important for protein expression. Luria broth (LB) medium is easy to make and is the most commonly used medium for culturing *E. coli*. However, *E. coli* growth in LB stops at a relatively low density, because it contains low amounts of carbohydrates and divalent cations [145]. The 2× yeast extract tryptone, terrific broth and super broth media can also be used and have been shown to be superior to LB for reaching higher cell densities [146]. As cell density increases, oxygen may limit *E. coli* growth and protein expression in batch culture [147]; additional agitation can be generated by using high shaking speeds, shaking in a baffled flask and oxygen-enriched air or pure oxygen [148]. It is also possible to avoid the formation of inclusion bodies by optimizing cell culture conditions. Protein expression in *E. coli* at 15–25°C is commonly induced to increase the solubility of recombinant proteins, and the induction temperature can be lowered to 6–10°C [149]. Uncontrolled pH culture conditions favour recombinant protein aggregation, but stable pH can be maintained by using buffers or through the automatic addition of base or acid [150]. The addition of the cofactors or binding partners required for protein folding to the cultivation media will enhance protein solubility and prevent inclusion body formation [151,152]. Alternatively, the addition of a mild detergent such as Triton X-100 in shaker flasks can enhance the solubility and secretion ratio of aggregation-prone protein [153]. In conclusion, media composition and culture conditions are critical factors for optimizing the expression of recombinant proteins. Although this is attained mostly by trial and error, it may be beneficial.

In contrast to the IPTG induction method, autoinduction was introduced as a convenient method for producing recombinant proteins without inducer addition at the small laboratory scale for *lac* operon-controlled expression systems [146]. Autoinduction medium contains glycerol, lactose and glucose at optimized levels, with glycerol used as the carbon source. Lactose is metabolized for autoinduction once glucose is depleted [154]. Thus, there is no need to monitor the growth, minimizing operator intervention from inoculation to cell harvest, which is preferable in high-throughput experiments. Additionally, there is tighter control of protein induction, improving the expression of toxic proteins. Another advantage of autoinduction is that the medium allows cultures to reach high cell densities and generally produces a greater proportion of soluble target proteins than IPTG-induced expression [155,156]. A disadvantage of autoinduction is that the medium is adversely affected by the aeration level. This can be overcome by using a glucose fed-batch medium, which attenuates oxygen-sensitivity and provides robust high-yield expression under high aeration rates [157]. In some cases, the use of autoinduction medium may not be optimal and is often replaced by other media and induction with IPTG to obtain better yields [158].

The simplest way to grow *E. coli* is batch cultivation, but control of the growth during this process is limited. High-throughput cultivation has undergone rapid evolution in recent years in reducing culture volume, applying in-process real-time monitoring or control at the micro scale, and realizing full automation of the systems [159,160]. A number of emerging cultivation platforms has been commercialized, including microtitre plate culture, micro scale bioreactors and in-parallel fermentation systems [160]. These platforms that significantly reduce culture volume have been adopted extensively to replace shaker flasks [161]. High-throughput cultivation technology, which enables researchers to handle a large number of samples under a range of fermentation conditions in a high-throughput format, can remarkably shorten the timeline from DNA to large-scale protein production [160].

6. High-throughput robotic platform for protein expression and purification

High-throughput platforms that can rapidly clone genes, pick colonies, isolate plasmid DNA, transform bacteria, and express and purify proteins have provided opportunities for executing complex molecular biological procedures with little human labour and minimal error rates. Several commercial robotic workstations are available for various purposes, including Equator GX8 Dispenser from Labcyte (Sunnyvale, CA, USA), MicroSys from Genomic Solutions (Ann Arbor, MI, USA), sciFLEXARRAYER dispenser from Scienion (Berlin, Germany) and other systems [162]. These platforms have been used to isolate plasmid DNA, transform bacteria, pick colonies and screen for protein expression [162,163], and a video showing the operation procedure for automatic protein purification is available [164]. Automatic platforms can cost hundreds of thousands of dollars and require routine maintenance, and organizations commonly hire specialists to care for these automated platforms. Thus, if a protein production process does not include adequate numbers of

samples to justify this level of spending, it may be prudent to continue to use a manual approach in parallel [165].

7. Conclusion and perspectives

Successful recombinant protein expression and purification is frequently indispensable for both basic research studies and biotechnological and commercial applications [166]. High-throughput protein expression and purification in *E. coli* has begun to revolutionize the manner in which studies are conducted in various research fields. Experiments that were typically performed manually to address one protein at a time over a period of several weeks can now be conducted for hundreds of proteins in as little as one week. However, limitations still exist and further improvements are possible.

In terms of obtaining target genes, *in silico* design followed by array-based *de novo* synthesis rather than PCR may become widely used in the future. The major challenges associated with *de novo* synthesis are sequence errors, availability and cost. However, if array-based gene synthesis can be commercialized, the costs could decrease by 3–5 orders of magnitude to 10^3 – 10^5 bp per dollar [25].

Cloning methods have seen rapid advances, and cloning systems used in both commercial and academic settings can be operated with high efficiency, fidelity and reliability, and at a reasonably low cost. The first requirement is to develop a highly flexible expression vector that is fully compatible with high-throughput procedures. An optimal vector must contain a strong but tunable promoter and tags with optimized N-terminal codons to facilitate protein expression, solubility and purification. Large N-terminal tags have been used to enhance translational initiation and promote solubility. However, the cleavage of these large tags may complicate the experiment being conducted and substantially add to the final cost compared to the use of short tags. Given that the downstream costs of testing the functions of individual proteins are often far higher than protein production costs, the cost will probably not dramatically affect experimental workflows. Moreover, new tags are being developed, but considerable room for improvement remains.

