SCUOLA
NORMALE
SUPERIORE

SELEZIONE PUBBLICA, PER TITOLI ED ESAMI, PER LA COPERTURA DI N. 1 POSTO DI CATEGORIA D – POSIZIONE ECONOMICA D1 – AREA TECNICA, TECNICO-SCIENTIFICA ED ELABORAZIONE DATI, A TEMPO INDETERMINATO E PIENO, PER LA STRUTTURA DI GESTIONE SAN CATALDO A SUPPORTO DEL LABORATORIO DI BIOLOGIA (BIO@SNS)

Criteri di valutazione delle prove

La prova consisterà nel descrivere in modo sintetico ad una traccia, sorteggiata tra tre possibili.
Per la valutazione la Commissione terrà conto dei seguenti criteri:

- correttezza metodologica
- ampiezza dell'informazione
- chiarezza di esposizione
- capacità di sintesi

Per la valutazione della prova orale la Commissione terrà conto dei seguenti criteri: chiarezza espositiva, completezza e approfondimento dell'argomento.
Verrà predisposto un argomento a cui verrà assegnato un punteggio massimo di 25 punti; inoltre verrà chiesta la traduzione dall'inglese all'italiano di un frammento di testo scientifico a cui sarà assegnato un punteggio massimo di 2 punti ed infine verrà posta una domanda di natura informatica a cui sarà assegnato il punteggio massimo di 3 punti.

Criteri di valutazione dei titoli

A. LAUREA: max 7 punti

- Diploma di Laurea rilasciato secondo l'ordinamento previgente ai DM 509/99 e DM 270/2004 o della Laurea Specialistica/Magistrale rilasciata secondo l'ordinamento di cui ai DM 509/99 e DM 270/2004: max 7 punti su 30, secondo la seguente suddivisione:

Voto laurea	Punti assegnati
110 + Lode	7
100-110	6
<=99	4,5

- Laurea di I livello rilasciata secondo l'ordinamento successivo ai DM 509/99 e DM 270/2004 (valutabile solo per i candidati che non siano anche in possesso di Laurea Specialistica/Magistrale): max 4 punti su 30, secondo la seguente suddivisione:

Voto laurea	Punti assegnati
110 + Lode	4
100-110	3
<=99	1

B. TITOLI ACCADEMICI POST-LAUREA : solo se giudicati attinenti alla professionalità oggetto di selezione, max 6 punti su 30 assegnati come di seguito indicato:

Tipologia titolo	Punti assegnati
Titolo di Dottore di Ricerca (PhD)	6
Master universitario di secondo livello	1
Master universitario di primo livello	0.5

C. ATTIVITA' LAVORATIVA prestata con rapporto di lavoro subordinato o autonomo presso università, soggetti pubblici o privati, giudicata attinente alla professionalità oggetto di selezione (sarà riconosciuto e valorizzato, in relazione alla durata, l'eventuale attività lavorativa prestata presso le Università): max 12 punti su 30 assegnati come di seguito indicato:

a. Attività di lavoro subordinato max 6 punti
c/o amministrazioni non universitarie e c/o privati: 0.1 p. per ogni semestre
c/o amministrazioni universitarie: 0.5 p. per ogni semestre
Non sono utili ai fini della valutazione stage, tirocini, borse di studio e collaborazioni studentesche.
Per periodi inferiori a 6 mesi il punteggio verrà assegnato forfettariamente:
c/o amministrazioni non universitarie e c/o privati: 0.01
c/o amministrazioni universitarie: 0.05

b. Attività di lavoro autonomo max 6 punti
c/o amministrazioni non universitarie e c/o privati: 0.1 p. per ogni semestre
c/o amministrazioni universitarie: 0.5 p. per ogni semestre
Non sono utili ai fini della valutazione stage, tirocini, borse di studio e collaborazioni studentesche.
Per periodi inferiori a 6 mesi il punteggio verrà assegnato forfettariamente:
c/o amministrazioni non universitarie e c/o privati: 0.01
c/o amministrazioni universitarie: 0.05

D. PUBBLICAZIONI PERTINENTI: max 5 punti su 30 assegnati, come di seguito indicato:
Saranno valutate pubblicazioni in cui il candidato è autore/coautore che dimostrino la sua conoscenza e/o competenza in uno degli ambiti indicati nell'art. 1 del bando: max 1,00 p. a pubblicazione

Quesiti delle prove scritte

TEMA 1
Il/La candidato/a descriva in maniera sintetica la modalità di preparazione di una proteina ricombinante.

TEMA 2
Il/La candidato/a descriva in maniera sintetica la modalità di preparazione di una muteina ricombinante.

TEMA 3
Il/La candidato/a descriva in maniera sintetica la modalità di preparazione di un anticorpo ricombinante.

Quesiti dei colloqui

PROVA N. 1
- Descrizione della metodologia sperimentale di Western Blot
- Prova di Inglese: il candidato legga e traduca in italiano parte del testo proposto allegato
- Prova di informatica: descrizione dei programmi di riferimento per:
  gestione delle sequenze di acidi nucleici

PROVA N. 2
- Descrizione della tecnica sperimentale dell'elettroforesi
- Prova di Inglese: il candidato legga e riassuma per la commissione in italiano parte del testo proposto allegato
- Prova di informatica: descrizione dei programmi di riferimento per la gestione delle immagini

PROVA N. 3
- Descrizione dell'uso della PCR (Polymerase Chain Reaction)
- Prova di Inglese: il candidato legga e riassuma per la commissione in italiano parte del testo proposto allegato
- Prova di informatica: descrizione dei programmi di riferimento per effettuare una ricerca bibliografica on-line

PROVA N. 4
- metodiche di controllo dell'attività fisiologica di proteine ed anticorpi ricombinanti;
- Prova di Inglese: il candidato legga e riassuma per la commissione in italiano parte del testo proposto allegato
- Prova di informatica: descrizione dei programmi di riferimento per determinare caratteristiche salienti di una proteina (punto isoelettrico, epsilon molare)

PROVA N. 5

- coltivazione di cellule eucariotiche e procariotiche (cellule di mammifero, batteri)
- Prova di Inglese: il candidato legga e riassuma per la commissione in italiano parte del testo proposto allegato
- Prova di informatica: descrizione dei programmi di riferimento per effettuare una analisi statistica

TeaserProduction of recombinant proteins is essential for drug development and discovery, but can often be problematic; why do these processes fail and how can these problems be overcome?

Recombinant protein production in bacterial hosts

Tim W. Overton

Bioengineering, School of Chemical Engineering, The University of Birmingham, Birmingham, B15 2TT, UK

Tim W. Overton is a lecturer in the School of Chemical Engineering at The University of Birmingham. A biochemist and molecular microbiologist by training, his research focuses on bioprocess optimisation using molecular microbiology and flow cytometry approaches. Currently, he is investigating recombinant protein production and biofilm-mediated biotransformation processes and the elimination of bacteria from varied feedstreams. He is the director of the MSc programme in Biochemical Engineering, which trains science and engineering graduates to become the next generation of bioprocess engineers, with a focus on biopharmaceutical engineering.

The production of recombinant proteins is crucial for both the development of new protein drugs and the structural determination of drug targets. As such, recombinant protein production has a major role in drug development. Bacterial hosts are commonly used for the production of recombinant proteins, accounting for approximately 30% of current biopharmaceuticals on the market. In this review, I introduce fundamental concepts in recombinant protein production in bacteria, from drug development to production scales. Recombinant protein production processes can often fail, but how can this failure be minimised to rapidly deliver maximum yields of high-quality protein and so accelerate drug discovery?

Proteins and peptides are the mainstay of the biopharmaceutical sector; over 200 protein drug products are currently on the market [1,2], and more are currently undergoing preclinical and clinical trials. However, proteins are complex molecules in terms of their structure and function and, unlike many pharmaceuticals, cannot be synthesised chemically. Therefore, proteins are manufactured in biological processes, usually inside host cells (although a growing number of cell-free expression technologies are available). These proteins, synthesised in a host cell frequently of a different species to their origin, are termed 'recombinant proteins' (see Glossary) because the DNA encoding them has been recombined or engineered.

Recombinant proteins are required at different stages of the drug discovery process and in different quantities. Initial drug development studies frequently involve the structural determination of proteins that are drug targets, for example human membrane proteins [3]; such structural studies are often required in the development of protein and small molecule drugs. These studies typically require small quantities of recombinant protein (on the milligram scale). Further in the development process, larger quantities of protein drug are required for preclinical and clinical trials, synthesised under current good manufacturing practice (cGMP) conditions [4]. Systems must be in place to enable the synthesis of recombinant proteins in this range of scales so that drug development can proceed. Large drug companies undertake much of this work in-house, whereas other companies might contract out recombinant protein production to contract manufacturers.

E-mail address: t.w.overton@bham.ac.uk.

