J Ind Microbiol Biotechnol (2011) 38:1891–1910
DOI 10.1007/s10295-011-1034-4

REVIEW

Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering

Hendrik Waegeman · Wim Soetaert

Received: 27 June 2011 / Accepted: 20 August 2011 / Published online: 8 September 2011

Abstract Different hosts have been used for recombinant protein production, ranging from simple bacteria, such as *Escherichia coli* and *Bacillus subtilis*, to more advanced eukaryotes as *Saccharomyces cerevisiae* and *Pichia pastoris*, to very complex insect and animal cells. All have their advantages and drawbacks and not one seems to be the perfect host for all purposes. In this review we compare the characteristics of all hosts used in commercial applications of recombinant protein production, both in the area of biopharmaceuticals and industrial enzymes. Although the bacterium *E. coli* remains a very often used organism, several drawbacks limit its possibility to be the first-choice host. Furthermore, we show what *E. coli* strains are typically used in high cell density cultivations and compare their genetic and physiological differences. In addition, we summarize the research efforts that have been done to improve yields of heterologous protein in *E. coli*, to reduce acetate formation, to secrete the recombinant protein into the periplasm or extracellular milieu, and to perform post-translational modifications. We conclude that great progress has been made in the incorporation of eukaryotic features into *E. coli*, which might allow the bacterium to regain its first-choice status, on the condition that these research efforts continue to gain momentum.

Keywords Heterologous protein · Recombinant protein production · *Escherichia coli* · Engineering · Biopharmaceutical · Industrial enzyme

H. Waegeman (✉) · W. Soetaert
Centre of Expertise–Industrial Biotechnology and Biocatalysis, Department of Biochemical and Microbial Technology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
e-mail: Hendrik.Waegeman@UGent.be

Introduction

The development of recombinant DNA technology in the early 1970s [31] enabled the expression of heterologous genes into pro- or eukaryotic hosts which do not naturally harbor these pieces of DNA. As a result, an organism can produce a protein which it does not produce in natural circumstances, and the protein can originate from a totally different, sometimes not even closely related species. The ability to produce recombinant proteins led to a tremendous rise in biopharmaceutical and industrial enzyme applications.

The first biopharmaceutical produced by recombinant DNA technology was human insulin, a medicine used for the treatment of diabetes [60]. After approval by the US Food and Drug Administration (FDA) in 1982, it became the first biopharmaceutical compound of its kind to enter the market. Since then, the number of applications for the clinical use of biopharmaceuticals has increased steadily. To date, the application of more than 150 recombinant proteins has been licensed by the FDA and European Medicines Agency (EMEA) to be used as a pharmaceutical [52]. The global market for biopharmaceuticals is estimated to be worth approximately €50–60 billion[1] [174]. Moreover, in the upcoming years, this market is expected to grow at a rate of approximately 9% annually [61].

Recombinant DNA technology also facilitated the expansion of the industrial enzyme market. Before the 1980s, the vast majority of enzymes were still naturally derived from plants and animals, resulting in low yields and concomitant high prices. By cloning the plant and animal genes into appropriate expression vectors, the

[1] However this value can vary greatly, depending on how a biopharmaceutical is defined and the sources consulted.

enzymes could be produced in much larger amounts by microbial fermentation, often circumventing the laborious downstream processing that was associated with conventional extraction methods. For example, for over hundreds of years, the enzyme chymosin or rennin has been used as a coagulation agent of milk proteins in the production of cheese. Until 1990, chymosin was isolated from the abomasum or fourth stomach of the calf and lamb. Today, it is produced by genetically engineered strains, such as *Escherichia coli*, *Aspergillus niger*, *Kluyveromyces lactis*, or *Pichia pastoris* [96]. Industrial enzymes are used in various industrial segments. Proteases, amylases, lipases, and cellulases are used for stain removal and color clarification in the detergent industry. Amylases, amyloglucosidases, and pullulanases are applied for starch liquefaction and saccharification. Glucose can be converted to the sweeter-tasting fructose by glucose isomerase. Pectinases are used in the clarification of fruit juices and many more examples exist in other industrial fields, e.g., the pulp and paper, textile, and leather industries [91].

The industrial enzyme market rose from €0.3 billion in 1994 to €1.4 billion in 2000 and €1.8 billion in 2009. By 1994, over 50% of the industrial enzyme market value was generated by recombinant processes [43, 74]. Taking into account the great progress that has been achieved since then, this number is probably even much higher today.

This review will describe what factors determine the choice of the host in recombinant bioprocesses, both in the domains of biopharmaceuticals and industrial enzymes. More specifically, we will focus on the use of *Escherichia coli* as a host and the different engineering strategies that have been applied to increase recombinant protein production (RPP) in this bacterium.

Plasmid-based RPP

In the first attempts to produce recombinant proteins, the gene of interest was typically cloned into an expression vector, generally a plasmid. Expression vectors are extrachromosomal self-replicating DNA elements that contain an origin of replication (ori), a selection marker (usually antibiotic resistance), transcriptional promoters, translation initiation regions (TIRs) as well as transcriptional and translational terminators [157]. To date, the vast majority of recombinant protein expression is plasmid-based, primarily because higher gene dosages can be obtained compared with when the recombinant gene is chromosomally integrated into the host's DNA, i.e., multiple copies of the plasmid are present in the cell, e.g., 15–20 in pBR322 and 500–700 in the pUC plasmid series. Furthermore, genetic manipulations are much more straightforward and less time consuming on plasmids than in the genome [124].

Nonetheless, plasmids can only be used in bacterial hosts and to a lesser extent in yeast or fungi. For higher eukaryotes, chromosomal integration of the gene remains the only solution. In addition, many research groups encountered a reduction in expression of the desired protein during long cultivations when using plasmid-based systems, which is caused by plasmid loss and overgrowth of plasmid-free bacteria [142] or loss of functional T7 RNA polymerase as a result of chromosomal mutations [171]. Moreover, plasmid maintenance imposes a metabolic burden on the host cells, which results in a reduced growth rate and viability [123]. Finally, fragments larger than 15 kb cannot be cloned into conventional plasmids [130].

To increase plasmid stability and evolutionary robustness of the strain, three basic design principles should be respected: (1) high expression of genetic circuits comes with the cost of low evolutionary stability, (2) repeated sequences should be avoided, and (3) the use of inducible promoters generally increases evolutionary stability [152]. Furthermore, to overcome plasmid instability problems or metabolic burden, the gene of interest can be chromosomally integrated. If only one gene copy is incorporated, recombinant protein yields remain rather low because of a very low gene dosage. However, when multiple copies are inserted, gene dosage is increased, which results in a linear increase in RPP [23, 135, 155]. This approach is much more labor-intensive than the plasmid-based system, but generally results in more stable production.

Alternatively, bacterial artificial chromosomes (BACs) can be used as expression vectors. BACs are derived from the F factor of *E. coli* and are capable of maintaining human genomic DNA fragments of more than 300-kb pairs [146]. The advantage compared with plasmids is that the cloning capacity is increased and compared with chromosomal integration that the expression is less affected by the surrounding chromatin of the insertion site in the host genome, which results in a more stable expression. Despite these enviable characteristics, BACs have been rarely used in RPP. One recent successful application involves the production of the Fc fragment of human immunoglobulin IgGI in human embryonic kidney cells (HEK 293 cell line) [12].

Hosts for RPP

Besides the choice between a chromosomal or plasmid-based approach, the selection of a suitable host strain is one of the most important aspects in the design of recombinant protein bioprocesses.

In the ideal case, a strain should grow rapidly on cheap media, it should be very easy to genetically modify, and should be able to perform difficult glycosylation reactions

or assist in disulfide bond formation. Furthermore, the recombinant protein should be easily extractable from the cell or the extracellular milieu in a minimum number of steps. It is also important that the strain is generally regarded as safe, because this status facilitates the applicability of the recombinant protein. Taking all these factors into account, this should result in a bioprocess that is economically superior to the competing process it wants to replace [47, 103, 138].

The strains that are most commonly used in the biopharmaceutical and industrial enzyme industries are displayed in Fig. 1, and their most important features are shown in Table 1. The main difference between the biopharmaceutical and the industrial enzyme market is that industrial enzymes must be produced at much lower cost than biopharmaceuticals. The former are typically produced at €7−14 per kilogram [166] and margins on sales are rather low, whereas biopharmaceuticals can be sold at prices that are typically 100–1,000 times higher [134]. This implies that yields and production rates can be lower, while economical viability is still guaranteed. Consequently, in the biopharmaceutical market, slower-growing organisms like plants or animals can be utilized.