Currently, certain post-translational modifications can be achieved in *E. coli* by coexpressing the corresponding enzymes. However, such coexpression invariably affects the growth rate of *E. coli*, and several vectors cannot be readily coexpressed in a single strain. One solution is to integrate genes encoding post-translational modification factors into the genome to create 'eukaryotic-like' *E. coli*. Moreover, according to previous studies, tuning or precisely controlling the transcript levels of target proteins is critical for expressing membrane proteins. Membrane protein production is not always successful when the strategy involves coexpressing proteins that function in membrane protein biogenesis. Thus, it is crucial to understand the protein biogenesis mechanism and the physiological response of *E. coli* to membrane protein production. The combination of physiological, genetic and 'omics' technologies has improved the understanding of the biogenesis process and has provided rationale for the forward engineering of expression hosts.

Finally, robotic platforms for protein expression and purification are available but are too expensive for most laboratories. However, the protocols and systems currently in use provide an approach required for the cloning,

expression and purification of hundreds of proteins in parallel within a few days. The limitations of the protein production process are nearly impossible to solve in a simple and global manner, cases of failure are rarely reported and experience gained does not effectively help guide subsequent efforts. Therefore, a searchable protein expression database that includes strains, vectors, tags, promoters, and cases of success and failure to guide the journey from trial

and error towards rational design would be more beneficial to the scientific community than a robotic platform.

Competing interests. We declare we have no competing interests.

Funding. Our research was supported by the Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ00999302), RDA, Republic of Korea.

Acknowledgements. We would like to thank the reviewers for their valuable and insightful comments, which helped to improve the paper.