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2013.11.008

GLOSSARY

Affinity chromatography separation of proteins based on binding affinity. For example, MBP has high binding affinity to amylose; therefore, amylose can be used to attract MBP and selectively purify it from other proteins.
Bacteriophage a virus that infects bacteria.
Bioreactor a tank, usually made of stainless steel or glass, that is used for growth of organisms. Bioreactors are usually designed to maintain culture conditions within certain parameters (e.g. temperature and pH control), mix the contents and supply the cells with oxygen and nutrients. Bioreactors also enable containment, ensuring that the cells are not contaminated by external factors or vice versa.
Current good manufacturing practice (cGMP) the set of guidelines that governs the safe production of pharmaceuticals for human use, as defined by regional drug regulatory bodies, such as the FDA (http://www.fda.gov) and Medicines and Healthcare Products Regulatory Agency (MHRA [4]).
Fab fragment a truncated antibody fragment comprising one light chain (V_L and C_L domains) and one truncated heavy chain (V_H and C_H1 domains), disulfide-bonded together. The Fab fragment contains the antigen-binding portion of the antibody.
Glycosylation oligosaccharide moieties are frequently attached to proteins of eukaryotic origin. This process occurs in the endoplasmic reticulum and is often essential for correct protein function. Antibodies usually require glycosylation for correct function. Lack of glycosylation, or incorrect glycosylation (the incorrect sugar molecules being attached to the protein), leads to loss of protein function and frequently rapid clearance of protein drugs from the body.
Gram negative/positive classification of bacteria depending upon cell wall structure. Gram-negative bacteria have two membranes separated by a periplasm, which contains a thin layer of peptidoglycan, a structural polysaccharide. Gram-positive bacteria (e.g. *Bacillus* spp) have cell walls comprising a single membrane and a thicker layer of peptidoglycan.
Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducer molecule for the Lac and pET expression systems. IPTG binds to the Lac repressor LacI, whose natural ligand is the disaccharide lactose, resulting in repression of DNA binding. IPTG is used in recombinant protein production as, unlike lactose, it is non-metabolisable and, therefore, is not degraded by bacteria over time.
Metabolic burden the stress caused by requirements for energy generated by metabolism (in the form of ATP) and metabolic intermediates within the cell.
Plasmid a circular DNA fragment, usually 1–100 kb in size, that is replicated independently of the host cell chromosome.
Post-translational modification biochemical modifications to a protein that occur following translation. These include glycosylation, phosphorylation, and acylation [5].
Promoter the region of DNA upstream of a gene or set of genes that specifies when transcription occurs.
RNA polymerase a multisubunit enzyme complex that catalyses the process of transcription. RNA polymerase recognises and binds to a promoter region upstream of a gene and then generates the RNA molecule corresponding to the DNA sequence of the gene.
RNA polymerase sigma S subunit (RpoS) a master regulator of the general stress response in *E. coli*.
Single-chain variable fragment (scFv) comprises the variable domains of the light (V_L) and heavy (V_H) chains of an antibody, linked together by a flexible peptide linker. It contains the antigen-binding portion of the antibody.
Site-directed mutagenesis specific changes to DNA nucleotide sequences that are made to alter the peptide sequence of a protein, the promoter sequence or codon usage.
Transformation the process by which bacteria are induced to take up plasmid DNA. When containing the plasmid, the bacteria are termed 'transformants'.

The first decision that must be made when making a recombinant protein is choice of the host system. Before choosing a host, the chemical properties of the desired recombinant protein are investigated; if the desired protein is glycosylated or otherwise extensively post-translationally modified [5], then a eukaryotic expression system is usually chosen, because such modifications are frequently essential for correct protein function and bacteria are currently unable to incorporate such modifications. However, disulfide bond formation (DSB) is possible in bacteria [6]. The development of eukaryotic-like post-translational modification in bacterial hosts is a current area of research, recently reviewed by Nothaft and Szymanski [7]. Some possible eukaryotic expression systems are outlined in Box 1.

For proteins that are not required to be synthesised in a glycosylated or extensively post-translationally modified form, bacteria are an excellent expression system because of their relative sim-

BOX 1

Eukaryotic hosts: their advantages and drawbacks.

A central problem with bacteria as hosts for recombinant protein production is their inability to post-translationally modify proteins in the way that human cells can, for example glycosylation (attachment of antennae of specific sugar epitopes to proteins) [5]. One reason for this is their different cellular structure; bacteria lack the endoplasmic reticulum and Golgi apparatus, the sites of post-translational modification. If a recombinant protein is required to be glycosylated or otherwise modified (e.g. phosphorylated or lipidated), then a eukaryotic host is usually used. Many glycosylated and otherwise post-translationally modified protein biotherapeutics are generated in cultured mammalian cell lines such as CHO cells (recently reviewed in [69]). Manufacture in these hosts is more expensive and complicated than bacterial processes, owing to the higher cost of culture media, low cell tolerance for changes in reaction conditions and slow growth rates. Typical protein drug products of CHO cells include monoclonal antibody therapies for treating cancer and erythropoietin [1].
A halfway house between mammalian cells and bacteria remains yeast systems, typified by the bakers' and brewers' yeast *Saccharomyces cerevisiae* and the methylotrophic (methanol-utilising) *Pichia pastoris* (reviewed in [70]). Given that yeasts are eukaryotic, they are able to generate some post-translational modifications such as glycosylation; current research is focusing on engineering yeast glycosylation patterns to mimic human cells. Yeast are also simple and quick to grow and can generate high yields of recombinant proteins. Yeast-generated protein drugs on the market include vaccines and insulin [1,14]. Ultimately, the choice of host cell system is usually a compromise between the ease and cost of growth and the overall yield and function of the generated recombinant protein.

plicity, both biologically (in terms of biochemistry and physiology) and from a process perspective [8]. Bacterial processes also tend to be cheaper than eukaryotic cell processes because of lower media costs and shorter process times. The most commonly used bacterium for recombinant protein production is *Escherichia coli*, an enteric bacterium that has a long pedigree of safe use in laboratories and industry [9]. *E. coli* is a particularly suitable host because it is well characterised physiologically and metabolically, it was among the first organisms to have its entire genome sequenced [10] and many molecular biology tools are available for engineering its DNA sequences to generate novel functionality [11]. *E. coli* is a cost-effective host for recombinant protein production and is recognised by drug regulatory authorities.

The first recombinant human protein to be generated in *E. coli* was somatostatin in 1977 [12]. Genentech received approval from the US Food and Drug Administration (FDA) for recombinant human insulin in 1982, the first commercial recombinant drug, also made in *E. coli* [13]. Since then, many recombinant drugs synthesised in *E. coli* have been approved for human use, including thrombolytics, hormones, growth factors, interferons and antibody fragments (reviewed in [1]). Currently, approximately 30% of recombinant biologic products on the market are manufactured in *E. coli* [14]. Recent FDA approvals for recombinant protein drugs made in *E. coli* include Voraxaze® (BTG International, 2012), used for the treatment of delayed methotrexate clearance in patients with impaired renal function; the interferon β-1b Extavia® (Novartis, 2009) used in multiple sclerosis treatment; and the gout treatment Krystexxa® (Savient Pharmaceuticals, 2010). *E. coli* remains a key host for the production of biopharmaceutical products, both those currently on the market and at various stages of the drug development pathway. Financial pressures might also lead to a resurgence of *E. coli* as a host for biologic manufacture, as the number of patients able to afford more expensive protein biopharmaceuticals generated in mammalian cell systems falls.

Other bacterial species can also be used for the production of recombinant proteins. *Bacillus* species in particular have found use in the production of industrially and pharmaceutically relevant recombinant proteins [8,15]. *Lactococcus lactis* has also been identified as a possible host for recombinant protein production, and has a long history of use in food fermentation [16,17]. However, at present, no recombinant biopharmaceuticals approved for human drug use are made in non-*E. coli* bacteria.

In this foundational review, I discuss the outline of the process of recombinant protein production in *E. coli* and consider the choices that have to be made in the design of recombinant protein production strategies. Given that many recombinant proteins are difficult to manufacture, I then focus on the troubleshooting of recombinant protein production processes, highlighting potential solutions and ways to improve recombinant protein yield and activity. Finally, I discuss future directions for recombinant protein production in bacteria.

Outline of the process and relation to stages in drug development

The bacterial recombinant protein production process is outlined in Fig. 1. The gene encoding the desired protein is first cloned into the multiple cloning site (MCS) of an expression vector under the control of a promoter that will regulate expression of the gene. If the gene contains introns, it is usually cloned from a cDNA library because bacteria cannot excise introns. The plasmid vector is transformed into a strain of *E. coli* that is capable of recombinant protein production, and the transformants are grown in liquid culture. At a specific stage of growth, production of the recombinant protein is induced by the addition of a chemical inducer that will activate the promoter on the expression vector. Thus, the recombinant gene is expressed and the recombinant polypeptide chain folds into the recombinant protein of interest. The recombinant protein can then be released from the cell, captured and purified [18].