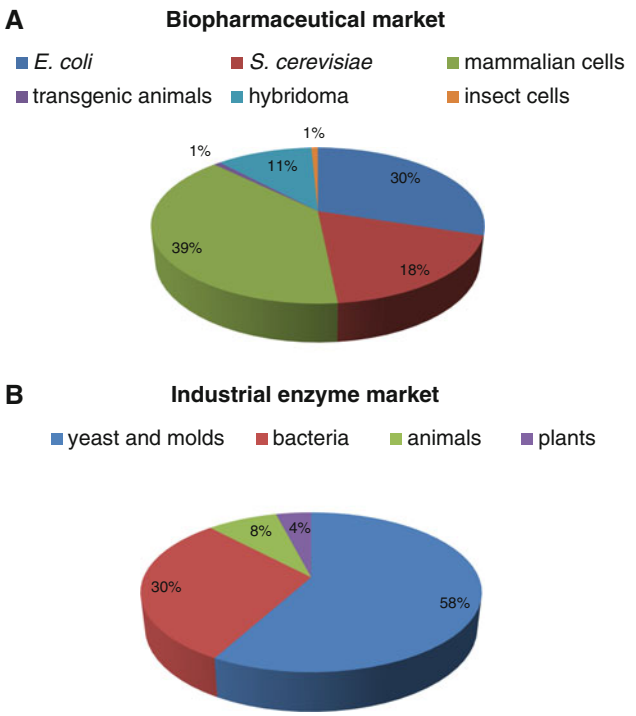

Fig. 1 Strains typically used in the biopharmaceutical and industrial enzyme market. The data for the biopharmaceutical market (**a**) are based on recombinant proteins approved as drugs by the FDA and EMEA [52]; the data for the strains used in the enzyme market (**b**) are adopted from [43]

Bacteria

Escherichia coli remains one of the most commonly used bacteria, primarily because of its well-characterized genetics, its ability to grow rapidly on cheap substrates, and the many molecular tools available. Nonetheless, *E. coli* suffers some major drawbacks as well. Many of the post-translational modifications that are found in higher organisms like glycosylation steps and disulfide bond formation cannot be performed, i.e., *E. coli* does not glycosylate its proteins, and disulfide bond formation, which is necessary for correct folding of many eukaryotic proteins, is obstructed by the reduced state of the *E. coli* cytoplasm. Another severe disadvantage is the difficulty of the strain to secrete the recombinant protein in the culture medium [7, 85, 106]. Even more problematic is that in many cases recombinant proteins end up in inclusion bodies, and therefore need to be refolded in a subsequent step [8].

Recently, in academia and industry, more attention has been devoted to the use of Gram-positive *Bacillus* strains for the production of recombinant proteins. *Bacillus* strains share a number of advantages with *E. coli*, but become more and more attractive because of their ability to secrete the recombinant protein into the culture medium in high amounts. Furthermore, *Bacillus* strains are Gram-positive bacteria and hence do not contain an outer membrane which consists of lipopolysaccharides (LPS). These structures often contain endotoxins which are pyrogenic to humans. Disadvantages and points of particular interest are plasmid instability, the availability of expression vectors, protease activity, and the difficulty of high cell density cultivation [129]. The most often used strains include *Bacillus megaterium*, *Bacillus subtilis*, and *Bacillus brevis* [161].

Other hosts that have been exploited for RPP include *Pseudomonas putida* for antibody fragments [34] and *Ralstonia eutropha* for organophosphohydrolase production [9]. Many more bacterial systems are described, but more than often the limited information about their genetics or metabolism or the unavailability of expression vectors or promoter systems hinders the expansion of their application.

Yeast

To date, yeast strains have been extensively applied as hosts both for the production of biopharmaceuticals and industrial enzymes (see Fig. 1). As non-pathogenic organisms with a long history in agriculture and the food industry, yeasts have been the first choice hosts in many commercial applications [57].

The most often used strain is *Saccharomyces cerevisiae*, which like *E. coli* is a ‘pet’ organism in many laboratories.

Table 1 Features of strains used in RPP

Features	Bacteria	Yeasts	Filamentous fungi	Higher eukaryotes
Most relevant species	*E. coli*	*Saccharomyces cerevisiae*	*Aspergillus* strains	Chinese hamster ovary cells (CHO)
	Bacillus strains	*Pichia pastoris*	*Trichoderma* strains	
Applications	Biopharmaceuticals (mainly *E. coli*)	Mainly biopharmaceuticals	Mainly industrial enzymes	Mainly biopharmaceuticals
	Industrial enzymes (mainly *Bacillus*)	Some industrial enzymes		
Production rate	Very high	High	Moderate to high	Low
Substrate cost	Low	Low	Low	High
Secretion	Into periplasm (Gram-negative bacteria) or culture medium (Gram-positive bacteria)	Into culture medium	Into culture medium	Into culture medium
Genetic manipulation	Very easy	Easy	Rather easy	Difficult
Folding	Often problematic and refolding necessary	Refolding sometimes required	Refolding sometimes required	No folding problems encountered
Glycosylation of the protein	Rarely occurs	Possible, but glycoproteins typically contain high amounts of mannose units	Possible, better mimics human glycosylation than yeast, but does not contain many human sugar units	Approaches human-like glycosylation

It shares more or less the same advantages with *E. coli*, i.e., it is easy to cultivate and has well-characterized genetics and metabolism. Because it is a eukaryote, it can, however, perform some of the post-translational modifications that are typical of higher organisms.

Whereas *S. cerevisiae* was the most preferred yeast in the early days of RPP, other yeast strains are also being explored as hosts today. During the last two decades there has been a tremendous increase in the use of methylotrophic yeasts, such as *Hansenula polymorpha*, *Candida boidinii*, and especially *Pichia pastoris* [33, 76, 105]. The popularity of *Pichia* can be attributed to the easy implementation of genetic, biochemical, and molecular biological techniques that have been developed for *S. cerevisiae* over the past decades, while protein expression levels are often 10- to 100-fold higher [73].

In contrast to bacteria, yeast strains can perform eukaryotic glycosylation patterns, making them particularly suitable for the production of more complex, mammalian proteins. Glycosylation increases the stability and bioactivity of the foreign protein, and the better the recombinant glycosylation pattern mimics the native one, the more stable and active the protein seems to be [153]. Both *S. cerevisiae* and *P. pastoris* co- and post-translationally add oligosaccharide chains to the proteins and in both strains these sugars are mainly mannose units. However, *S. cerevisiae* tends to hyperglycosylate the protein, adding up to 150 mannose residues per side chain, whereas *P. pastoris* only adds 8–14 mannose residues on average [68]. Maybe a more important structural difference in glycosylation is that in contrast to *S. cerevisiae*, *P. pastoris* does not add α1,3-terminal mannose units to the oligosaccharide chains [33], and it was discovered that mannose-terminating glycans are responsible for breakdown in the blood and immunogenic responses in patients [35]. Although *Pichia*-derived proteins perform slightly better in terms of immunogenicity than those produced in *Saccharomyces*, the high mannose content and the absence of typically human intermediate and terminal sugar residues such as sialic acid remain bottlenecks in the production of safe and active mammalian therapeutic proteins. This explains the various attempts to engineer the glycosylation pattern of *Pichia* to obtain more human-like glycans [14, 19, 24, 69, 170].

Although *Pichia* has been extensively used as a host in research, e.g., more than 400 proteins have been produced in this yeast [20], the commercial applications are rather limited. In 2009 the first approval for a biopharmaceutical derived from *Pichia* was granted. It concerns Kalbitor® and is used for the treatment of acute attacks of hereditary angioedema,[2]

Filamentous fungi

Filamentous fungi are important mainly as hosts for the production of industrial enzymes, and they are rarely applied for the production of biopharmaceuticals. Of the

[2] http://www.biopharma.com/approvals.html consulted on 15 June 2011.

Table 2 The industrial enzyme market: type of proteins and strains used

Total number of enzymes	260
Enzymes that are homologous proteins	57%
Enzymes that are heterologous proteins	43%
Enzymes produced by *Aspergillus* strains	33%
Enzymes produced by *Trichoderma* strains	11%
Enzymes produced by *Penicillium* strains	5%
Enzymes produced by *Saccharomyces* strains	1%
Enzymes produced by *Kluyveromyces* strains	1%
Enzymes produced by *Bacillus* strains	18%
Enzymes produced by other bacterial strains	9%

Numbers are based on the Amfep list of commercial enzymes http://www.amfep.org from 2009. Numbers from 2004 can be found elsewhere [143]

260 enzymes listed by the Association of Manufacturers and Formulators of Enzyme Products (Amfep) in 2009,[3] about 50% were produced in filamentous fungi (Table 2). It is noteworthy that the largest group comprises homologous proteins (57%). Many filamentous fungi species naturally secrete large amounts of industrially relevant enzymes into the growth medium, explaining their wide use for enzyme production. Furthermore filamentous fungi grow relatively fast on inexpensive substrates and are GRAS organisms (Generally Recognized As Safe) [118]. The most important species are *Aspergillus*, *Trichoderma*, and *Penicillium*, which naturally secrete enzymes such as amylases, cellulases, lipases, and proteases.

In most engineering strategies, multiple copies of the gene of interest (homologous or heterologous) are introduced into a host strain that has a high natural ability to secrete proteins and that can be easily cultivated and genetically engineered. Furthermore, the use of heterologous gene products is steadily increasing. For example, in 2004, 35% of all enzymes were heterologous proteins [143], whereas in 2009 this number reached 43%.

Filamentous fungi have also been explored for the production of biopharmaceuticals; however, to date this has not resulted in any successful applications.

Higher eukaryotes

Although yeasts and filamentous fungi can grow relatively fast on inexpensive substrates, their inability to execute human-like glycosylation patterns obstructs their wide application in the biopharmaceutical industry. There have also been attempts to use higher eukaryotes, such as insects and plants, but in most cases with limited industrial success. As shown in Fig. 1, the majority of approved biopharmaceuticals are produced in mammalian cell systems, only because of their ability to perform complex human-like post-translational modifications (PTMs). Of all PTMs, glycosylation is by far the most notorious and most difficult to engineer. There are five types of glycosylation, in order of importance classified as N-, O-, P-, C-, and G-linked, all involving the addition of an oligosaccharide structure to the protein core but with different binding sites [125]. The most widely occurring type of glycosylation is N-linked and comprises the attachment of glycans to nitrogen of asparagine or arginine side chains [71, 125, 149, 176].