References

- Hsiao A, Kuo MD. 2006 High-throughput biology in the postgenomic era. *J. Vasc. Interv. Radiol.* **17**, 1077–1085. (doi:10.1097/01.rvi.0000228840.12260.ff)
- Cameron DE, Bashor CJ, Collins JJ. 2014 A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390. (doi:10.1038/nrmicro3239)
- Pepperkok R, Ellenberg J. 2006 High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.* **7**, 690–696. (doi:10.1038/nrm1979)
- Khow O, Suntrarachun S. 2012 Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac. J. Trop. Biomed.* **2**, 159–162. (doi:10.1016/S2221-1691(11)60213-X)
- Rosano GL, Ceccarelli EA. 2014 Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, 172. (doi:10.3389/fmicb.2014.00172)
- Lesley SA. 2001 High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expr. Purif.* **22**, 159–164. (doi:10.1006/prep.2001.1465)
- Stevens RC. 2000 Design of high-throughput methods of protein production for structural biology. *Structure* **8**, R177–R185. (doi:10.1016/S0969-2126(00)00193-3)
- Vincentelli R, Cimino A, Geerlof A, Kubo A, Satou Y, Cambillau C. 2011 High-throughput protein expression screening and purification in *Escherichia coli*. *Methods* **55**, 65–72. (doi:10.1016/j.ymeth.2011.08.010)
- Bird LE. 2011 High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods* **55**, 29–37. (doi:10.1016/j.ymeth.2011.08.002)
- Mlynek G, Lehner A, Neuhold J, Leeb S, Kostan J, Charnagalov A, Stolt-Bergner P, Djinić-Carugo K, Pinotsis N. 2014 The center for optimized structural studies (COSS) platform for automation in cloning, expression, and purification of single proteins and protein–protein complexes. *Amino acids* **46**, 1565–1582. (doi:10.1007/s00726-014-1699-x)
- Dieckman L, Gu M, Stols L, Donnelly MI, Collart FR. 2002 High throughput methods for gene cloning and expression. *Protein Expr. Purif.* **25**, 1–7. (doi:10.1006/prep.2001.1602)
- Weinzierl ROJ. 2013 The RNA polymerase factory and archaeal transcription. *Chem. Rev.* **113**, 8350–8376. (doi:10.1021/cr400148k)
- Jia B, Li Z, Liu J, Sun Y, Jia X, Xuan YH, Zhang J, Jeon CO. 2015 A zinc-dependent protease AMZ-tk from a thermophilic archaeon is a new member of the archaeometzincin protein family. *Front. Microbiol.* **6**, 1380. (doi:10.3389/fmicb.2015.01380)
- Jia B, Liu J, Van Duyet L, Sun Y, Xuan YH, Cheong GW. 2015 Proteome profiling of heat, oxidative, and salt stress responses in *Thermococcus kodakarensis* KOD1. *Front. Microbiol.* **6**, 605. (doi:10.3389/fmicb.2015.00605)
- Jia B, Cheong GW, Zhang S. 2013 Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles* **17**, 193–203. (doi:10.1007/s00792-012-0509-1)
- Holz C, Lueking A, Bovekamp L, Gutjahr C, Bolotina N, Lehrach H, Cahill DJ. 2001 A human cDNA expression library in yeast enriched for open reading frames. *Genome Res.* **11**, 1730–1735. (doi:10.1101/gr.181501)
- Büssow K, Nordhoff E, Lubbert C, Lehrach H, Walter G. 2000 A human cDNA library for high-throughput protein expression screening. *Genomics* **65**, 1–8. (doi:10.1006/geno.2000.6141)
- Marsischky G, LaBaer J. 2004 Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res.* **14**, 2020–2028. (doi:10.1101/gr.2528804)
- Cao Y, Sun J, Zhu J, Li L, Liu G. 2010 PrimerCE: designing primers for cloning and gene expression. *Mol. Biotechnol.* **46**, 113–117. (doi:10.1007/s12033-010-9276-3)
- Camilo CM, Lima GM, Maluf FV, Guido RV, Polikarpov I. 2015 HTP-OligoDesigner: an online primer design tool for high-throughput gene cloning and site-directed mutagenesis. *J. Comput. Biol.* **23**, 27–29. (doi:10.1089/cmb.2015.0148)
- Otto P, Larson B, Krueger S. 2002 Automated high-throughput purification of PCR products using Wizard® MagneSil™ paramagnetic particles. *J. Lab. Autom.* **7**, 120–124. (doi:10.1016/s1535-5535-04-00232-1)
- Xiong AS, Yao QH, Peng RH, Duan H, Li X, Fan HQ, Cheng ZM, Li Y. 2006 PCR-based accurate synthesis of long DNA sequences. *Nat. Protoc.* **1**, 791–797. (doi:10.1038/nprot.2006.103)
- Yehezkel TB, Linshiz G, Buaron H, Kaplan S, Shabi U, Shapiro E. 2008 *De novo* DNA synthesis using single molecule PCR. *Nucleic Acids Res.* **36**, e107. (doi:10.1093/nar/gkn457)
- Klein JC, Lajoie MJ, Schwartz JJ, Strauch EM, Nelson J, Baker D, Shendure J. 2016 Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res.* **44**, e43. (doi:10.1093/nar/gkv1177)
- Kosuri S, Church GM. 2014 Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507. (doi:10.1038/nmeth.2918)
- Quan J, Saaem I, Tang N, Ma S, Negre N, Gong H, White KP, Tian J. 2011 Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.* **29**, 449–452. (doi:10.1038/nbt.1847)
- Presnyak V *et al.* 2015 Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124. (doi:10.1016/j.cell.2015.02.029)
- Festa F, Steel J, Bian X, Labaer J. 2013 High-throughput cloning and expression library creation for functional proteomics. *Proteomics* **13**, 1381–1399. (doi:10.1002/pmic.201200456)
- Koehn J, Hunt I. 2009 High-throughput protein production (HTPP): a review of enabling technologies to expedite protein production. *Methods Mol. Biol.* **498**, 1–18. (doi:10.1007/978-1-59745-196-3_1)
- Blommel PG, Martin PA, Wrobel RL, Steffen E, Fox BG. 2006 High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system. *Protein Expr. Purif.* **47**, 562–570. (doi:10.1016/j.pep.2005.11.007)
- Nagase T *et al.* 2008 Exploration of human ORFeome: high-throughput preparation of ORF clones and efficient characterization of their protein products. *DNA Res.* **15**, 137–149. (doi:10.1093/dnares/dsn004)
- Engler C, Kandzia R, Marillonnet S. 2008 A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647. (doi:10.1371/journal.pone.0003647)
- Whitman L, Gore M, Ness J, Theodorou E, Gustafsson C, Minshull J. 2013 Rapid, scarless cloning of gene fragments using the electra vector system. *Genet. Eng. Biotechnol.* **33**, 42. (doi:10.1089/gen.33.11.20)
- Chen WH, Qin ZJ, Wang J, Zhao GP. 2013 The MASTER(methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly. *Nucleic Acids Res.* **41**, e93. (doi:10.1093/nar/gkt122)