The overall process is similar in principle for the production of recombinant proteins for drug screening, clinical trials or final production following approval. Given that a key consideration during the development of protein biopharmaceuticals is the ability to produce the chosen protein on a commercial scale, problems in the production of small quantities of a candidate recombinant protein drug for screening studies would count against it when selecting candidates to take forward to later stages of the drug development process.

The scale of the bacterial growth is the major factor that changes in recombinant protein production from screening to commercial production. Initial growth experiments are usually done on a small scale, typically 10–100 ml. Early experiments focus on the production of relatively low concentrations of the recombinant protein of interest and the folding state and biological activity of the produced protein. These expression-screening studies can be done manually in small conical flasks grown in a shaking incubator, or using high-throughput automated systems that enable scores of simultaneous cultures to be automatically controlled and monitored. These high-throughput systems often assess not only the production of the recombinant protein in the bacteria, but also early stages of capture and purification, giving an indication of the overall processability of the recombinant protein (reviewed in [19]).

Following initial batch cultures on a small scale, production is usually shifted to stirred-tank bioreactors operating on fed-batch regimens [20]. This operation strategy is used to increase greatly the generated biomass by feeding large quantities of growth substrate, such as glucose or glycerol, to the growing bacteria as required. Gradual feeding of glucose in particular is necessary because *E. coli* preferentially converts glucose to acetate when exposed to high glucose concentrations in a process known as overflow metabolism, thereby lowering the pH and inhibiting growth. Gradual feeding enables the total mass of glucose added to the bacteria to be high while maintaining a continually low glucose concentration. Fed-batch operation enables cell densities of greater than 100 g dry cell weight per litre to be achieved, thus massively increasing the recombinant protein yield per litre. Growth in bioreactors proceeds at varying scales, depending upon the stage in the drug development process. Initial bioreactor growth is typically on the 1–10 l scale, followed by scale-up to pilot scale (approximately 50–500 l) and production scale (10 000 l and more). Exact scales are also dependent upon the quantity of protein required at each stage of the development process.

The final stage of recombinant protein production is release and purification, which are not covered extensively here; there are recent reviews and texts on these subjects [18,21,22]. However, it

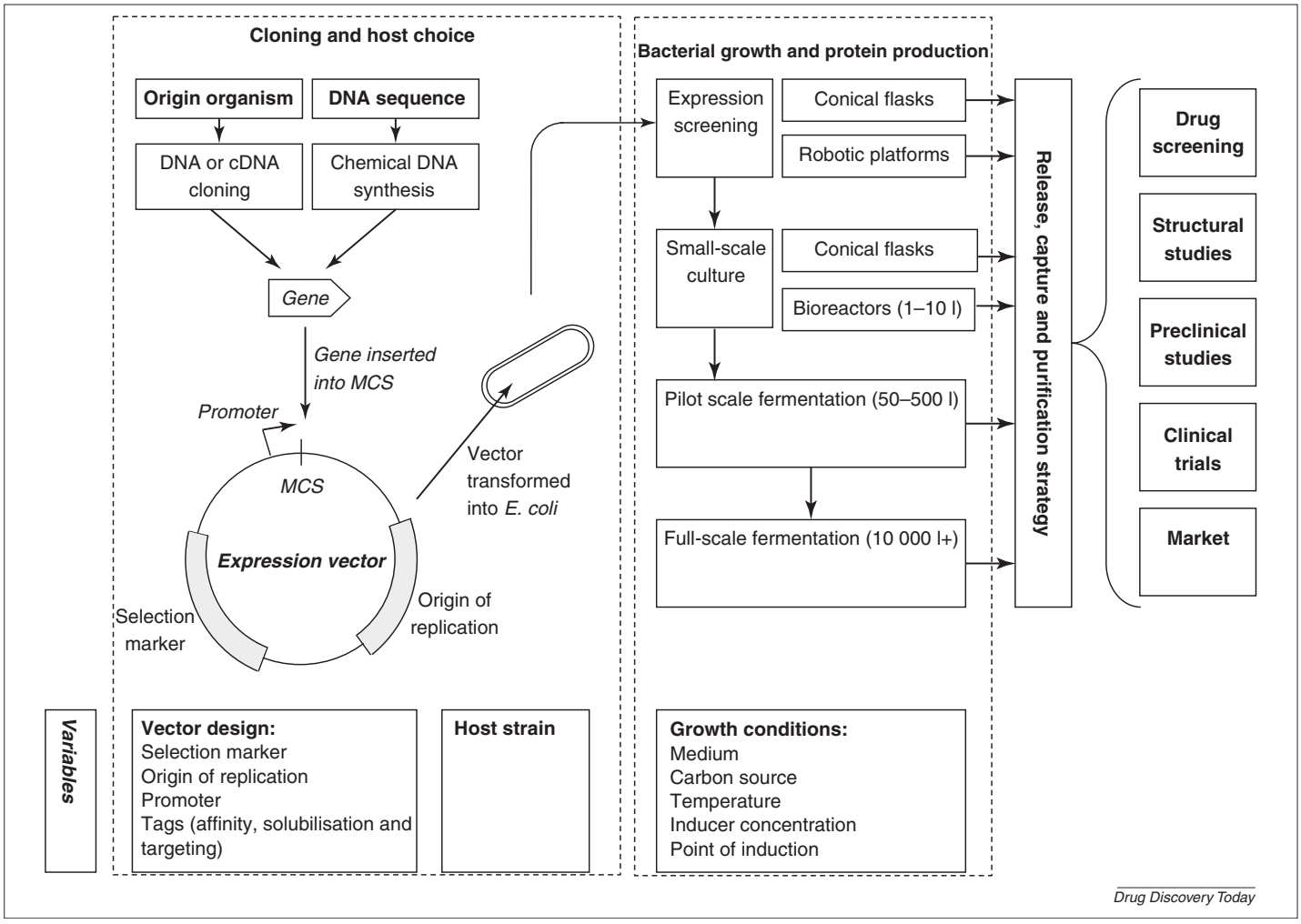

FIGURE 1

Outline of recombinant protein production process in *Escherichia coli*. The process is described in detail in the main text. Abbreviation: MCS, multiple cloning site.

should be noted that considerable efforts can be made in the design of bacterial growth strategies to simplify protein release and purification; some of these considerations are discussed below.

Process decisions at the DNA level

When designing the recombinant protein production process, some of the first decisions to be made relate to the characteristics of the expression vector (Fig. 1), specifically the promoter from which the recombinant gene of interest will be expressed, the selection marker that enables maintenance of the plasmid in the host cell, and the origin of replication, which regulates the number of copies of the plasmid in each host cell.

The promoter from which expression of the recombinant gene of interest is regulated is important to the expression system. Promoters are usually regulated by a chemical inducer, which activates transcription when added to the culture; this enables temporal control of protein production and, frequently, separation of the cell growth and protein production phases of the process. Desired characteristics of promoters are tight control, so that the promoter is switched entirely 'off' in the absence of inducer, and regulatable expression levels that are dependent upon the concentration of inducer molecule added to the culture. Some commonly used promoter systems are described in Table 1.

The pET system, based on the T7 RNA polymerase, is commonly used in recombinant protein production (Fig. 2 [23]). The pET system relies upon an engineered *E. coli* host that carries a chromosomal copy of the gene encoding the RNA polymerase of bacteriophage T7. This RNA polymerase gene is usually under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter, such as *lacUV5* [24]; this construct is stably incorporated into the bacterial chromosome at the DE3 locus. In the absence of IPTG, the *lac* promoter is bound by the Lac repressor LacI, which represses transcription; T7 RNA polymerase is not synthesised. Upon addition of IPTG, LacI is released from the *lac* promoter region and the T7 RNA polymerase gene is transcribed and translated. The T7 RNA polymerase is then able to activate transcription from the T7 promoter located on the pET expression plasmid, downstream of which is cloned the recombinant gene of interest. Crucially, the T7 promoter on the pET expression plasmid is not strongly activated by the *E. coli* RNA polymerase, so expression of the gene of interest is dependent upon IPTG. The pET system is commercially available from Novagen.