The differences in N-glycosylation patterns between the yeast *P. pastoris* and mammalian cells are displayed in Fig. 2. The first steps of N-linked glycosylation are executed in the endoplasmic reticulum (ER) and are similar for yeast and higher eukaryotes. These steps consist of the addition of two *N*-acetylglucosamine and eight mannose units to the protein. The resulting glycoprotein is transported to the Golgi apparatus for further protein processing. However, these later glycosylation steps completely differ between yeasts and mammalian cells. Whereas in yeast, additional mannose units are added to the oligosaccharide structure, in mammalian cells the opposite occurs—mannose units are cleaved off and galactose and sialyl units are added [18, 58].

Chinese hamster ovary (CHO) cells are by far the most commonly used higher eukaryotes for the production of glycoproteins. Almost all approved biopharmaceuticals produced in mammalian cell systems rely on the use of CHO cells [52]. Although CHO cells are maybe the best system to achieve human-like glycosylation patterns, there are also some major drawbacks associated with this host. Animal cells grow very slowly and need expensive serum-based media, resulting in a low productivity and, hence, in an economically harsh process. Moreover, correct glycosylation of the protein seems to be very sensitive towards culture parameters such as pH, dissolved oxygen and carbon dioxide, and medium component concentrations [78].

Escherichia coli strains used in RPP

During the last decade several organisms have been further explored as hosts for RPP. However, the easily cultivable and genetically well-known bacterium *E. coli* remains very popular in a lot of applications. One of the key advantages of *E. coli* is that the microorganism can be grown to high cell densities in appropriate bioreactors by high cell density cultivation (HCDC) techniques, which allow the production of high amounts of heterologous protein [99].

Throughout the years, various *E. coli* strains with different genotypes have been examined for their potential to

[3] http://www.amfep.org consulted on 17 June 2011.

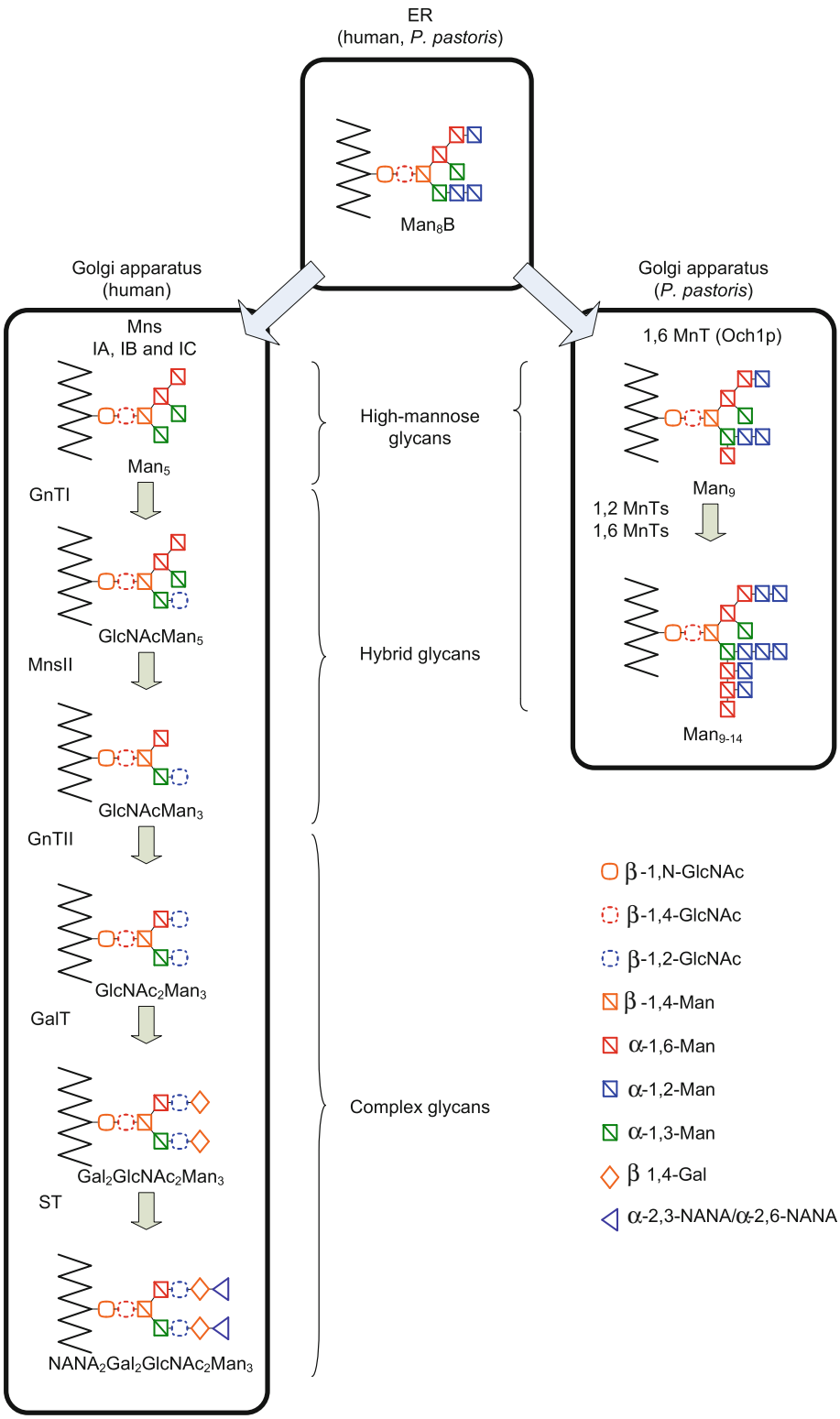

Fig. 2 Glycosylation steps in humans and *P. pastoris*. *Mns* α-1,2-mannosidase, *MnsII* mannosidase II, *GnTI* β-1,2-*N*-acetylglucosaminyltransferase I, *GnTII* β-1,2-*N*-acetylglucosaminyltransferase I, *GalT* β-1,4-galactosyltransferase I, *ST* sialyltransferase, *MnT* mannosyltransferase. Based on [58]

produce large amounts of these recombinant proteins. An overview of what strains are typically used in HCDCs is given in Table 3; the genotype characteristics of the strains are listed in Table 4. By far the most commonly used strain is *E. coli* BL21, with 15 of the 43 proteins being produced in this host. The primary reason for this is that BL21 strains are deficient in the proteases Lon and OmpT, which decreases the breakdown of recombinant protein and results in higher yields [62, 64]. Second, *E. coli* BL21 produces a substantially lower amount of acetate compared

Table 3 *E. coli* strains used in RPP processes

E. coli strain	Recombinant protein	Application	Titer ($g\ L^{-1}$)	Protein size (kDa)	Ref.
BL21 (DE3)	2-Deoxyribose-5-phosphate aldolase	Synthetic chemistry enzyme	5.12	±25	[127]
	Carboxylesterase B1	Hydrolyzes insecticide residues	NA[a]	64	[132]
	Insulin-like growth factor-2	Biopharmaceutical	9.69	6.5	[79]
	Human β-endorphin	Biopharmaceutical	5.6[b]	40[c]	[83]
	rh-GCSF[d]	Biopharmaceutical	1.75	18.8	[4]
	HPPCn[e]	Biopharmaceutical	0.64	30	[102]
	Human interferon-γ	Biopharmaceutical	4.2	25	[5]
	Human leptin	Biopharmaceutical	9.7	16	[82]
	Human mini-proinsulin	Biopharmaceutical	7	±5	[145]
	Human tumor necorsis factor-α	Biopharmaceutical	NA	±18	[128]
	Hydroxynitrile lyase	Chemical industry	NA[f]	±30	[140]
	Phytase	Feed additive	NA[g]	±50	[92]
	Resilin	Elastic protein for material industry	0.3	28	[89]
	Streptokinase	Biopharmaceutical	1.12	47	[67]
	Transglucosidase	Chemical industry	NA[h]	90[i]	[180]
BL21 (G2)	Carbamoylase	Chemical industry	NA[j]	±43	[21]
BLR (DE3)	Asparaginase	Biopharmaceutical	5.24	141	[88]
	Bioadhesive precursor (BP) protein	Biopharmaceutical	14[k]	26	[90]
DH5α	Mouse endostatin	Biopharmaceutical	0.04	20	[182]
	Protective antigen protein	Biopharmaceutical	0.125	83	[22]
C41 (DE3)	rHsp60[l]	Biopharmaceutical	0.0004	58	[133]
C600	TRAIL[m]	Biopharmaceutical	4.51	19	[104]
HB101	Alkaline phosphatase	Model enzyme	5.2	±48	[25]
	C8/119S[n]	Biopharmaceutical	NA[o]	15	[95]
	Insulin-like growth factor-I fused protein	Biopharmaceutical	1.24	15	[120]
	N-Carbamyl-D-amino acid amidohydrolase	Chemical industry	NA[p]	±33	[115]
HMS174 (DE3)	Human interleukin-7	Biopharmaceutical	NA[q]	17.4	[122]
JM101	Exoglucanase	Cellulose degradation	NA[r]	40	[55]
	hEGF[s]	Biopharmaceutical	0.325	6.2	[151]
MG1655	Aminolevulinate synthase	Insecticide, antimicrobial drug	5.2	±45	[181]
Origami (DE3)	AMA1[t]	Biopharmaceutical	NA[u]	±48	[48]
	Bacteriocin DvnRV41 fusion protein	Antimicrobial agent	0.296[v]	±20	[185]
	Human interferon α-4	Biopharmaceutical	NA[w]	18.4	[183]
	Pig liver esterase	Chemical industry	NA[x]	±60	[16]
Rosetta (DE3)	Glycine oxidase	Chemical industry	NA[y]	±41	[109]
	RADD[z]	Biopharmaceutical	0.068	±12	[179]
Rosetta-gami (DE3)	Insulin-like growth factor 1 fusion protein	Biopharmaceutical	2.1	24.8	[188]
TG1	rhBMP-2[aa]	Biopharmaceutical	8.6[ab]	±14	[164]
	Human interferon-α	Biopharmaceutical	4[ac]	4	[6]
	Phytolacca insularis protein	Biological pesticide	0.466	±48	[98]