35. Chao R, Yuan Y, Zhao H. 2015 Recent advances in DNA assembly technologies. *FEMS Yeast Res.* **15**, 1–9. (doi:10.1111/1567-1364.12171)
36. Esposito D, Garvey LA, Chakiath CS. 2009 Gateway cloning for protein expression. *Methods Mol. Biol.* **498**, 31–54. (doi:10.1007/978-1-59745-196-3_3)
37. Cheo DL, Titus SA, Byrd DRN, Hartley JL, Temple GF, Brasch MA. 2004 Concerted assembly and cloning of multiple DNA segments using *in vitro* site-specific recombination: functional analysis of multi-segment expression clones. *Genome Res.* **14**, 2111–2120. (doi:10.1101/gr.2512204)
38. Zhang Y, Werling U, Edelmann W. 2012 SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55. (doi:10.1093/nar/gkr1288)
39. Motohashi K. 2015 A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis. *BMC Biotechnol.* **15**, 47. (doi:10.1186/s12896-015-0162-8)
40. Okegawa Y, Motohashi K. 2015 Evaluation of seamless ligation cloning extract preparation methods from an *Escherichia coli* laboratory strain. *Anal. Biochem.* **486**, 51–53. (doi:10.1016/j.ab.2015.06.031)
41. Okegawa Y, Motohashi K. 2015 A simple and ultra-low cost homemade seamless ligation cloning extract (SLiCE) as an alternative to a commercially available seamless DNA cloning kit. *Biochem. Biophys. Rep.* **4**, 148–151. (doi:10.1016/j.bbrep.2015.09.005)
42. Aslanidis C, de Jong PJ. 1990 Ligation-independent cloning of PCR products(LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074. (doi:10.1093/nar/18.20.6069)
43. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CAIII, Smith HO. 2009 Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345. (doi:10.1038/nmeth.1318)
44. Irwin CR, Farmer A, Willer DO, Evans DH. 2012 In-fusion® cloning with vaccinia virus DNA polymerase. *Methods Mol. Biol.* **890**, 23–35. (doi:10.1007/978-1-61779-876-4_2)
45. Klock HE, Koesema EJ, Knuth MW, Lesley SA. 2008 Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71**, 982–994. (doi:10.1002/prot.21786)
46. Li MZ, Elledge SJ. 2007 Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256. (doi:10.1038/nmeth1010)
47. Bryksin AV, Matsumura I. 2010 Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* **48**, 463–465. (doi:10.2144/000113418)
48. Stevenson J, Krycer JR, Phan L, Brown AJ. 2013 A practical comparison of ligation-independent cloning techniques. *PLoS ONE* **8**, e83888. (doi:10.1371/journal.pone.0083888)
49. Thomas S, Maynard ND, Gill J. 2015 DNA library construction using Gibson Assembly®. *Nat. Methods* **12**, 11. (doi:10.1038/nmeth.f.384)
50. Camilo CM, Polikarpov I. 2014 High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning(LIC). *Protein Expr. Purif.* **99**, 35–42. (doi:10.1016/j.pep.2014.03.008)
51. Schmid-Burgk JL, Schmidt T, Kaiser V, Honing K, Hornung V. 2013 A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.* **31**, 76–81. (doi:10.1038/nbt.2460)
52. Yuan H, Peng L, Han Z, Xie JJ, Liu XP. 2015 Recombinant expression library of *Pyrococcus furiosus* constructed by high-throughput cloning: a useful tool for functional and structural genomics. *Front. Microbiol.* **6**, 943. (doi:10.3389/fmicb.2015.00943)
53. Lee N, Franchlyn C, Hamilton EP. 1987 Arabinose-induced binding of AraC protein to ara2 activates the araBAD operon promoter. *Proc. Natl Acad. Sci. USA* **84**, 8814–8818. (doi:10.1073/pnas.84.24.8814)
54. Guzman LM, Belin D, Carson MJ, Beckwith J. 1995 Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130.
55. Otto CM, Niagro F, Su X, Rawlings CA. 1995 Expression of recombinant feline tumor necrosis factor is toxic to *Escherichia coli*. *Clin. Diagn. Lab. Immunol.* **2**, 740–746.
56. Francis DM, Page R. 2010 Strategies to optimize protein expression in *E. coli*. *Curr. Protoc. Protein Sci.* **5**, 21–29. (doi:10.1002/0471140864.ps0524s61)
57. Graslund S *et al.* 2008 Protein production and purification. *Nat. Methods* **5**, 135–146. (doi:10.1038/nmeth.f.202)
58. Studier FW, Moffatt BA. 1986 Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130. (doi:10.1016/0022-2836(86)90385-2)
59. Pan SH, Malcolm BA. 2000 Reduced background expression and improved plasmid stability with pET vectors in BL21(DE3). *BioTechniques* **29**, 1234–1238.
60. Stano NM, Patel SS. 2004 T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *J. Biol. Chem.* **279**, 16 136–16 143. (doi:10.1074/jbc.M400139200)
61. Miroux B, Walker JE. 1996 Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298. (doi:10.1006/jmbi.1996.0399)
62. Wagner S *et al.* 2008 Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl Acad. Sci. USA* **105**, 14 371–14 376. (doi:10.1073/pnas.0804090105)
63. Kozak M. 1983 Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**, 1–45.
64. Cheong DE, Ko KC, Han Y, Jeon HG, Sung BH, Kim GJ, Choi JH, Song JJ. 2015 Enhancing functional expression of heterologous proteins through random substitution of genetic codes in the 5' coding region. *Biotechnol. Bioeng.* **112**, 822–826. (doi:10.1002/bit.25478)
65. Mutalik VK *et al.* 2013 Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360. (doi:10.1038/nmeth.2404)
66. Liebeton K, Lengefeld J, Eck J. 2014 The nucleotide composition of the spacer sequence influences the expression yield of heterologously expressed genes in *Bacillus subtilis*. *J. Biotechnol.* **191**, 214–220. (doi:10.1016/j.jbiotec.2014.06.027)
67. Egbert RG, Klavins E. 2012 Fine-tuning gene networks using simple sequence repeats. *Proc. Natl Acad. Sci. USA* **109**, 16 817–16 822. (doi:10.1073/pnas.1205693109)
68. Mirzadeh K, Martinez V, Toddo S, Guntur S, Hergard MJ, Elofsson A, Norholm MH, Daley DO. 2015 Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector-coding sequence junction. *ACS Synth. Biol.* **4**, 959–965. (doi:10.1021/acssynbio.5b00033)
69. Kozak M. 2005 Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37. (doi:10.1016/j.gene.2005.06.037)
70. Goltermann L, Borch Jensen M, Bentin T. 2011 Tuning protein expression using synonymous codon libraries targeted to the 5' mRNA coding region. *Protein Eng. Des. Sel.* **24**, 123–129. (doi:10.1093/protein/gzq086)
71. Bentele K, Saffert P, Rauscher R, Ignatova Z, Bluthgen N. 2013 Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* **9**, 675. (doi:10.1038/msb.2013.32)
72. Goodman DB, Church GM, Kosuri S. 2013 Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475–479. (doi:10.1126/science.1241934)
73. Tuller T *et al.* 2010 An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354. (doi:10.1016/j.cell.2010.03.031)
74. Boël GB *et al.* 2016 Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **529**, 358–363. (doi:10.1038/nature16509)
75. Castillo-Mendez MA, Jacinto-Loeza E, Olivares-Trejo JJ, Guarneros-Pena G, Hernandez-Sanchez J. 2012 Adenine-containing codons enhance protein synthesis by promoting mRNA binding to ribosomal 30S subunits provided that specific tRNAs are not exhausted. *Biochimie* **94**, 662–672. (doi:10.1016/j.biochi.2011.09.019)
76. Motohashi K, Okegawa Y. 2014 Method for enhancement of plant redox-related protein expression and its application for *in vitro* reduction of chloroplastic thioredoxins. *Protein Expr. Purif.* **101**, 152–156. (doi:10.1016/j.pep.2014.07.001)
77. Krishna Rao, DV, Rao JV, Narasu ML, Bhujanga Rao AK. 2008 Optimization of the AT-content of codons immediately downstream of the initiation codon