Although popular, the pET system has various drawbacks. The levels of T7 RNA polymerase generated in the cell can be high, leading to a massive amount of transcription of the recombinant gene of interest. Control of expression is frequently leaky, in that

TABLE 1

Commonly used promoter systems for the regulation of recombinant protein production in *Escherichia coli*

Promoter system	Source	Basis of regulation	Notes	Refs
pET system (DE3/T7)	Engineered from *E. coli lac* promoter, T7 RNA polymerase gene and T7 promoters	See main text and Fig. 2		[23]
Other T7 systems	Various	As pET, but production of the T7 RNA polymerase is induced by signals other than IPTG		
***lac* systems (*Plac, PlacUV5*)**	Natural or modified versions of *E. coli lac* promoter	Expression is repressed by the *lac* repressor, LacI. This repression is lifted by addition of a ligand of LacI, which is naturally lactose, but a non-metabolisable analogue, such as IPTG, can be used		[71]
tac/trc	Hybrid of *lacUV5* and *trp* promoters	As *lac* systems	Observed to be more effective than *lac* promoter systems	[72]
pBAD	Arabinose operon of *E. coli*	*araBAD* promoter is repressed by the arabinose repressor AraC. Repression is lifted upon addition of arabinose	Frequently has tight control of expression	[73]
λ p_L	Promoter and repressor cI from λ phage	p_L promoter is repressed by cI repressor protein. A temperature-sensitive version of cI (cI857) is stable at 30°C but unstable and, thus, lifts repression at 42°C	Induction requires growth at 42°C, which might not be beneficial for correct protein folding	[74]

low levels of recombinant gene expression can occur in the absence of IPTG. This can result from low levels of T7 RNA polymerase production in the absence of IPTG or low levels of expression from the T7 promoter on the pET vector caused by *E. coli* RNA polymerase. This can be a specific problem when the recombinant protein being generated causes depression of cell growth (discussed in more detail under 'toxic proteins', below.) A possible remedy for this is to include a *lac* operator sequence overlapping the T7 promoter on the expression vector, so that the Lac repressor binds to the promoters regulating expression of both the T7 RNA polymerase and the recombinant gene, and silencing 'leaky' expression [25]. Alternatively, additional plasmids (pLysE and/or pLysS) can be used to express the T7 lysozyme in host strains, which is a natural repressor of T7 RNA polymerase [23].

No one promoter system is suitable for all recombinant proteins; promoters are chosen based on their characteristics for each target protein. Promoter development continues, with an aim to generate promoter systems that are minimally leaky, tuneable (such that different inducer concentrations give rise to different levels of expression) and enable a high level of recombinant protein production [26].

Maintenance of plasmids is a source of metabolic burden to the host bacterium; therefore, it is usually selected for by the use of an antibiotic and corresponding resistance gene on the plasmid. Cells not carrying the plasmid should be unable to grow in medium containing the antibiotic. Penicillin derivatives, such as ampicillin, are commonly used; however, the penicillin resistance gene product, β-lactamase, is a periplasmic protein, and frequently leaks through the outer membrane into the extracellular milieu, where it degrades the antibiotic in the medium, thus decreasing the antibiotic concentration and the selective pressure. Over the course of a cultivation, the number of cells bearing plasmid can decrease (frequently down to 0%); this loss is exacerbated by the fact that plasmid-free cells are at a competitive advantage compared with plasmid positive cells because of a decreased metabolic burden, having to neither replicate the plasmid DNA nor generate recombinant protein. Most plasmid retention assays are time consuming and take hours to generate data; therefore, analysis of plasmid retention cannot be used to make changes to the process in real time to increase plasmid retention. Common solutions to the problems of β-lactamase systems are the use of a penicillin derivative that is less susceptible to degradation, such as carbenicillin, or the use of alternative antibiotic resistance cassettes, such as those encoding resistance to tetracycline, kanamycin or chloramphenicol.

A general disadvantage of antibiotic selection is the need to eliminate residual antibiotic completely from the generated recombinant protein drug before administration to humans (e.g. to eliminate the risk of allergic reactions). In addition, increasing levels of antibiotic resistance in bacteria in the environment are threatening human health and limiting the clinical effectiveness of antibiotics in treating bacterially caused disease [27]; therefore process industries are looking to phase out the use of antibiotics as much as possible. For these reasons, metabolic selection markers are gaining acceptance in the field [28]. This mechanism involves a host bacterium that is lacking an essential gene, which is carried on the expression plasmid. A metabolic selection marker might be a gene involved in DNA metabolism or cell wall synthesis.

One example is the repressor titration system described by Cranenburgh *et al.* [29]. Host cells carry a chromosomal copy of the essential *dapD* gene regulated by a LacI-repressed, IPTG-inducible promoter. To grow, bacteria must either be provided with IPTG (enabling production of the DapD enzyme) or the product of the pathway that DapD is a member of [diaminopimelate (DAP, required for cell wall crosslinking)]. Plasmids carry the Lac operator sequence, which binds the LacI protein within the bacteria, thus relieving repression and enabling DapD to be synthesised.

The origin of replication is a DNA sequence carried on a plasmid that instructs the host bacterium to replicate the plasmid [30]. It also determines how many copies of the plasmid should be present per bacterium (the copy number); this can vary between 1 and several hundred or more. Higher copy numbers generate a higher gene dosage, with more copies of the recombinant gene in each bacterium and, thus, can result in higher recombinant protein production; however, this will also generate a higher metabolic burden for the host bacterium, because of the energy and metabolites needed for both synthesis of the recombinant protein and replication of the plasmid DNA. This metabolic burden can increase cellular stress levels, and can lead to plasmid rejection, where bacteria eliminate the plasmid DNA to decrease the metabolic burden, thereby decreasing recombinant protein production. In each case, the balance between high copy number for high recombinant protein productivity and low copy number for

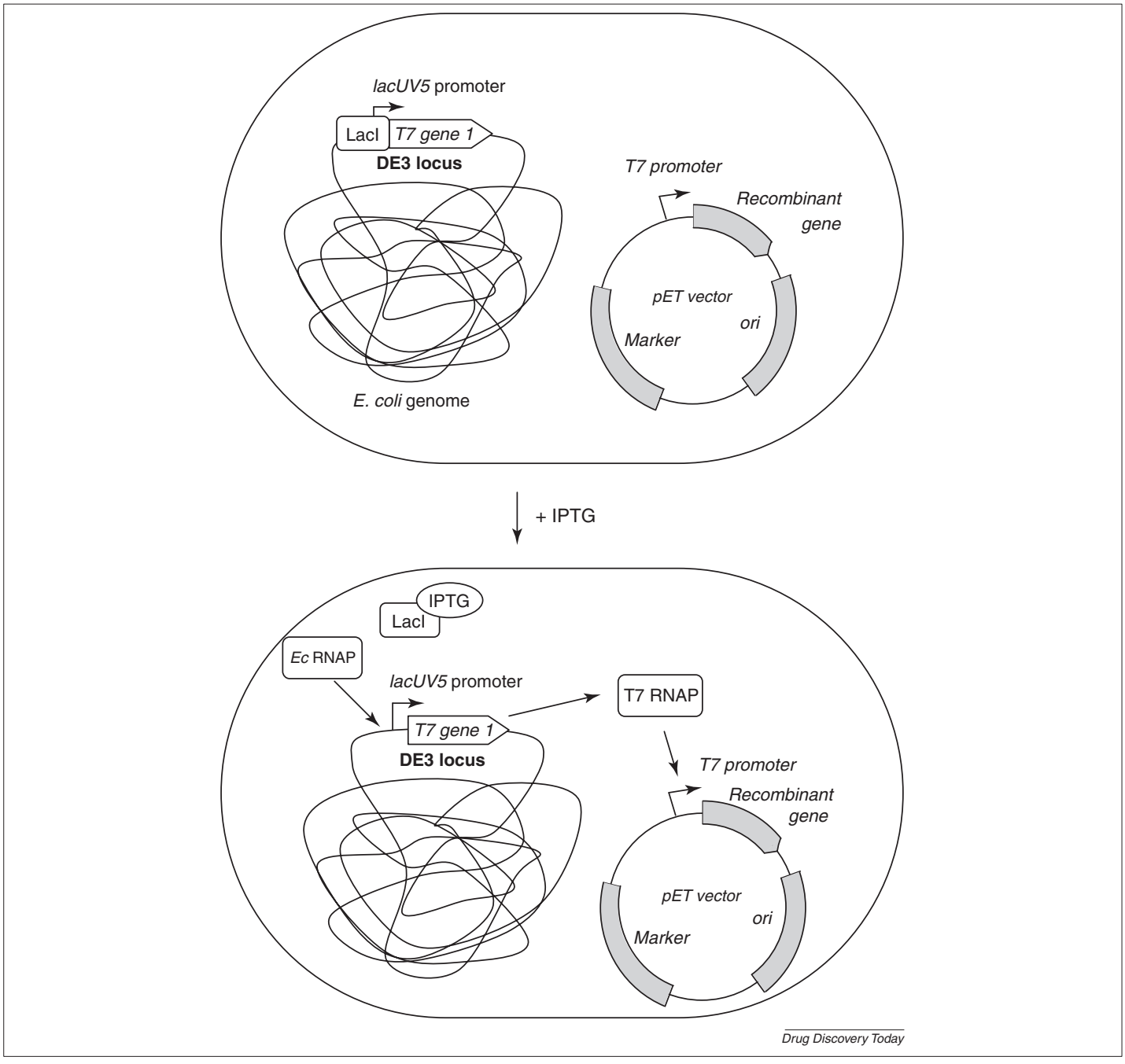

FIGURE 2

The pET system. The recombinant gene of interest is carried on a plasmid vector under the control of a T7 promoter, which cannot be activated by the native *Escherichia coli* RNA polymerase enzyme. The T7 RNA polymerase (RNAP) enzyme is required for transcription; this is encoded on a portion of the host cell genome known as the DE3 locus. Expression of the T7 RNA polymerase is regulated by the *lacUV5* promoter, which is repressed by the Lac repressor LacI **(a)**. To express the recombinant gene, isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the system **(b)**, which is a ligand for the Lac repressor LacI. When bound to IPTG, LacI vacates the *lacUV5* promoter, enabling *E. coli* RNAP to transcribe the T7 gene 1, encoding the T7 RNAP. T7 RNAP is then able to activate the promoter on the expression vector and transcribe the recombinant gene. Abbreviation: ori, origin of replication.

decreased metabolic burden must be determined and optimised. As well as the copy number of the plasmid, a further consideration is that of compatibility; if two plasmids are required in a bacterial host, then they must have compatible origins of replication, such that they both can be maintained.