Table 3 continued

E. coli strain	Recombinant protein	Application	Titer (g L^{-1})	Protein size (kDa)	Ref.
W3110	L-*N*-Carbamoylase	Chemical industry	3.8	±43	[177]
	Human insulin-like growth factor I fusion protein	Biopharmaceutical	4.3	±26	[27]
	Penicillin acylase	Semisynthetic antibiotic production	NA[ad]	95	[100]
	P64k, dihydrolipoamide dehydrogenase	Vaccine preparations	2.1	64	[50]

All production processes refer to HCDCs. When '±' is added to the protein size no exact value was mentioned in the cited reference and an estimation based on SDS–PAGE results or amino acid length is given

NA not available

[a] 99 U mL^{-1}

[b] As ompF fusion protein, ompF was cleaved by Factor Xa to yield β-endorphin

[c] As fusion protein, β-endorphin = 4 kDa

[d] Human granulocyte colony-stimulating factor

[e] Human hepatic growth factor hepatopoietin Cn

[f] 180 U mL^{-1}

[g] 120 U mL^{-1}

[h] 501 nkat g^{-1}

[i] As a fusion protein, truncated to *Pseudomonas syringae* ice nucleation protein

[j] 14 U mL^{-1}

[k] Assuming that 50% of the biomass is protein

[l] *S. enteritidis* recombinant heat shock protein 60

[m] Truncated tumor necrosis factor related apoptosis-inducing ligand

[n] Major house dust mite allergen der f2 mutant

[o] Produced in inclusion bodies, 2.3 g purified protein from 300 L culture

[p] 22.9 U mL^{-1}

[q] Produced in inclusion bodies, 10−12 g L^{-1} as S-200 purified IL-7 monomer

[r] 4.5 U mL^{-1}

[s] Human epidermal growth factor

[t] Apical membrane antigen 1

[u] 70 mg purified from 100 L culture

[v] As fusion protein

[w] 21.7% of total protein

[x] 1.4 U mL^{-1}

[y] 1.2 U mL^{-1}

[z] Recombinant human ADAM15 disintegrin domain

[aa] Recombinant human bone morphogenetic protein-2

[ab] 0.750 g L^{-1} after purification from inclusion bodies

[ac] 0.300 g L^{-1} after purification from inclusion bodies

[ad] 0.9 U mL^{-1}

with K12-derived strains [144]. Acetate inhibits growth and recombinant protein formation, even at low concentrations of 0.5–1 g L^{-1} [32, 114]. The use of the low-acetate producer BL21 therefore generally results in higher protein yields.

Moreover, many of the other *E. coli* strains used in RPP are BL21-derived strains. For example, *E. coli* BLR is a BL21 *recA* mutant strain. RecA is involved in DNA repair and homologous recombination [158], which implies that when RecA is active, it can provoke homologous recombination between chromosomal and plasmid DNA. Since most recombinant proteins are expressed from plasmids, plasmid stability can be a very important issue in large-scale production processes. Zhao et al. [189] investigated the effects of this *recA* mutation and observed that *E. coli* BLR exhibits a superior plasmid stability compared with its

Table 4 Genotype characteristics of *E. coli* strains used in RPP processes

E. coli strain	Chromosomal genotype	Comments
BL21 (DE3)	*E. coli* B F^-*dcm ompT hsdS*$_B$($r_B^-m_B^-$) *gal lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Contains λDE3 phage with IPTG-regulated T7 polymerase
BL21 (G2)	*E. coli* B F^-*dcm ompT hsdS*$_B$($r_B^-m_B^-$) *gal lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Thermoregulated instead of IPTG-regulated T7 polymerase
BLR (DE3)	*E. coli* B F^-*dcm ompT hsdS*$_B$($r_B^-m_B^-$) *gal lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])* Δ*(srl-recA)306*::Tn*10*(Tc^R)	BL21 *recA* derivative
DH5α	*E. coli* K12 F^- *fhuA2* Δ*(argF-lacZ)U169* Φ80 Δ *(lacZ)M15 gyrA96 phoA glnV44 recA1 relA1 endA1 thi-1 hsdR17*	Typical plasmid storage strain
C41 (DE3)	*E. coli* B F^-*dcm ompT hsdS*$_B$($r_B^-m_B^-$) *gal lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	BL21 mutant designed for production of membrane proteins [112]
C600	*E. coli* K12 F^- *supeE44 hsdR thi-1 thr-1 leuB6 lacY1 tonA21*	K12 derivative
HB101	F^-*hsd20*($r_B^-m_B^-$) *recA13 rpsL20 leu proA2*	Hybrid between *E. coi* K12 and B
HMS174 (DE3)	*E. coli* K12 F^- *recA1 hsdR*($r_{K12}^-m_{K12}^+$) (*Rif R*)	*recA* mutation in a K12 background
JM101	*E. coli* K12 *glnV44 thi-1* Δ*(lac-proAB) F'[lacI*^{q}*Z* Δ*M15 traD36 proAB*$^+$]	Original blue/white cloning strain
MG1655	*E. coli* K12 F^- λ^- *ilvG- rfb-50 rph-1*	K12 wild-type strain
Origami (DE3)	*E. coli* K12 Δ*(ara-leu)7697* Δ*lacX74* Δ*phoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac*$^+$ *lacI*^{q} *pro] (DE3) gor522::Tn10 trxB pLysS (Cam*^{R} *Str*^{R} *Tet*^{R}*)*	K12 strain with mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes
Rosetta (DE3)	*E. coli* B F^- *dcm ompT hsdS* $_B$($r_B^-m_B^-$) *gal lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])RARE (Cam*^{R}*)*	BL21 derivative for expression of eukaryotic proteins that contain codons rarely used in *E. coli*
Rosetta-gami (DE3)	Δ*(ara-leu)7697* Δ*lacX74* Δ*phoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[lac*$^+$ *lacI*^{q} *pro] gor522::Tn10 trxB pLysSRARE (Cam*^{R}, *Str*^{R}, *Tet*^{R}*)*	K12 Rosetta–origami hybrid
TG1	*E. coli* K12 *F' [traD36 proAB*$^+$ *lacI*^{q} *lacZ* Δ*M15]supE thi-1* Δ*(lac-proAB)* Δ*(mcrB- hsdSM)5* ($r_K^-m_K^-$)	Strain useful for phage display
W3110	*E. coli* K12 F^- λ^-*rph-1 INV(rrnD, rrnE)*	K12 wild-type strain

IPTG isopropyl-β-D-thiogalactoside

parent strain BL21. They noticed that in BL21 the percentage of plasmid-carrying cells dropped from 87 to 0% within the first 4 h after induction with IPTG. However, in *E. coli* BLR this value remained at 100% [189]. Another example is *E. coli* C41, a mutant BL21 strain selected for the production of membrane proteins [112]. The best known BL21-derived strain is *E. coli* Rosetta (Merck KGaA, Darmstadt, Germany). Rosetta strains harbor a chloramphenicol-resistant plasmid, pRARE, which contains multiple genes corresponding to tRNAs for codons that are rarely used in *E. coli*. Consequently, since codon usage between eukaryotes and *E. coli* can vary greatly, this strain is predominately utilized to enhance the expression of eukaryotic proteins. Besides co-expressing tRNAs, codon bias can also be addressed by optimization of the coding sequence, i.e., by substitution of lesser-used codons with synonymous ones that are more frequently used in *E. coli*. Both strategies generally result in enhanced yields of eukaryotic proteins. However, the former strategy is more economical when a large number of genes are to be expressed, whereas the latter usually yields better results and in addition offers the opportunity of altering other features such as mRNA secondary structures [17]. Although to date few recombinant proteins have been produced by HCDCs using *E. coli* Rosetta as a host (only two examples in Table 3), a lot of effort has been put into comparing the production capabilities of the strain with that of BL21. Tegel et al. [159, 160] tested a set of 68 recombinant proteins in both strains and concluded that mainly proteins that are badly produced in BL21 were better expressed in *E. coli* Rosetta. Furthermore, the purity of the proteins is greatly improved in Rosetta. Also Brenac et al. [15] observed a higher production of a recombinant mouse protein in Rosetta strains compared with BL21.

Escherichia coli K12 strains have also been used in heterologous protein production, despite their attributed disadvantages such as high acetate production and higher protease activity. Daughter strains MG1655 and W3110 have been applied as hosts in some applications, however, to a lesser extent than BL21 (see Table 3). These K12 strains were probably used because the corresponding laboratories had a lot of experience in their cultivation.