- and evaluation of culture conditions for high-level expression of recombinant human G-CSF in *Escherichia coli*. *Mol. Biotechnol.* **38**, 221–232. (doi:10.1007/s12033-007-9018-3)
78. Care S, Bignon C, Pelissier MC, Blanc E, Canard B, Coutard B. 2008 The translation of recombinant proteins in *E. coli* can be improved by *in silico* generating and screening random libraries of a $-70/+96$ mRNA region with respect to the translation initiation codon. *Nucleic Acids Res.* **36**, e6. (doi:10.1093/nar/gkm1097)
 79. Salis HM, Mirsky EA, Voigt CA. 2009 Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950. (doi:10.1038/nbt.1568)
 80. Na D, Lee D. 2010 RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* **26**, 2633–2634. (doi:10.1093/bioinformatics/btq458)
 81. Seo SW, Yang JS, Kim I, Yang J, Min BE, Kim S, Jung GY. 2013 Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* **15**, 67–74. (doi:10.1016/j.ymben.2012.10.006)
 82. Bonde MT, Pedersen M, Klausen MS, Jensen SI, Wulff T, Harrison S, Nielsen AT, Herrgard MJ, Sommer MOA. 2016 Predictable tuning of protein expression in bacteria. *Nat. Methods* **13**, 233–236. (doi:10.1038/nmeth.3727)
 83. Reeve B, Hargest T, Gilbert C, Ellis T. 2014 Predicting translation initiation rates for designing synthetic biology. *Front. Bioeng. Biotechnol.* **2**, 1. (doi:10.3389/fbioe.2014.00001)
 84. Sambrook JF, Maniatis E. 1989 *Molecular cloning: a laboratory manual*, 2nd edn. New York, NY: Cold Spring Laboratory Press.
 85. Kimple ME, Sondek J. 2004 Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* **9**, 9. (doi:10.1002/0471140864.ps0909s36)
 86. Costa S, Almeida A, Castro A, Domingues L. 2014 Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front. Microbiol.* **5**, 63. (doi:10.3389/fmicb.2014.00063)
 87. Zhao X, Li G, Liang S. 2013 Several affinity tags commonly used in chromatographic purification. *J. Anal. Methods Chem.* **2013**, 581093. (doi:10.1155/2013/581093)
 88. Gaberc-Porekar V, Menart V. 2001 Perspectives of immobilized-metal affinity chromatography. *J. Biochem. Biophys. Methods* **49**, 335–360. (doi:10.1016/S0165-022X(01)00207-X)
 89. Schmidt PM, Sparrow LG, Attwood RM, Xiao X, Adams TE, McKimm-Breschkin JL. 2012 Taking down the FLAG! How insect cell expression challenges an established tag-system. *PLoS ONE* **7**, e37779. (doi:10.1371/journal.pone.0037779)
 90. Schmidt TGM, Skerra A. 2007 The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2**, 1528–1535. (doi:10.1038/nprot.2007.209)
 91. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. 2004 Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32. (doi:10.1186/1472-6750-4-32)
 92. Zuo X, Li S, Hall J, Mattern MR, Tran H, Shoo J, Tan R, Weiss SR, Butt TR. 2005 Enhanced expression and purification of membrane proteins by SUMO fusion in *Escherichia coli*. *J. Struct. Funct. Genomics* **6**, 103–111. (doi:10.1007/s10969-005-2664-4)
 93. Hammon J, Palanivelu DV, Chen J, Patel C, Minor DL. 2009 A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci.* **18**, 121–133. (doi:10.1002/pro.18)
 94. Lilius G, Persson M, Bulow L, Mosbach K. 1991 Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. *Eur. J. Biochem.* **198**, 499–504. (doi:10.1111/j.1432-1033.1991.tb16042.x)
 95. Wu X, Oppermann U. 2003 High-level expression and rapid purification of rare-codon genes from hyperthermophilic archaea by the GST gene fusion system. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **786**, 177–185. (doi:10.1016/S1570-0232432#(02)00810-3)
 96. Coleman MA, Lao VH, Segelke BW, Beernink PT. 2004 High-throughput, fluorescence-based screening for soluble protein expression. *J. Proteome Res.* **3**, 1024–1032. (doi:10.1021/pr049912g)
 97. Hewitt SN, Choi R, Kelley A, Crowther GJ, Napuli AJ, Van Voorhis WC. 2011 Expression of proteins in *Escherichia coli* as fusions with maltose-binding protein to rescue non-expressed targets in a high-throughput protein-expression and purification pipeline. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 1006–1009. (doi:10.1107/s1744309111022159)
 98. Esposito D, Chatterjee DK. 2006 Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353–358. (doi:10.1016/j.copbio.2006.06.003)
 99. Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L. 2013 The Fh8 tag: a fusion partner for simple and cost-effective protein purification in *Escherichia coli*. *Protein Expr. Purif.* **92**, 163–170. (doi:10.1016/j.pep.2013.09.013)
 100. McCluskey AJ, Poon GM, Garipey J. 2007 A rapid and universal tandem-purification strategy for recombinant proteins. *Protein Sci.* **16**, 2726–2732. (doi:10.1110/ps.072894407)
 101. Li Y. 2010 Commonly used tag combinations for tandem affinity purification. *Biotechnol. Appl. Biochem.* **55**, 73–83. (doi:10.1042/ba20090273)
 102. van Geest M, Lolkema JS. 2000 Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol. Mol. Biol. Rev.* **64**, 13–33. (doi:10.1128/mmbr.64.1.13-33.2000)
 103. Drew D *et al.* 2005 A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* **14**, 2011–2017. (doi:10.1110/ps.051466205)
 104. Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW. 2006 Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3**, 303–313. (doi:10.1038/nmeth0406-303)
 105. Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S. 2008 GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 784–798. (doi:10.1038/nprot.2008.44)
 106. Backmark AE, Olivier N, Snijder A, Gordon E, Dekker N, Ferguson AD. 2013 Fluorescent probe for high-throughput screening of membrane protein expression. *Protein Sci.* **22**, 1124–1132. (doi:10.1002/pro.2297)
 107. Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P. 2004 Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr. Purif.* **36**, 207–216. (doi:10.1016/j.pep.2004.04.020)
 108. Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. 2005 NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **307**, 1317–1321. (doi:10.1126/science.1106392)
 109. Leviatan S, Sawada K, Moriyama Y, Nelson N. 2010 Combinatorial method for overexpression of membrane proteins in *Escherichia coli*. *J. Biol. Chem.* **285**, 23 548–23 556. (doi:10.1074/jbc.M110.125492)
 110. Knaust RK, Nordlund P. 2001 Screening for soluble expression of recombinant proteins in a 96-well format. *Anal. Biochem.* **297**, 79–85. (doi:10.1006/abio.2001.5331)
 111. Kim Y, Ganesan P, Ihee H. 2013 High-throughput instant quantification of protein expression and purity based on photoactive yellow protein turn off/on label. *Protein Sci.* **22**, 1109–1117. (doi:10.1002/pro.2286)
 112. Goldenzweig A *et al.* 2016 Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* **63**, 337–346. (doi:10.1016/j.molcel.2016.06.012)
 113. Yuan TZ *et al.* 2015 Shear-stress-mediated refolding of proteins from aggregates and inclusion bodies. *ChemBiochem* **16**, 393–396. (doi:10.1002/cbic.201402427)
 114. Waugh DS. 2011 An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **80**, 283–293. (doi:10.1016/j.pep.2011.08.005)
 115. Vergis JM, Wiener MC. 2011 The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expr. Purif.* **78**, 139–142. (doi:10.1016/j.pep.2011.04.011)
 116. Davis GJ, Wang QM, Cox GA, Johnson RB, Wakulchik M, Dotson CA, Villarreal EC. 1997 Expression and purification of recombinant rhinovirus 14 3CD proteinase and its comparison to the 3C proteinase. *Arch. Biochem. Biophys.* **346**, 125–130. (doi:10.1006/abbi.1997.0291)
 117. Joseph BC, Pichaimuthu S, Srimeenakshi S, Murthy M, Selvakumar K, Ganesan M, Manjunath SR. 2015