Apart from the basic components of the expression vector, further DNA components can be cloned upstream or downstream of the recombinant gene of interest. These features are discussed below, in the context of their function and use in recombinant protein production.

Process decisions at the bacterial and process levels

As well as choosing the plasmid carrying the recombinant gene of interest, the choice of *E. coli* host strain can also be important for the success of the protein production process. Strain choice must first be guided by the requirements of the expression plasmid system: for example, is a chromosomally encoded T7 RNA polymerase required, as in pET vectors? Following such considerations, the main choice is frequently between two *E. coli* strains: K-12 and BL21. *E. coli* K-12 was initially isolated in 1922, whereas BL21 was developed from *E. coli* B, isolated in 1918 [31]. From a bioprocessing perspective, BL21 is a desirable strain because it frequently exhibits a higher biomass yield and lower acetate production compared with K-12 [32,33]. The most commonly used host for pET vector systems is BL21 (DE3), although K-12 DE3 derivatives are also available.

The choice of growth medium is crucial for process success, but must be made with consideration of regulatory requirements. The growth medium must contain all the nutrients required for bacterial growth [34]. This typically includes carbon and nitrogen sources, amino acids and micronutrients, such as vitamins (e.g. biotin and nicotinic acid) and metal ions (e.g. Fe, Ni, Co, Mo and Mn). Bacteria such as *E. coli* are metabolically versatile and can frequently interconvert biochemicals (e.g. by generating amino acids from sugars and ammonia) but during rapid growth it is often preferential to supply amino acids and other biochemicals in the growth medium to prevent limitation.

Growth media can be split into two broad categories: complex media, which contain components whose exact chemical composition is unknown, such as meat protein and yeast hydrolysates; and defined media, whose exact chemical composition is defined according to bacterial requirements. Generally speaking, complex media are cheaper and simpler to manufacture compared with defined media. Design of a defined medium requires an exact knowledge of the chemical requirements of the bacterium. Limitation of a component in defined medium will result in growth arrest; however, definition of limiting nutrients (especially micronutrients, such as vitamins or cofactors) can be time consuming. For these reasons, development of defined media is difficult and many media will contain one undefined component.

Many complex media contain hydrolysed protein mixtures from animal origin (e.g. tryptone or peptones) and, therefore, are unsuitable for use in the manufacture of human drugs. Yeast- or soybean-derived alternatives are available. In addition, complex media components of undefined composition can vary between batches, so that bacterial growth also varies between batches, an undesirable situation leading not only to decreases in biomass yield, but also differences in recombinant protein yield. Protein yield can also vary substantially if a strain is grown in a different medium; one could envisage initial optimisation experiments in a complex medium followed by production in defined media, but this could lead to changes in recombinant protein yield and folding state.

The other most-commonly addressed medium question is that of carbon and energy sources [34]. Glucose and glycerol are the most commonly used feedstocks. Glucose is the preferred energy

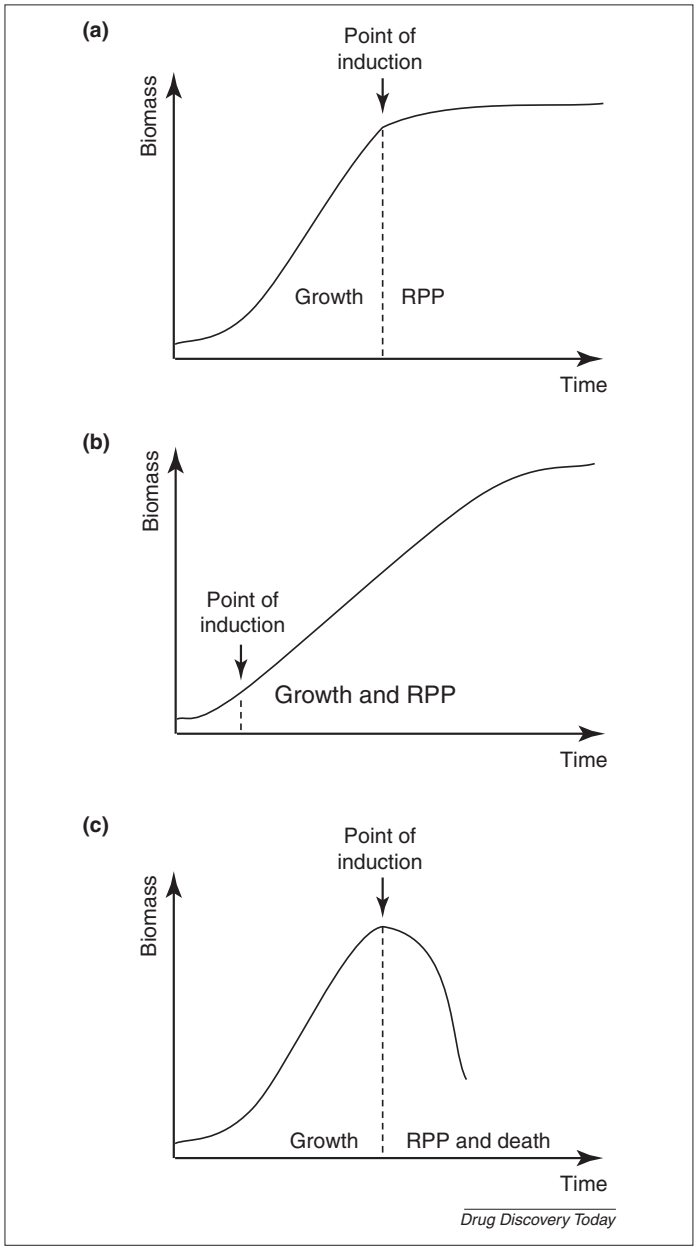

FIGURE 3

Growth and induction strategies. Induction at a high biomass **(a)** enables separation of growth and recombinant protein production (RPP) phases. Following induction, growth often slows as metabolic resources are channelled to protein production. Induction at low biomass **(b)** can only proceed if metabolic resources are evenly split between biomass accumulation of protein production; if this does not occur, then growth and/or recombinant protein production will be impaired. When 'toxic' proteins are made **(c)**, growth is severely inhibited or cell death occurs.

source for *E. coli* and can be metabolised via respiration aerobically and anaerobically (as long as an electron acceptor is supplied), as well as via fermentative pathways to mixed acids. This latter pathway is undesirable because it leads to both lower biomass and energy yields, and lowers the pH of the growth medium, thus repressing growth. Glucose can also be metabolised to acids via overflow metabolism, where flux through glycolysis exceeds that through the TCA cycle, leading to acid generation. Overflow metabolism proceeds when the glucose concentration in the medium exceeds a certain concentration; for this reason, careful control of glucose concentration must be maintained by use of fed-batch growth regimens [35]. Although more expensive and generating a lower biomass yield than glucose, glycerol cannot be readily fermented by *E. coli* and so does not generate acid as a by-product; therefore, it is the preferred carbon source for many processes.

Recombinant protein production can either be induced following biomass accumulation (by inducing protein production at a high biomass concentration), or allowed to proceed concurrently with bacterial growth by inducing protein production at a low biomass concentration (Fig. 3). Both strategies have found success in different processes; generally speaking, proteins that are known to inhibit growth are synthesised after biomass accumulation (Fig. 3a), whereas proteins whose synthesis does not cause growth inhibition can be made alongside growth (Fig. 3b). The possible uses of each method are discussed below.

Potential problems and solutions: growth-rate depression, zero productivity and recombinant protein 'toxicity'

A central part of the design and development of recombinant protein production methods is anticipating and solving problems. Few recombinant proteins can be synthesised perfectly on the first attempt and, therefore, some form of optimisation and problem solving must be undertaken. Here, I outline common problems and discuss possible solutions. Some commonly encountered problems are described in Table 2. Problems can be separated into two broad groups: metabolic problems caused by the process of protein production; and cellular responses to the recombinant protein itself.