Furthermore, in two applications, mutant strains had to be created and in contrast with K12 strains, at the time of publication no genome information was available for B strains [177, 181]. Lack of genome sequence information complicates mutant creation. The genome of *E. coli* K12 was sequenced in 1997 [13]; however, for BL21, sequence information only became available in 2009 [80]. A very interesting K12 strain is *E. coli* TG1. TG1 produces a very low amount of acetate both in batch and glucose-limited cultures, which beneficially influences biomass yield and protein production [86, 94]. Despite these attractive characteristics, the applications using *E. coli* TG1 as a host are limited to three examples (see Table 3). *E. coli* C600, DH5α, and JM101 are also K12 strains but are only used in a handful of applications. Maybe the most important K12-derived strain is the redox-modified *E. coli* Origami (Merck KGaA, Darmstadt, Germany). Origami strains contain mutations in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, and thereby promote disulfide bond formation, a process which is normally (in the wild-type strain) strongly disfavored because of the reduced state of the cytoplasm [11]. All protein production examples listed in Table 3 using Origami refer to proteins for which disulfide bonds are important for the activity and stability of the protein. *E. coli* Rosetta-gami (Merck KGaA, Darmstadt, Germany) strains are Origami strains that harbor the pRARE plasmid and thereby combine the advantages of Origami strains (disulfide bond formation) with that of Rosetta (providing tRNAs for rare codons). Zhang et al. [188] compared the performance of BL21, Rosetta, and Rosetta-gami for the production of human insulin-like growth factor 1 (hIGF-1) and observed that Rosetta-gami clearly outperformed the other two strains.

Furthermore, from Table 3 it can be concluded that *E. coli* strains are mainly used for the production of small proteins. Only 8 of the 44 proteins are larger than 50 kDa, whereas 14 are smaller than 20 kDa.

Engineering of *E. coli* for increased RPP

Many efforts have been made to increase RPP in *E. coli* or to express proteins that are generally difficult to produce in this host. These engineering attempts can be divided into the following classes:

1. Engineering of stress-related phenomena that generally result in breakdown of the recombinant protein
2. Optimization of inclusion body formation
3. Reduction of acetate formation by genetic engineering
4. Engineering of *E. coli* for the extracellular production of proteins
5. Introduction of PTM steps: expression of heterologous proteins that contain disulfide bonds and introduction of glycosylation reactions

Engineering of stress-related phenomena

For the production of heterologous proteins, the cDNA of the gene of interest is usually cloned into plasmids under the control of a strong inducible promoter, in order to achieve high transcription rates and high protein concentrations. The use of high-copy plasmids and strong promoters can result in such strong protein formation that more than 50% of the cellular protein consists of recombinant protein [8, 56]. However, the folding machinery of the cell is not adapted to fold such substantially higher amounts of protein. When a protein or a part of a protein fails to reach its native conformation in time, a stress response is initiated [59], leading to the transcription of stress-related proteins, like proteases, eventually resulting in breakdown of the heterologous protein. When the protein of interest requires additional PTM steps that *E. coli* cannot perform, the possibility of misfolding and stress increases even further. Misfolded proteins initiate a heat-shock like response, which comprises the upregulation and activition of a set of proteins that are normally active when cells are exposed to a sudden elevated temperature [75, 126]. A second possibility, besides proteolytic breakdown, is that the misfolded protein aggregates in inclusion bodies. Recent research has shown that inclusion bodies can be considered as transient reservoirs of aggregated, unfolded polypeptides that are still under the quality control surveillance of cell chaperones and proteases [169]. This aggregation is usually not desired.

In addition to misfolding, other factors can also induce stress responses. The presence of multicopy plasmids or gene overexpression causes metabolic burden and growth retardation [10]. High expression of foreign genes requires energy and building blocks that are normally reserved for the general cellular activities of the host cell, such as biosynthesis of DNA, membranes, and enzymes. The consumption of cellular resources for heterologous protein production can outbalance the cellular functions dramatically, resulting in starvation and even suicidal behavior [46, 97].

Most engineering strategies focus on prevention of misfolding, neutralization of the increased protease activity or stress response, or optimization of heterologous gene expression [30].

Proper folding can be promoted by co-expression of the chaperones DnaK and GroEL or their co-chaperones or a combination of these chaperones [37, 39, 93, 162]. Proper folding prevents the recombinant protein from provoking a

severe stress response or from being deposited in inclusion bodies. This strategy has been successful in many applications, leading to improved solubilities, enhanced yields, and enhanced specific activities [147, 154, 175]. However, this approach cannot be generalized because it often results in other adverse effects such as the formation of soluble aggregates, lower specific activities, and enhanced proteolysis resulting in lower yields [108]. This is because chaperone activities are also linked to proteolysis, in addition to their valuable function in folding of the recombinant protein. Finding the optimal composition and strength of the co-expressed chaperones is case-dependent and usually determined by trial and error.

A second often used strategy is that the increased proteolysis is counteracted by deleting genes that code for proteases. *E. coli* proteases are well documented [62, 64–66, 139], making the targets for knockouts easy to identify. However, proteases are essential to cellular quality control of homologous proteins and thus cell viability, which limits the possibility of deleting many or all protease genes. Meerman and Georgiou [110] constructed a family of 25 isogenic strains deficient in all known cell envelope proteases (DegP, Protease III, Tsp(Prc), and OmpT), as well as the heat shock sigma factor *rpoH15* mutant allele, and characterized their growth in both shake flasks and fermentors. Deletion of DegP, OmpT, and Lon have led to the best results in terms of improved protein production without impairment on growth [63, 113]. Another possibility is to alter the sensitivity of the recombinant protein towards proteases by identifying and engineering of the proteolytic sites on the protein [116]. This is a very valuable alternative to knocking out protease genes, because this strategy does not impede the normal cellular function of the proteases.

Besides neutralizing the increased protease activity, the stress response can also be counteracted by modifying the expression of stress-related sigma factors. Many recombinant proteins are produced by fed-batch cultivations and as a result of the low specific growth rate applied in the feed phase, a stationary-phase-like condition is initiated, which leads to an increased stress response [186]. The transition from exponential growth phase to stationary phase is transcriptionally regulated by the sigma factor σ^S, encoded by the *rpoS* gene [72]. By deleting the *rpoS* gene, Jeong et al. [81] successfully improved the production of human leptin in fed-batch cultivations.

Choosing the proper *E. coli* host strain is also a strategy to decrease stress response and proteolytic activity. Seo et al. [141] compared four *E. coli* strains: production hosts JM105 and BL21, and cloning hosts HB101 and TOP10. They observed that BL21 exhibited the lowest cellular stress level and the highest protein expression. BL21 is also known to be deficient in the proteases *lon* and *ompT* [64].

Optimization of inclusion body formation

In some applications, misfolding of proteins and the accompanying formation of inclusion bodies can also be desired. In this strategy, the protein of interest is produced in inclusion bodies during the fermentation, the inclusion bodies are subsequently washed and solubilized, and the recombinant protein is refolded and purified. The advantages are that inclusion bodies can accumulate in the cytoplasm to much higher amounts than soluble proteins resulting in higher yields, and that for some specific proteins, purification can be very simple (e.g., purification and refolding in only one step [6, 164]). Furthermore, inclusion bodies generally contain very high amounts (more than 95%) of pure unfolded recombinant protein. Finally, when captured in inclusion bodies, the recombinant protein is considered to be protected against proteases [26, 150, 165]. Consequently, this strategy has been applied in numerous examples [6, 79, 87, 122, 164]. Main drawbacks remain the absence of generally applicable and inexpensive methods to refold the protein and the low overall yields of downstream processing [150]. Although in some occasions these problems might have been successfully tackled, industry still seeks processes which are economical and applicable for a wide range of proteins [49].

Reduction of acetate formation by genetic engineering

One of the key problems observed in HCDCs of *E. coli* for the production of recombinant proteins is acetate toxicity [26]. Acetate is produced as a result of overflow metabolism when cells grow very rapidly and cannot metabolize the delivered carbon source fast enough, even under fully aerobic conditions [3, 77]. However, even low concentrations of acetate, e.g., 0.5 g L^{-1}, can hamper growth and obstruct the production of recombinant proteins [114]. Consequently, many engineering approaches have focused on the reduction of acetate formation in order to improve RPP in *E. coli*. These approaches can be classified as follows: (a) deletion of acetate pathway genes, (b) avoiding overflow metabolism by engineering of the glucose uptake system, and (c) avoiding overflow metabolism by redirecting central metabolic fluxes. These different approaches are summarized in Fig. 3.

Maybe a first instinctive approach to decrease acetate formation is to knockout genes that encode for acetate pathway enzymes, e.g., *ackA* (acetate kinase), *pta* (phosphate acetyltransferase), and *poxB* (pyruvate oxidase) [32, 40, 44, 45, 184]. These endeavors resulted in a decreased acetate yield, however at the expense of pyruvate, lactate, or formate formation, which are also undesired by-products.