- An overview of the parameters for recombinant protein expression in *Escherichia coli*. *J. Cell Sci. Ther.* **6**, 5. (doi:10.4172/2157-7013.1000221)
118. Wu X, Jörnvall H, Berndt KD, Oppermann U. 2004 Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochem. Biophys. Res. Commun.* **313**, 89–96. (doi:10.1016/j.bbrc.2003.11.091)
 119. Prinz WA, Aslund F, Holmgren A, Beckwith J. 1997 The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15 661–15 667. (doi:10.1074/jbc.272.25.15661)
 120. Zhang HC, Cisneros RJ, Dunlap RB, Johnson LF. 1989 Efficient synthesis of mouse thymidylate synthase in *Escherichia coli*. *Gene* **84**, 487–491. (doi:10.1016/0378-1119(89)90525-8)
 121. Weiner M, Anderson C, Jerpseth B, Wells S, Johnson-Browne B, Vaillancourt P. 1994 Studier pET system vectors and hosts. *Strateg. Mol. Biol.* **7**, 41–43.
 122. Loyevsky M *et al.* 2003 Expression of a recombinant IRP-like *Plasmodium falciparum* protein that specifically binds putative plasmodial IREs. *Mol. Biochem. Parasitol.* **126**, 231–238. (doi:10.1016/S0166-6851(02)00278-5)
 123. Schlegel S, Lofblom J, Lee C, Hjelm A, Klepsch M, Strous M, Drew D, Slotboom DJ, de Gier JW. 2012 Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *J. Mol. Biol.* **423**, 648–659. (doi:10.1016/j.jmb.2012.07.019)
 124. 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**, 188–199. (doi:10.1046/j.1365-2958.1999.01465.x)
 125. Hochkoeppler A. 2013 Expanding the landscape of recombinant protein production in *Escherichia coli*. *Biotechnol. Lett.* **35**, 1971–1981. (doi:10.1007/s10529-013-1396-y)
 126. Murata T, Shinozuka Y, Obata Y, Yokoyama KK. 2008 Phosphorylation of two eukaryotic transcription factors, Jun dimerization protein 2 and activation transcription factor 2, in *Escherichia coli* by Jun N-terminal kinase 1. *Anal. Biochem.* **376**, 115–121. (doi:10.1016/j.ab.2008.01.038)
 127. Wacker M *et al.* 2002 N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793. (doi:10.1126/science.298.5599.1790)
 128. Ihssen J, Kowarik M, Dilettoso S, Tanner C, Wacker M, Thony-Meyer L. 2010 Production of glycoprotein vaccines in *Escherichia coli*. *Microb. Cell Fact.* **9**, 61. (doi:10.1186/1475-2859-9-61)
 129. Lizak C, Fan YY, Weber TC, Aebi M. 2011 N-Linked glycosylation of antibody fragments in *Escherichia coli*. *Bioconjug. Chem.* **22**, 488–496. (doi:10.1021/bc100511k)
 130. O'Brien SP, DeLisa MP. 2012 Functional reconstitution of a tunable E3-dependent sumoylation pathway in *Escherichia coli*. *PLoS ONE* **7**, e38671. (doi:10.1371/journal.pone.0038671)
 131. Magnani R, Chaffin B, Dick E, Bricken ML, Houtz RL, Bradley LH. 2012 Utilization of a calmodulin lysine methyltransferase co-expression system for the generation of a combinatorial library of post-translationally modified proteins. *Protein Expr. Purif.* **86**, 83–88. (doi:10.1016/j.pep.2012.09.012)
 132. Duronio RJ, Jackson-Machelski E, Heuckeroth RO, Olins PO, Devine CS, Yonemoto W, Slice LW, Taylor SS, Gordon JL. 1990 Protein N-myristoylation in *Escherichia coli*: reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl Acad. Sci. USA* **87**, 1506–1510. (doi:10.1073/pnas.87.4.1506)
 133. Ren Y, Yao X, Dai H, Li S, Fang H, Chen H, Zhou C. 2011 Production of N α -acetylated thymosin α 1 in *Escherichia coli*. *Microb. Cell Fact.* **10**, 26. (doi:10.1186/1475-2859-10-26)
 134. Gubellini F, Verdon G, Karpowich NK, Luff JD, Boel G, Gauthier N, Handelsman SK, Ades SE, Hunt JF. 2011 Physiological response to membrane protein overexpression in *E. coli*. *Mol. Cell Proteomics* **10**, M111.007930. (doi:10.1074/mcp.M111.007930)
 135. Bill RM *et al.* 2011 Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* **29**, 335–340. (doi:10.1038/nbt.1833)