From a metabolic level, recombinant protein production represents a mode of growth that *E. coli* would never encounter in its natural habitat, the mammalian gut; that is, a high growth rate under conditions of high nutrient concentration coupled with the generation of a single protein at high levels, comprising up to 50% of the total cellular protein [36]. Given this reasoning, it is not surprising that recombinant protein production can fail. Upon induction of protein production, growth rates typically fall

TABLE 2

Problems that can occur during recombinant protein production in bacteria

Level of process	Problem encountered	Diagnostic method	Comments and possible solutions
Cell viability	Extensive cell death throughout growth	Monitoring biomass during growth	A general sign of stress or process failure, which might indicate a 'toxic' protein. Possibly remedied by decreasing promoter basal expression, lowering plasmid copy number and, thus, gene dose, lowering temperature or lowering gene expression by using a lower inducer concentration or weaker promoter
	Extensive cell death on induction of protein production	Monitoring biomass during growth	
DNA stability	Loss of expression plasmid	Replica plating	Plasmid loss frequently suggests that host cells are experiencing stress or metabolic burden. Possibly remedied by lowering plasmid copy number and, thus, gene dose, lowering temperature, lowering inducer concentration or changing the antibiotic resistance from ampicillin
Transcription	mRNA degradation	Northern blot, RT-PCR or qPCR[a]	mRNA stability can be improved by removing sequence features that trigger degradation. Can also be a sign of rare codons
Translation	Translation stalls owing to rare codons, leading to truncated proteins	Bioinformatic analysis of codon bias of recombinant gene	Codon-optimise recombinant gene for host cell or provide rare codons on additional plasmid
Protein folding	Protein misfolding leading to low protein activity	Protein activity assay	Protein folding can be enhanced by lower growth temperature, slower gene expression by use of weaker promoter or lower inducer concentration, coexpression of chaperones, choice of specialised host strain or use of a solubilisation fusion tag
	IB formation	Separation of soluble and insoluble cellular proteins	
	Proteolysis	SDS-PAGE[b] analysis	Proteolysis can be decreased in host strains with deletions in protease genes (e.g. BL21) or by targeting protein to periplasm (which has fewer proteases than cytoplasm)
	Failure to form disulphide bonds	SDS-PAGE analysis	Target recombinant protein to periplasm or use a host strain capable of forming DSB cytoplasmically
Protein translocation	Protein fails to translocate to periplasm in significant quantities	Subcellular fractionation	Translocation can often be a rate-limiting step; therefore, the rate of protein production could be decreased by lowering inducer concentration or decreasing growth temperature

[a] *Abbreviations*: IB, inclusion body; qPCR, quantitative PCR; RT-PCR, reverse-transcriptase PCR; SDS-PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis.

(Fig. 3a); in many cases, this is a simple case of metabolic economics, whereby cellular energy (in the form of ATP), reducing power (NADH) and metabolites (especially amino acids) need to be shared between not only biomass accumulation, but also production of the recombinant protein. This competition can lead to stress responses, such as the stringent response, a reaction to limitations in intracellular amino acid concentration [37], and the general stress response mediated by the RNA polymerase sigma S subunit (RpoS) [38]. Both of these responses have the capacity to further decrease both growth and recombinant protein production [39]. Amino acid shortages can be remedied by supplementation, but more general metabolic issues, such as ATP shortage, need to be counteracted by decreasing the rate of recombinant protein production. This can be achieved by using a lower concentration of inducer, decreasing plasmid copy number or using a weaker promoter.

More extreme than a simple drop in growth rate upon induction is the phenomenon of 'toxic' recombinant proteins, which has been encountered on numerous occasions: upon transformation with the plasmid containing the recombinant gene, bacteria will either not grow at all, or will grow poorly, and the recombinant protein of interest is not made at all [40]. In many cases, the plasmid encoding the recombinant protein will be lost from the bacteria. Although commonly referred to as 'toxic' recombinant proteins, further investigation reveals several potential solutions, depending upon the root cause of the problem (Table 2). Typically, analysis methods will include monitoring the quantity of expression vector and mRNA transcript corresponding to the recombinant protein in the bacteria. This can reveal the point in the production process (Fig. 1) at which productivity is lost. In some cases, the problem is not toxicity caused by the recombinant protein but metabolic problems generated by the transcription and translation process. Again, minimisation of stress and lowering the rate of recombinant protein synthesis by decreasing inducer concentration and/or temperature can often aid productivity [41]. An alternative generic approach to improving production of 'toxic' proteins is absolute separation of biomass production and recombinant protein production phases (Fig. 3c), typified by decreasing basal promoter activity and lowering plasmid copy number and, thus, gene dose. In such cases, once protein production is induced, cell death can occur quickly, so the point of harvest must be carefully defined as to maximise recombinant protein yield.

One potential problem with producing eukaryotic proteins in bacteria is of differing codon usage. Given that the genetic code is degenerate, many amino acids are encoded by multiple codons (up to six). Each organism preferentially utilises certain codons; therefore, a commonly used codon for arginine in humans (AGG) is rarely used in *E. coli* (for only 2% of arginine codons). Thus, expression of a human gene containing multiple AGG codons would rapidly deplete the low level of arginine tRNAs corresponding to the AGG codon, stalling translation and, thus, stopping recombinant protein production, leading to truncated proteins and potentially affecting mRNA stability.

Two solutions are available for this problem [42]. One could provide rare tRNAs on a plasmid or chromosomal insertion to supplement the small pools of aminoacyl-tRNAs corresponding to rare codons. Such plasmids are available commercially, such as those in the Novagen Rosetta™ strains. Alternatively, the recombinant gene could be optimised to replace rare codons with those more commonly encountered in *E. coli*. This was historically a more complicated route, because it could require modification of dozens of nucleotides via site-directed mutagenesis, but the popularity of chemical DNA synthesis routes where genes can be made to order means that codons can be optimised *in silico* before the optimised gene is manufactured and cloned into the vector.

Potential problems and solutions: protein folding and misfolding

Even if the recombinant protein is generated, there is no guarantee that it will fold correctly, owing in part to the inherent differences between the chemical environment in bacteria and the endoplasmic reticulum of eukaryotic cells. Frequently, if a recombinant protein misfolds, it will form inclusion bodies (IBs), dense particles comprising unfolded and partially folded proteins in varying proportion [43]. Although first thought to contain entirely non-functional protein, recent studies have demonstrated that IBs can contain some entrapped functional protein (e.g. GFP [44]). IBs can easily be separated from other bacterial components following growth by cell lysis and centrifugal separation (relying upon their high density), so can be a source of relatively pure recombinant protein [43]. However, to become functional, IB proteins must be fully denatured (e.g. using urea or guanidinium hydrochloride) then refolded to a functional form following removal of the denaturant. In some cases, this offers a good production and purification route; Retavase® (Cornerstone Therapeutics Inc.), a derivative of human tissue plasminogen activator used for treatment of acute myocardial infarction, is generated in this way. However, although denaturing IBs is a simple process, refolding proteins to a functional form is protein dependent. Extreme care must be taken when using IB expression and refold pathways that the refolded protein is homogenously and correctly folded, and folded in the same manner as the native protein. The refold route is impossible for most recombinant proteins, therefore IB formation must be avoided and enhancing recombinant protein folding becomes a major objective [45]. This is especially true for the production of recombinant proteins for structural determination.

There has been great interest in improving recombinant protein folding by the manipulation of the bacterial heat shock response. This is a natural response in bacteria, triggered by the detection of misfolded proteins in the cytoplasm and periplasm, and mediated via signalling networks (reviewed in [46,47]). This results in the synthesis of two classes of protein, proteases and chaperones [48]. Generally speaking, proteases are responsible for degrading misfolded proteins, thus improving overall protein quality, whereas chaperones function by assisting correct folding of newly synthesised polypeptides or attempting to refold misfolded proteins. It could be hypothesised that recombinant polypeptides misfold because they are synthesised at a rate faster than the rate of chaperone-mediated folding, effectively overloading the chaperone-mediated folding pathway; therefore, increasing the quantity of chaperones in the bacterium could improve folding.

Martínez-Alonso *et al.* reviewed this field [49] and summarised the current state of the heat shock manipulation strategy. The hypothesis that increasing the level of cellular chaperones would improve protein folding is only true for certain recombinant

proteins; for others, decreasing chaperone levels improves folding, and increasing chaperone levels decreases protein-folding quality. The only certain generic conclusion from this body of work is that modification of chaperone and/or protease expression in host cells needs to be optimised for each recombinant protein, and cannot currently be predicted on the basis of recombinant protein primary structure or function. Again, this is an area of process optimisation that requires a parallel, trial-and-error approach.