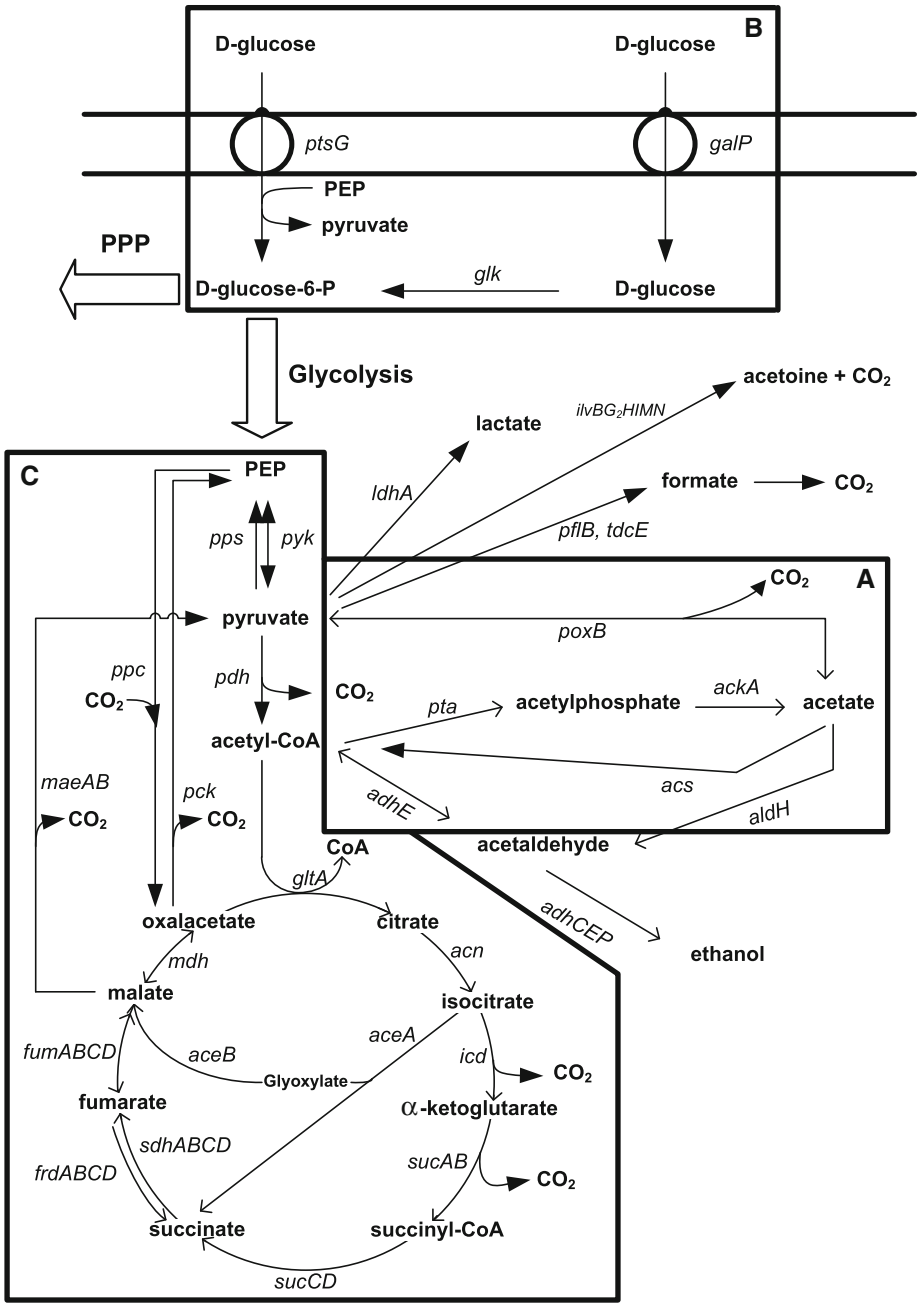

Fig. 3 Strategies for acetate reduction in *E. coli*, **a** knocking out genes that encode for acetate pathway enzymes, **b** engineering the glucose uptake system, **c** redirecting central metabolic fluxes. *aceA* isocitrate lyase, *aceB* malate synthase, *ackA* acetate kinase, *acn* aconitase, *acs* acetyl-CoA synthase, *adhCEP* ethanol dehydrogenase, *adhE* aldehyde dehydrogenase, *fumABCD* fumarase, *galP* galactose permease, *glk* glucokinase, *icd* isocitrate dehydrogenase, *ilvBG$_2$HIMN* acetolacetate decarboxylase, *ldhA* lactate dehydrogenase, *maeAB* malic enzyme, *mdh* malate dehydrogenase, *pck* phosphoenolpyruvate carboxykinase, *pdh* pyruvate dehydrogenase, *pflB, tdcE* pyruvateformate lyase, *poxB* pyruvate oxidase, *ppc* phosphoenolpyruvate carboxylase, *pps* phosphoenolpyruvate synthase, *pta* acetylphosphotransferase, *pyk* pyruvate kinase, *sdhABCD* succinate dehydrogenase, *sucAB* α-ketoglutarate dehydrogenase, *sucCD* succinate thiokinase

A second, and in many cases successful, approach is to engineer the glucose uptake system to overcome overflow metabolism. Overflow metabolism occurs when high glycolytic fluxes, due to the rapid uptake of glucose, are not further processed in the tricarboxylic acid (TCA) cycle. Consequently, a bottleneck occurs at the pyruvate node, and pyruvate is converted to acetate instead of being processed to acetyl-CoA. By deleting one of the phosphotransferase system (PTS) genes, e.g., *ptsG*, *ptsH*, or *ptsI*, the uptake through the major glucose transporter is severely impeded, resulting in a reduced glycolytic flux and a reduced acetate production [29, 148, 178]. However, one of the major drawbacks of this approach is the strong reduction in growth rate. By overexpressing the alternative glucose transporter gene *galP* (galactose permease) and exploiting the native glucose kinase (Glk) transporter, the wild-type growth rate can be restored [36]. The resulting strain *E. coli* W3110 Δ*ptsHGalP*$^+$ exhibits a very low acetate yield and an increased recombinant protein yield compared with the W3110 wild type, without a reduction in growth rate. Similarly, Wong et al. [178] restored glucose transport by co-expressing *glf*, encoding a passive

glucose transporter of *Zymomonas mobilis*; however, this only resulted in a decreased acetate formation in M9 minimal media, not in LB media.

Instead of limiting the glucose uptake through the PTS, it is also possible to redirect metabolic fluxes around the phosphoenolpyruvate–pyruvate–oxaloacetate node to overcome acetate formation. Farmer and Liao [51] increased anaplerotic and glyoxylate fluxes by overexpressing phosphoenolpyruvate carboxylase (Ppc) and by deleting the FadR regulator. This strategy resulted in an over 75% decrease in acetate yield compared with its wild type. Alternatively, the overexpression of a heterologous anaplerotic pyruvate carboxylase from *Rhizobium etli* resulted in a 57% reduction in acetate formation and a concomitant 68% increase in β-galactosidase production [107]. Recently, it was shown that when the genes encoding the global regulator ArcA and the local regulator IclR are deleted, an over 70% decrease in acetate formation can be achieved in *E. coli* K12. This decrease in acetate formation could be explained by the activation of the glyoxylate pathway (Δ*iclR*) and relief of repression on the TCA cycle enzymes (Δ*arcA*) [173], and was accompanied by an increase in biomass yield to 0.63 c-mole/c-mole glucose[4] which approximates the maximum theoretical yield as determined by Varma et al. [167, 168].

Finally, it is possible to combine different strategies to reduce acetate formation. For example, De Mey et al. [40, 41] reduced acetate and other by-product formation by deleting *ack*, *pta*, and *poxB*, combined with the overexpression of *ppc*.

An approach which is not covered in this review since it does not concern the engineering of *E. coli*, is to decrease acetate production by manipulation of the cultivation conditions, e.g., overflow metabolism can also be circumvented by setting a fixed low growth rate in fed-batch processes [101]. More information about these approaches and about minimizing acetate in *E. coli* fermentation can be found elsewhere [42].

Extracellular production of proteins

The secretory production of recombinant proteins into the culture medium has several advantages compared with the conventional production in inclusion bodies or the cytoplasm. First, purification of the protein is much easier owing to reduced contamination of various cellular components, which decreases the cost of downstream processing. Moreover, in the culture medium, recombinant proteins are safe from cytoplasmic or periplasmic proteases. In addition, when the protein of interest is toxic to the host cell, it will affect cellular metabolism less than if it were produced intracellularly. Finally, extracellular proteins are easier to measure, which could be an issue when optimization is still ongoing and convenient protocols need to be developed, e.g., in directed evolution [119].

Escherichia coli contains different secretion systems for the transport of proteins from the cytoplasm to the periplasm or extracellular environment, designated as type I through V (for more background information, see [163]). Types VI and VII are identified in bacteria, but not in *E. coli* [1, 53]. It is possible that proteins are exported in one step across the inner and outer membranes as in types I, III, and IV, or in two steps, in which the protein is first transported into the periplasmic space via the general secretion pathway (Sec) or twin-arginine translocation (Tat) system and subsequently translocated across the outer membrane via the type II or V system. Types I, II and the Sec machineries have been applied the most intensively for recombinant production processes with *E. coli* [111], primarily because they were the identified first. A system that has attracted much interest in recent years is the Tat system. In contrast to other secretion systems, the Tat machinery can secrete large, folded proteins. Furthermore, for relatively simple proteins there seems to be some kind of proofreading/quality control that ensures that proteins are correctly folded before they are secreted [117, 137]. The combination of the Tat system for transport across the inner membrane and the type V machinery for transport across the outer membrane could be very promising, since the type V system exhibits a relatively simple secretion mechanism and has a high transport capacity [84].