 136. Marreddy RKR, Geertsma ER, Poolman B. 2011 Recombinant membrane protein production: past, present and future. In *Supramolecular structure and function 10* (eds J Brnjas-Kraljević, G Pifat-Mrčljak), pp. 41–74. Dordrecht, The Netherlands: Springer.
 137. Luijckx J, Yu Z, Wagner S, de Gier JW. 2012 Biogenesis of inner membrane proteins in *Escherichia coli*. *Biochim. Biophys. Acta* **1817**, 965–976. (doi:10.1016/j.bbabi.2011.12.006)
 138. Kwon SK, Kim SK, Lee DH, Kim JF. 2015 Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Sci. Rep.* **5**, 16076. (doi:10.1038/srep16076)
 139. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren PÅ, van Wijk KJ, de Gier JW. 2007 Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* **6**, 1527–1550. (doi:10.1074/mcp.M600431-MCP200)
 140. Chen Y, Song J, Sui SF, Wang DN. 2003 DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in *Escherichia coli*. *Protein Expr. Purif.* **32**, 221–231. (doi:10.1016/S1046-5928(03)00233-X)
 141. Link AJ, Skretas G, Strauch EM, Chari NS, Georgiou G. 2008 Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in *Escherichia coli*. *Protein Sci.* **17**, 1857–1863. (doi:10.1110/ps.035980.108)
 142. Skretas G, Makino T, Varadarajan N, Pogson M, Georgiou G. 2012 Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in *Escherichia coli*. *Metab. Eng.* **14**, 591–602. (doi:10.1016/j.jmb.2012.05.001)
 143. Skretas G, Georgiou G. 2009 Genetic analysis of G protein-coupled receptor expression in *Escherichia coli*: Inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. *Biotechnol. Bioeng.* **102**, 357–367. (doi:10.1002/bit.22097)
 144. Schlegel S, Hjelm A, Baumgarten T, Vikstrom D, de Gier JW. 2014 Bacterial-based membrane protein production. *Biochim. Biophys. Acta* **1843**, 1739–1749. (doi:10.1016/j.bbamcr.2013.10.023)
 145. Sezonov G, Joseleau-Petit D, D'Ari R. 2007 *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749. (doi:10.1128/jb.01368-07)
 146. Studier FW. 2005 Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234. (doi:10.1016/j.pep.2005.01.016)
 147. Losen M, Frölich B, Pohl M, Büchs J. 2004 Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnol. Prog.* **20**, 1062–1068. (doi:10.1021/bp034282t)
 148. Choi JH, Keum KC, Lee SY. 2006 Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chem. Eng. Technol.* **61**, 876–885. (doi:10.1016/j.ces.2005.03.031)
 149. Song JM, An YJ, Kang MH, Lee YH, Cha SS. 2012 Cultivation at 6–10°C is an effective strategy to overcome the insolubility of recombinant proteins in *Escherichia coli*. *Protein Expr. Purif.* **82**, 297–301. (doi:10.1016/j.pep.2012.01.020)
 150. Castellanos-Mendoza A, Castro-Acosta RM, Olvera A, Zavala G, Mendoza-Vera M, García-Hernández E, Alagón A, Trujillo-Roldán MA, Valdez-Cruz NA. 2014 Influence of pH control in the formation of inclusion bodies during production of recombinant sphingomyelinase-D in *Escherichia coli*. *Microb. Cell Fact.* **13**, 1–14. (doi:10.1186/s12934-014-0137-9)
 151. Yang Q, Xu J, Li M, Lei X, An L. 2003 High-level expression of a soluble snake venom enzyme, glosheobin, in *E. coli* in the presence of metal ions. *Biotechnol. Lett.* **25**, 607–610. (doi:10.1023/A:1023067626846)
 152. Sørensen HP, Mortensen KK. 2005 Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **4**, 1–8. (doi:10.1186/1475-2859-4-1)
 153. Duan X, Zou C, Wu J. 2015 Triton X-100 enhances the solubility and secretion ratio of aggregation-prone pullulanase produced in *Escherichia coli*. *Bioresour. Technol.* **194**, 137–143. (doi:10.1016/j.biortech.2015.07.024)
 154. Mondal S, Shet D, Prasanna C, Atreya HS. 2013 High yield expression of proteins in *E. coli* for NMR studies. *Adv. Biosci. Biotechnol.* **4**, 751–767. (doi:10.4236/abb.2013.46099)
 155. Correa A, Oppezzo P. 2015 Overcoming the solubility problem in *E. coli*: available approaches for recombinant protein production. *Methods Mol. Biol.* **1258**, 27–44. (doi:10.1007/978-1-4939-2205-5_2)