A more generic method of improving recombinant protein folding is that of stress minimisation. This method relies upon generating the recombinant polypeptide more slowly, so that it can fold at a rate commensurate with the bacterial folding apparatus and not overload the chaperones. This strategy has been used in several different ways, each of which decreases the rate of polypeptide synthesis to the advantage of overall protein folding and yield. The high expression rates generated by the pET system frequently lead to decreases in host viability. Miroux and Walker [50] isolated mutant BL21 host strains that were able to survive the production of a recombinant protein that caused the death of most host bacteria; these mutants were later found to contain mutations in the *lac* promoter that controls expression of the T7 RNA polymerase, thus naturally decreasing the rate of recombinant gene expression [51]. These strains (named OverExpress™) are available commercially. This strategy was similarly used by Alfasi *et al*. [52] to isolate hosts for the generation of readily misfolding recombinant proteins using fluorescence-activated cell sorting (FACS) as a screening tool.

As well as molecular biology approaches to stress minimisation, culture conditions can also be modified to slow polypeptide production rates, enabling enhanced protein folding to occur. Decreased growth temperatures have previously been shown to improve overall yields of correctly folded protein as well as improving cell viability [53]. It was later shown that decreasing growth temperature and lowering inducer concentration, thereby decreasing the rate of transcription of the recombinant gene, enabled bacteria to accumulate greater yields of recombinant protein in a correctly folded form, both in small-scale batch cultures and high cell-density fed-batch fermentations [41]. Furthermore, a greater proportion of bacteria in the culture were viable and plasmid positive. Some of this improvement is probably also because of the decreased metabolic burden imposed on bacteria following induction with lower concentrations of inducer, as discussed above. These methods were utilised to produce correctly folded recombinant proteins that were either impossible to generate in a correctly folded form or that led to the death of the host bacteria using conventional growth approaches (e.g. 37°C and high inducer concentration). Although more time consuming, these stress minimisation methods prove that speed is not always the primary factor in process development; product yield and activity are crucial.

Potential solutions: tags for solubilisation, translocation and purification

As well as modifications to host strain and process conditions, the expression vector can be modified with tags to enhance recombinant protein productivity and folding. The nucleotide sequences encoding these tags are fused directly to the recombinant gene that is being synthesised, at either the N or C terminus, such that, when transcribed and translated, the recombinant protein and tag are generated as a single polypeptide chain. These tags enable recombinant proteins to be purified more simply, to be fused to additional proteins or to be targeted to different cellular compartments. Commercial expression vectors are available that contain such tags.

Affinity tags are a commonly used mechanism to enable rapid isolation of recombinant proteins from bacteria [54,55]. The commonly used His-tag comprises several histidine residues (usually six) attached to either the N or C terminus of the recombinant protein. Once translated, these histidine residues allow complexation of metal ions, such as nickel(II) or copper(II). By use of a chromatography resin containing these metal ions, the recombinant protein can be specifically separated from other cellular components in a process called immobilised metal affinity chromatography (IMAC; reviewed in [56]). This can be done on a high-throughput basis, enabling parallel purification of many proteins. Affinity tags are also useful for the detection of recombinant proteins using commercially available antibodies in ELISA or immunoblotting assays. When using purification tags for protein drugs, it is essential to ensure that the presence of the tag does not affect protein structure, folding or activity. Affinity tags are not usually utilised for the production of protein drugs for human use, owing to the requirement to cleave the tag before administration, thus increasing the complexity of downstream processing [54]. For this reason, affinity tags are frequently used for rapid generation of small quantities of multiple proteins for early stages of drug screening and in other research and development settings.

Solubilisation tags are a class of peptide sequences that readily fold into soluble proteins; when fused to recombinant proteins, they often enhance solubility [45,57]. Common examples are the NusA tag (a 54.8 kDa *E. coli* protein) and the maltose-binding protein (MBP, a 40 kDa *E. coli* protein). MBP is also an affinity tag, enabling protein purification using amylose affinity chromatography. However, production of these fusions potentially provides host cells with an additional metabolic burden over and above that imposed by the production of the recombinant protein alone. There is also a possibility that the solubility tag will fold correctly into a soluble form whereas the recombinant protein will remain misfolded. As with affinity tags, solubilisation tags must be cleaved from the recombinant protein before use; upon proteolytic cleavage from the recombinant protein of interest, the latter can spontaneously misfold.

The standard location of recombinant protein accumulation in *E. coli* is the cytoplasm; however, this is not the ideal environment for all recombinant proteins. The cytoplasm is a reducing environment and contains high levels of proteases, which can degrade recombinant proteins. In addition, release of recombinant protein from the cytoplasm requires cell breakage, which is an energy-intensive process that can generate heat, and results in a complex mixture of cellular components (proteins, nucleic acids and membrane components, including endotoxins, lipids and cell fragments) from which the recombinant protein must be purified before use in drug testing or clinical settings.

An alternative route is transport of recombinant protein into the periplasm, the region between the cytoplasmic and outer membranes of Gram-negative bacteria. This is usually achieved by attachment of a signal sequence to the N terminus of the

recombinant protein, which directs the recombinant protein to the Sec protein translocation apparatus, transporting the protein across the cytoplasmic membrane (reviewed in [58]). There are several potential advantages to this approach. The periplasm contains fewer proteases than the cytoplasm, thus recombinant proteins can accumulate with a decreased risk of proteolysis. Periplasmic targeting also enables the relatively simple release of the recombinant protein by stripping away the outer membrane; this is often achieved using osmotic or mild heat treatment [59,60]; therefore, it does not require the extensive input of energy required to break cells for release of cytoplasmic proteins. Such release can also make purification of the recombinant protein simpler, because the periplasm contains fewer proteins than the cytoplasm, is of lower volume and does not contain DNA, which can interfere with purification steps.

Periplasmic targeting is commonly used for the production of recombinant proteins that require DSBs for function, because the periplasm is an oxidising environment and the *E. coli* DSB chaperones are located in the periplasm (reviewed in [61]). The classic biopharmaceutical example is antibody fragments, such as Fab and single-chain variable fragment (scFv) fragments, which are often periplasmically targeted. Monoclonal antibodies are a popular class of biopharmaceuticals owing to their high specificity, but their production requires mammalian cell systems, making the process and, thus, the products expensive. Antibody fragments can be generated more cheaply because they can be made in bacteria, and have greater tissue penetration and mobility compared with full-length antibodies. Lack of an Fc region leads to shorter half-life compared with full-length antibodies, but this can be remedied by attachment of polyethylene glycol (PEG) moieties [62]. The Crohn's disease and rheumatoid arthritis treatment Cimzia® (UCB) is an anti-tumour necrosis factor (TNF)-α Fab' fragment made in *E. coli* and conjugated to PEG.

E. coli strains that are able to form DSB in the cytoplasm have also been developed, with mutations in genes involved in thioredoxin and glutathione reduction, giving rise to a more oxidising cytoplasm [63]. Such strains are available commercially from Novagen under the trade name Origami™.

In most periplasmic expression systems, polypeptide chains pass from the cytoplasm to the periplasm via the Sec apparatus in an unfolded form, one amino acid at a time [64]. This can be a drawback; if a recombinant protein folds completely or partially in the cytoplasm before reaching Sec, it might not be translocated. In addition, recombinant proteins might be unable to fold correctly in the periplasm and, therefore, can form periplasmic inclusion bodies. Recently, an alternative system for periplasmic targeting has been developed, based on the twin-arginine translocation (TAT) system, which enables translocation of fully folded proteins [65].

An exciting area of future development for bacterial hosts is true secretion of recombinant proteins [66]. Although common in eukaryotic hosts, such as Chinese hamster ovary (CHO) cells and *Pichia pastoris*, Gram-negative bacteria are generally poor hosts for the secretion of proteins into the extracellular milieu. Bacterial protein secretion systems do exist (reviewed in [67]), but they are typically used by pathogenic bacteria to secrete proteins required for pathogenesis. One example of this is injection of protein toxins into host intestinal cells by *E. coli* O157, which causes severe diarrhoea and can lead to kidney failure. However, groups are starting to investigate the opportunities to utilise these secretion mechanisms to transport recombinant proteins out of bacteria into the extracellular milieu, where protease activity is low and purification is theoretically simpler. This would also bring *E. coli* recombinant protein production processes in line with CHO cell processes in terms of product recovery. One such secretion system is the *E. coli* autotransporter system, which has been developed to enable the export of recombinant proteins [68].

Concluding remarks

Given the advantages discussed above, bacterial hosts are desirable systems for the production of recombinant proteins, and much work is currently being done to extend their application by enabling more efficient protein synthesis, folding and release. Although it is unlikely that bacteria will be able to generate any protein that is desired as a biopharmaceutical drug molecule or drug target, advances have been made in the synthesis of functional proteins that were previously hard (or impossible) to manufacture. High-throughput screening approaches for expression optimisation have enabled an acceleration of protein production and, thus, development of both biopharmaceutical and small-molecule drugs. Likewise, intensification of growth and optimisation of protein yield and folding have delivered ever-higher yields of functional protein in shorter timeframes. Advances in bacterial post-translational modification and protein release systems (such as periplasmic release and secretion systems) will enable further improvements to be made by simplifying protein production processes, enabling even more rapid synthesis of target proteins and, thus, further accelerating drug development.