Generally in all secretion systems, the proteins are synthesized in the cytoplasm as a kind of preprotein and are linked to a specific amino acid sequence or signal sequence, which allows recognition and transport by the secretion system. During transport, the signal sequence is cleaved off by a signal peptidase to yield a mature protein. The selection of the optimal signal peptide is crucial for efficient secretion; however, in most cases optimization is performed by trial and error [28]. Most engineering approaches focus on (1) selection and modification of the signal peptide, (2) co-expression of proteins that assist in translocation and folding, (3) improvement of periplasmic release when transport occurs in two steps, and (4) protection of the target protein from degradation and contamination [187].

In the literature, there are more than 50 successful examples of secretory production of recombinant proteins in *E. coli*, overviews of which can be found elsewhere [28, 84, 114, 119].

[4] 1 c-mole equals 1 mole multiplied by the number of C-atoms in the molecule. This concept is used to express mass balances in biochemistry.

Introduction of PTM steps

As mentioned earlier, the most important drawback of *E. coli* remains that this host cannot carry out PTMs. Consequently, many research efforts have focused on this topic. Two PTMs which are utterly vital for the stability and activity of recombinant proteins are disulfide bond formation and glycosylation.

Disulfide bond formation in the cytoplasmic environment of eukaryotes or prokaryotes is strongly disfavored because of the high reducing power of this compartment. As a result, the cytoplasmic thiol–disulfide redox potential is too low to allow the formation of disulfide bonds, i.e., the reaction is not thermodynamically feasible. Furthermore, the enzymes that could carry out disulfide bond formation, i.e., thioredoxins and glutaredoxins are maintained in a reduced state by thioredoxin reductase and glutathione, respectively [131]. To allow the catalyzed formation of disulfide bonds, complex pathways have evolved in more oxidizing cellular compartments, such as the ER of eukaryotes, the mitochondrial intermembrane, and the periplasm of prokaryotes [70].

Several engineering strategies have been developed to address this issue in *E. coli*, as excellently reviewed by de Marco [38]. First, disulfide bond formation can be facilitated by secreting the protein into the more oxidizing periplasmic space by exploiting the Sec or Tat secretion system (see Sect. "Extracellular production of proteins"). Second, the redox state of the cytoplasm can be altered by mutations in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes. However, the resulting strains grow very slowly and require the addition of an exogenous reductant, like dithiothreitol (DTT), to achieve a reasonable growth rate [131]. Bessette et al. [11] isolated strains still sharing the *trxB-gor* knockouts, but which suppress the slow-growing phenotype. Two years later, Ritz et al. [136] showed that the recovery of growth was due to a mutation in the *ahpC* gene, a mutation which converted the functionality of the enzyme from a peroxiredoxin into a disulfide reductase. The resulting strains are commercialized by Merck as Origami strains. The successful applications of Origami strains in HCDCs are listed in Table 3. A third strategy includes the cytoplasmic overexpression of periplasmic disulfide oxidoreductases, enzymes which enhance the rate of disulfide bond isomerization, e.g., DsbC. This strategy was implemented in the growth-optimized *trxB-gor* knockout strain and resulted in a 20-fold increase of production of the truncated form of the human tissue plasminogen activator (vtPA) [11]. This strain is commercially available as *E. coli* SHuffle (New England Biolabs). It is, however, not essential that the naturally occurring reduction pathways are disrupted, as recent research has shown that the introduction of a sulfhydryl oxidase Erv1p into the cytoplasm of wild-type *E. coli* results in equally high or higher levels of active recombinant proteins, compared with Δ*gor*Δ*trxB* strains [70]. Erv1p is a FAD-dependent catalyst of disulfide bond formation, found in the intermembrane space of mitochondria of *S. cerevisiae*.

Besides the proper formation of disulfide bonds, *E. coli* also lacks glycosylation pathways. In the early days of RPP, the possibility of adding glycans to proteins was only attributed to eukaryotes. However, N- and O-glycoproteins have now been described for all domains of life, including archaea and bacteria [121]. The first-found and best-characterized example of bacterial glycosylation is the gene cluster *pgl* in *Campylobacter jejuni* [156]. This gene cluster was successfully transferred in *E. coli*, with the result that this bacterium is now capable of producing N-linked glycoproteins [172]. The *pgl* cluster contains 12 genes, five of which encode putative glycosyltransferases involved in the assembly of a heptasaccharide on a lipid carrier and one gene encodes an enzyme responsible for the en bloc transfer from the carrier to the protein. The cluster also encodes for a UDP-*N*-acetylglucosamine C-6 dehydratase, a C-4 aminotransferase, and an *N*-acetyltransferase [2]. Moreover, recent research has yielded the construction of a simple, genetically encoded glycosylation tag, which in combination with the *pgl* system, is sufficient to promote covalent attachment of multiple glycans to a single protein carrier, thereby expanding the glycosylation possibilities for recombinant *E. coli* [54].

To date, engineering attempts have resulted in bacterial glycans which are structurally rather different from their eukaryotic counterparts. Promisingly, however, the PglB glycosyltransferase is relatively promiscuous in its choice of oligosaccharide chain and protein. Furthermore, the specific glycoform could be genetically encoded by co-expression of the right set of glycosyltransferases in addition to the expression of PglB. This implies that in the future other 'more eukaryotic' glycoproteins might be expressed in *E. coli* as well.

Conclusions

Although other expression hosts, such as mammalian cell systems, are being explored and applied for RPP, *E. coli* remains a very often used host. Its fast growth on cheap substrates, its well-characterized genetics, the many molecular toolboxes available, and the high cell densities that can be reached, favor its use. Despite these enviable characteristics, *E. coli* faces some major drawbacks as well. Proper folding is often problematic, with the result that the protein ends up in inclusion bodies, hindering recovery of activity. Furthermore, *E. coli* cells produce acetate as a

by-product in aerobic fermentations, which inhibits growth and the formation of the heterologous protein. Probably the most important obstruction for the universal use of *E. coli* in RPP is its difficulty in performing PTMs, such as the formation of disulfide bonds and glycosylation. Many engineering strategies have attempted to overcome these shortcomings. Stress-related phenomena have been mapped and by finding an optimal expression of chaperones, proper folding has been promoted. Moreover, by modifying the expression of stress-related sigma factors or by deleting specific proteases, the stress response associated with RPP could be controlled. In addition, secretion pathways have been characterized and exploited for the production of extracellular proteins, thereby avoiding proteolysis and inclusion body formation. Furthermore, many efforts have been made to minimize acetate production in *E. coli* cultivations, primarily by preventing the occurrence of overflow metabolism. Finally, in the last decade, research has also focused on the possibility of carrying out PTMs in *E. coli*. Strains have been developed with an altered cytoplasmic redox state to allow the formation of disulfide bonds and heterologous glycosylation pathways have been introduced. To summarize, all these research efforts have resulted in an extensive list of *E. coli* strains, many of which are commercially available. Taking into account all the research projects that focus on the production of more difficult eukaryotic proteins in *E. coli*, no doubt this host will remain an important workhorse in RPP in the future.

Acknowledgments This work was financially supported by the Special Research Fund (BOF) of Ghent University. The authors like to thank Jo Maertens and Joeri Beauprez for lively scientific discussions and critical comments.

References

1. Abdallah AM, van Pittius NCG, Champion PAD, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, Appelmelk BJ, Bitter W (2007) Type VII secretion—mycobacteria show the way. Nat Rev Microbiol 5:883–891
2. Abu-Qarn M, Eichler J, Sharon N (2008) Not just for eukarya anymore: protein glycosylation in bacteria and archaea. Curr Opin Struct Biol 18:544–550
3. Andersen KB, von Meyenburg K (1980) Are growth rates of *Escherichia coli* in batch cultures limited by respiration? J Bacteriol 144:114–123
4. Babaeipour V, Abbas MPH, Sahebnazar Z, Alizadeh R (2010) Enhancement of human granulocyte-colony stimulating factor production in recombinant *E. coli* using batch cultivation. Bioprocess Biosyst Eng 33:591–598
5. Babaeipour V, Shojaosadati SA, Khalilzadeh R, Maghsoudi N, Farnoud AM (2010) Enhancement of human gamma-interferon production in recombinant *E. coli* using batch cultivation. Appl Biochem Biotechnol 160:2366–2376
6. Babu KR, Swaminathan S, Marten S, Khanna N, Rinas U (2000) Production of interferon-α in high cell density cultures of recombinant *Escherichia coli* and its single step purification from refolded inclusion body proteins. Appl Microbiol Biotechnol 53:655–660
7. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Curr Opin Biotechnol 10:411–421
8. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol 22:1399–1408
9. Barnard GC, Henderson GE, Srinivasan S, Gerngross TU (2004) High level recombinant protein expression in *Ralstonia eutropha* using T7 RNA polymerase based amplification. Protein Expr Purif 38:264–271
10. Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS (1990) Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng 35:668–681
11. Bessette PH, Aslund F, Beckwith J, Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. Proc Natl Acad Sci U S A 96:13703–13708
12. Blaas L, Musteanu M, Eferl R, Bauer A, Casanova E (2009) Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. BMC Biotechnol 9:3
13. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462
14. Bobrowicz P, Davidson RC, Li H, Potgieter TI, Nett JH, Hamilton SR, Stadheim TA, Miele RG, Bobrowicz B, Mitchell T, Rausch S, Renfer E, Wildt S (2004) Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. Glycobiology 14:757–766
15. Brenac V, Mouz N, Schapman A, Ravault V (2006) Expression optimization and purification process development of an engineered soluble recombinant mouse linker of activation of T cells using surface enhanced laser desorption/ionization-mass spectrometry. Protein Expr Purif 47:533–541
16. Brüsehaber E, Schwiebs A, Schmidt M, Böttcher D, Bornscheuer UT (2010) Production of pig liver esterase in batch fermentation of *E. coli* Origami. Appl Microbiol Biotechnol 86:1337–1344
17. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. Protein Expr Purif 59:94–102
18. Butler M (2006) Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. Cytotechnology 50:57–76
19. Callewaert N, Laroy W, Cadirgi H, Geysens S, Saelens X, Jou WM, Contreras R (2001) Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1,2-alpha-D-mannosidase for N-glycan engineering in *Pichia pastoris*. FEBS Lett 503:173–178
20. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13:329–332
21. Chao YP, Law W, Chen PT, Hung WB (2002) High production of heterologous proteins in *Escherichia coli* using the thermoregulated T7 expression system. Appl Microbiol Biotechnol 58:446–453
22. Chauhan V, Singh A, Waheed SM, Singh S, Bhatnagar R (2001) Constitutive expression of protective antigen gene of *Bacillus anthracis* in *Escherichia coli*. Biochem Biophys Res Commun 283:308–315
23. Chen H, Lin M, Hou S (2008) Multiple-copy-gene integration on chromosome of *Escherichia coli* for beta-galactosidase production. Korean J Chem Eng 25:1082–1087