156. Correa A, Oppezzo P. 2011 Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening. *Biotechnol. J.* **6**, 715–730. (doi:10.1002/biot.201100025)
157. Ukkonen K, Mayer S, Vasala A, Neubauer P. 2013 Use of slow glucose feeding as supporting carbon source in lactose autoinduction medium improves the robustness of protein expression at different aeration conditions. *Protein Expr. Purif.* **91**, 147–154. (doi:10.1016/j.pep.2013.07.016)
158. Vincentelli R, Romier C. 2013 Expression in *Escherichia coli*: becoming faster and more complex. *Curr. Opin. Struct. Biol.* **23**, 326–334. (doi:10.1016/j.sbi.2013.01.006)
159. Rohe P, Venkanna D, Kleine B, Freudl R, Oldiges M. 2012 An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform. *Microb. Cell Fact.* **11**, 1–14. (doi:10.1186/1475-2859-11-144)
160. Long Q, Liu X, Yang Y, Li L, Harvey L, McNeil B, Bai Z. 2014 The development and application of high throughput cultivation technology in bioprocess development. *J. Biotechnol.* **192**, 323–338. (doi:10.1016/j.jbiotec.2014.03.028)
161. Grunzel P, Pilarek M, Steinbruck D, Neubauer A, Brand E, Kumke MU, Neubauer P, Krause M. 2014 Mini-scale cultivation method enables expeditious plasmid production in *Escherichia coli*. *Biotechnol. J.* **9**, 128–136. (doi:10.1002/biot.201300177)
162. Kong F, Yuan L, Zheng YF, Chen W. 2012 Automatic liquid handling for life science: a critical review of the current state of the art. *J. Lab. Auto.* **17**, 169–185. (doi:10.1177/2211068211435302)
163. Hughes SR, Butt TR, Bartoletti S, Riedmuller SB, Farrelly P. 2011 Design and construction of a first-generation high-throughput integrated robotic molecular biology platform for bioenergy applications. *J. Lab. Auto.* **16**, 292–307. (doi:10.1016/j.jala.2011.04.004)
164. Wiesler SC, Weinzierl RO. 2015 Robotic high-throughput purification of affinity-tagged recombinant proteins. *Methods Mol. Biol.* **1286**, 97–106. (doi:10.1007/978-1-4939-2447-9_9)
165. May M. 2016 Automated sample preparation. *Science* **351**, 300–302. (doi:10.1126/science.351.6270.300)
166. de Marco A. 2013 Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. *Microb. Cell Fact.* **12**, 101. (doi:10.1186/1475-2859-12-101)