Acknowledgements

The author would like to thank Chris Wyre and Tania Selas Castiñeiras for helpful comments on the manuscript.

References

1 Walsh, G. (2010) Biopharmaceutical benchmarks. *Nat. Biotechnol.* 28, 917–924
2 Walsh, G. (2012) New biopharmaceuticals. *BioPharm. Int.* 25, 34–38
3 Gordon, E. *et al.* (2008) Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr. Purif.* 62, 1–8
4 Medicines and Healthcare Products Regulatory Agency, (2007) *Rules and Guidance for Pharmaceutical Manufacturers and Distributors*. Pharmaceutical Press
5 Walsh, G. (2010) Post-translational modifications of protein biopharmaceuticals. *Drug Discov. Today* 15, 773–780
6 Collet, J.F. and Bardwell, J.C. (2002) Oxidative protein folding in bacteria. *Mol. Microbiol.* 44, 1–8
7 Nothaft, H. and Szymanski, C.M. (2013) Bacterial protein N-glycosylation: new perspectives and applications. *J. Biol. Chem.* 288, 6912–6920
8 Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 72, 211–222
9 Huang, C.J. *et al.* (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* 39, 383–399
10 Blattner, F.R. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462

11 Green, M.R. and Sambrook, J., eds) (2012) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press
12 Itakura, K. *et al.* (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056–1063
13 Johnson, I.S. (1983) Human insulin from recombinant DNA technology. *Science* 219, 632–637
14 Ferrer-Miralles, N. *et al.* (2009) Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* 8, 17
15 Pohl, S. and Harwood, C.R. (2010) Heterologous protein secretion by bacillus species from the cradle to the grave. *Adv. Appl. Microbiol.* 73, 1–25
16 Chen, R. (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol. Adv.* 30, 1102–1107
17 Morello, E. *et al.* (2008) *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* 14, 48–58
18 Balasundaram, B. *et al.* (2009) Advances in product release strategies and impact on bioprocess design. *Trends Biotechnol.* 27, 477–485
19 Peti, W. and Page, R. (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr. Purif.* 51, 1–10
20 Lee, S.Y. (1996) High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* 14, 98–105
21 Guiochon, G. and Beaver, L.A. (2011) Separation science is the key to successful biopharmaceuticals. *J. Chromatogr. A* 1218, 8836–8858
22 Carta, G. and Jungbauer, A. (2010) *Protein Chromatography: Process Development and Scale-Up*. John Wiley & Sons
23 Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37–44
24 Reznikoff, W.S. (1992) The lactose operon-controlling elements: a complex paradigm. *Mol. Microbiol.* 6, 2419–2422
25 Peränen, J. *et al.* (1996) T7 vectors with modified T7*lac* promoter for expression of proteins in *Escherichia coli*. *Anal. Biochem.* 236, 371–373
26 Brautaset, T. *et al.* (2009) Positively regulated bacterial expression systems. *Microb. Biotechnol.* 2, 15–30
27 World Health Organization, (2012) *The Evolving Threat of Antimicrobial Resistance: Options for Action*. World Health Organisation
28 Peubez, I. *et al.* (2010) Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microb. Cell Fact.* 9, 65
29 Cranenburgh, R.M. *et al.* (2001) Escherichia coli strains that allow antibiotic-free plasmid selection and maintenance by repressor titration. *Nucleic Acids Res.* 29, E26
30 Summers, D.K. (1996) *The Biology of Plasmids*. Blackwell Science
31 Daegelen, P. *et al.* (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J. Mol. Biol.* 394, 634–643
32 Waegeman, H. (2013) Increasing recombinant protein production in *Escherichia coli* K12 through metabolic engineering. *Nat. Biotechnol.* 30, 255–261
33 Phue, J.N. *et al.* (2005) Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. *Biotechnol. Bioeng.* 90, 805–820
34 Cote, R. and Gherna, R.L. (1999) Medium formulation and design, *E. coli* and *Bacillus* spp. In *Encyclopedia of Bioprocess Technology—Fermentation, Biocatalysis, and Bioseparation* (Flickinger, M.C. and Drew, S.W., eds), pp. 1676–1683, John Wiley & Sons
35 Shiloach, J. and Fass, R. (2005) Growing *E. coli* to high cell density: a historical perspective on method development. *Biotechnol. Adv.* 23, 345–357
36 Jevsevar, S. *et al.* (2005) Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnol. Prog.* 21, 632–639
37 Jin, D.J. *et al.* (2012) Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.* 36, 269–287
38 Battesti, A. *et al.* (2011) The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65, 189–213
39 Hoffmann, F. and Rinas, U. (2004) Stress induced by recombinant protein production in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.* 89, 73–92
40 Saïda, F. *et al.* (2006) Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr. Protein Pept. Sci.* 7, 47–56
41 Sevastsyanovich, Y. *et al.* (2009) Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins. *FEMS Microbiol. Lett.* 299, 86–94
42 Burgess-Brown, N.A. *et al.* (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. *Protein Expr. Purif.* 59, 94–102
43 Singh, S.M. and Panda, A.K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* 99, 303–310
44 García-Fruitós, E. *et al.* (2007) Localization of functional polypeptides in bacterial inclusion bodies. *Appl. Environ. Microbiol.* 73, 289–294
45 Sørensen, H.P. and Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* 4, 1
46 Guisbert, E. *et al.* (2008) Convergence of molecular, modelling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. *Microbiol. Mol. Biol. Rev.* 72, 545–554
47 Ades, S.E. (2008) Regulation by destruction: design of the σE envelope stress response. *Curr. Opin. Microbiol.* 11, 535–540
48 Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 22, 1399–1401
49 Martínez-Alonso, M. *et al.* (2010) Side effects of chaperone gene co-expression in recombinant protein production. *Microb. Cell Fact.* 9, 64
50 Miroux, B. and Walker, J.E. (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
51 Wagner, S. *et al.* (2008) Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14371–14376
52 Alfasi, S. *et al.* (2011) Use of GFP fusions for the isolation of *Escherichia coli* strains for improved production of different target recombinant proteins. *J. Biotechnol.* 156, 11–21
53 Vera, A. *et al.* (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.* 96, 1101–1106
54 Arnau, J. *et al.* (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* 48, 1–13
55 Waugh, D.S. (2005) Making the most of affinity tags. *Trends Biotechnol.* 23, 316–320
56 Cheung, R.C. *et al.* (2012) Immobilized metal ion affinity chromatography: a review on its applications. *Appl. Microbiol. Biotechnol.* 96, 1411–1420
57 Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533
58 Natale, P. *et al.* (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-distinct translocases and mechanisms. *Biochim. Biophys. Acta* 1778, 1735–1756
59 French, C. *et al.* (1996) Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm. *Enzyme Microb. Tech.* 19, 332–338
60 Popplewell, A.G. *et al.* (2005) Expression of antibody fragments by periplasmic secretion in *Escherichia coli*. *Methods Mol. Biol.* 308, 17–30
61 de Marco, A. (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* 8, 26
62 Chapman, A.P. (2002) PEGylated antibodies and antibody fragments for improved therapy: a review. *Adv. Drug Deliv. Rev.* 54, 531–545
63 Stewart, E.J. *et al.* (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.* 17, 5543–5550
64 Georgiou, G. and Segatori, L. (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr. Opin. Biotechnol.* 16, 538–545
65 Matos, C.F. *et al.* (2012) High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in *Escherichia coli*. *Biotechnol. Bioeng.* 109, 2533–2542
66 Mergulhão, F.J. *et al.* (2005) Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* 23, 177–202
67 Henderson, I.R. *et al.* (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* 68, 692–744
68 Sevastsyanovich, Y.R. *et al.* (2012) A generalised module for the selective extracellular accumulation of recombinant proteins. *Microb. Cell Fact.* 11, 69
69 Kim, J.Y. *et al.* (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* 93, 917–930
70 Mattanovich, D. *et al.* (2012) Recombinant protein production in yeasts. *Methods Mol. Biol.* 824, 329–538
71 Gronenborn, B. (1976) Overproduction of phage lambda repressor under control of the *lac* promotor of *Escherichia coli*. *Mol. Gen. Genet.* 148, 243–250
72 Brosius, J. *et al.* (1985) Spacing of the −10 and −35 regions in the *tac* promoter. Effect on its in vivo activity. *J. Biol. Chem.* 260, 3539–3541
73 Guzman, L.M. *et al.* (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130
74 Elvin, C.M. *et al.* (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* 87, 123–126