24. Choi BK, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. Proc Natl Acad Sci U S A 100:5022–5027
25. Choi JH, Jeong KJ, Kim SC, Lee SY (2000) Efficient secretory production of alkaline phosphatase by high cell density culture of recombinant *Escherichia coli* using the *Bacillus* sp. endoxylanase signal sequence. Appl Microbiol Biotechnol 53:640–645
26. Choi JH, Keum KC, Lee SY (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem Eng Sci 91:876–885
27. Choi JH, Lee SJ, Lee SJ, Lee SY (2003) Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. Appl Environ Microbiol 69:4737–4742
28. Choi JH, Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl Microbiol Biotechnol 64:625–635
29. Chou CH, Bennett GN, San KY (1994) Effect of modified glucose uptake using genetic engineering techniques on high-level recombinant protein production in *Escherichia coli* dense cultures. Biotechnol Bioeng 44:952–960
30. Chou CP (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Appl Microbiol Biotechnol 76:521–532
31. Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A 70:3240–3244
32. Contiero J, Beatty C, Kumari S, DeSanti C, Strohl W, Wolfe A (2000) Effects of mutations in acetate metabolism on high-cell-density growth of *Escherichia coli*. J Ind Microbiol Biotechnol 24:421–430
33. Cregg JM, Cereghino JL, Shi J, Higgins DR (2000) Recombinant protein expression in *Pichia pastoris*. Mol Biotechnol 16:23–52
34. Dammeyer T, Steinwand M, Krüger SC, Dübel S, Hust M, Timmis KN (2011) Efficient production of soluble recombinant single chain Fv fragments by a *Pseudomonas putida* strain KT2440 cell factory. Microb Cell Fact 10:11
35. Dasgupta S, Navarrete AM, Bayry J, Delignat S, Wootla B, André S, Christophe O, Nascimbeni M, Jacquemin M, Martinez-Pomares L, Geijtenbeek TBH, Moris A, Saint-Remy JM, Kazatchkine MD, Kaveri SV, Lacroix-Desmazes S (2007) A role for exposed mannosylations in presentation of human therapeutic self-proteins to CD4+ T lymphocytes. Proc Natl Acad Sci U S A 104:8965–8970
36. De Anda R, Lara AR, Hernández V, Hernández-Montalvo V, Gosset G, Bolívar F, Ramírez OT (2006) Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of *Escherichia coli* for recombinant protein production without impairment of growth rate. Metab Eng 8:281–290
37. de Marco A (2007) Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. Nat Protoc 2:2632–2639
38. de Marco A (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. Microb Cell Fact 8:26
39. de Marco A, De Marco V (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. J Biotechnol 109:45–52
40. De Mey M, Lequeux GJ, Beauprez JJ, Maertens J, Van Horen E, Soetaert WK, Vanrolleghem PA, Vandamme EJ (2007) Comparison of different strategies to reduce acetate formation in *Escherichia coli*. Biotechnol Prog 23:1053–1063
41. De Mey M, Lequeux GJ, Beauprez JJ, Maertens J, Waegeman HJ, Bogaert INV, Foulquié-Moreno MR, Charlier D, Soetaert WK, Vanrolleghem PA, Vandamme EJ (2010) Transient metabolic modeling of *Escherichia coli* MG1655 and MG1655 DeltaackA-pta, DeltapoxB Deltapppc ppc-p37 for recombinant beta-galactosidase production. J Ind Microbiol Biotechnol 37:793–803
42. De Mey M, Maeseneire SD, Soetaert W, Vandamme E (2007) Minimizing acetate formation in *E. coli* fermentations. J Ind Microbiol Biotechnol 34:689–700
43. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27:297–306
44. Diaz-Ricci JC, Regan L, Bailey JE (1991) Effect of alteration of the acetic acid synthesis pathway on the fermentation pattern of *Escherichia coli*. Biotechnol Bioeng 38:1318–1324
45. Dittrich CR, Vadali RV, Bennett GN, San KY (2005) Redistribution of metabolic fluxes in the central aerobic metabolic pathway of *E. coli* mutant strains with deletion of the ackA-pta and poxB pathways for the synthesis of isoamyl acetate. Biotechnol Prog 21:627–631
46. Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J Bacteriol 177:1497–1504
47. Durocher Y, Butler M (2009) Expression systems for therapeutic glycoprotein production. Curr Opin Biotechnol 20:700–707
48. Dutta S, Lalitha PV, Ware LA, Barbosa A, Moch JK, Vassell MA, Fileta BB, Kitov S, Kolodny N, Heppner DG, Haynes JD, Lanar DE (2002) Purification, characterization, and immunogenicity of the refolded ectodomain of the *Plasmodium falciparum* apical membrane antigen 1 expressed in *Escherichia coli*. Infect Immun 70:3101–3110
49. Eiberle MK, Jungbauer A (2010) Technical refolding of proteins: do we have freedom to operate? Biotechnol J 5:547–559
50. Espinosa R, Caballero E, Musacchio A, Silva R (2002) Production of a recombinant, immunogenic protein, P64k, of *Neisseria meningitidis* in *Escherichia coli* in fed-batch fermenters. Biotechnol Lett 24:343–346
51. Farmer WR, Liao JC (1997) Reduction of aerobic acetate production by *Escherichia coli*. Appl Environ Microbiol 63: 3205–3210
52. Ferrer-Miralles N, Domingo-Espín J, Corchero JL, Vázquez E, Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8:17
53. Filloux A, Hachani A, Bleves S (2008) The bacterial type VI secretion machine: yet another player for protein transport across membranes. Microbiology 154:1570–1583
54. Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP (2011) Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli*. Appl Environ Microbiol 77:871–881
55. Fu ZB, Ng KL, Lam CC, Leung KC, Yip WH, Wong WKR (2006) A two-stage refinement approach for the enhancement of excretory production of an exoglucanase from *Escherichia coli*. Protein Expr Purif 48:205–214
56. Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11
57. Gellissen G, Melber K, Janowicz ZA, Dahlems UM, Weydemann U, Piontek M, Strasser AW, Hollenberg CP (1992)

Heterologous protein production in yeast. Antonie Van Leeuwenhoek 62:79–93
58. Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 22:1409–1414
59. Gill RT, Valdes JJ, Bentley WE (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. Metab Eng 2:178–189
60. Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc Natl Acad Sci U S A 76:106–110
61. Goodman M (2009) Market watch: sales of biologics to show robust growth through to 2013. Nat Rev Drug Discov 8:837
62. Gottesman S (1989) Genetics of proteolysis in *Escherichia coli*. Annu Rev Genet 23:163–198
63. Gottesman S (1990) Minimizing proteolysis in *Escherichia coli*: genetic solutions. Methods Enzymol 185:119–129
64. Gottesman S (1996) Proteases and their targets in *Escherichia coli*. Annu Rev Genet 30:465–506
65. Gottesman S (2003) Proteolysis in bacterial regulatory circuits. Annu Rev Cell Dev Biol 19:565–587
66. Gottesman S, Maurizi MR (1992) Regulation by proteolysis: energy-dependent proteases and their targets. Microbiol Rev 56:592–621
67. Goyal D, Sahni G, Sahoo DK (2009) Enhanced production of recombinant streptokinase in *Escherichia coli* using fed-batch culture. Bioresour Technol 100:4468–4474
68. Grinna L, Tschopp J (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast *Pichia pastoris*. Yeast 5:107–115
69. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, Gerngross TU (2003) Production of complex human glycoproteins in yeast. Science 301:1244–1246
70. Hatahet F, Nguyen VD, Salo KEH, Ruddock LW (2010) Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. Microb Cell Fact 9:67
71. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. Science 291:2364–2369
72. Hengge R (2008) The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. Adv Exp Med Biol 631:40–53
73. Higgins DR (2001) Overview of protein expression in *Pichia pastoris*. Curr Protoc Protein Sci 5:7
74. Hodgson J (1994) The changing bulk biocatalyst market. Nat Biotechnol 12:789–790
75. Hoffmann F, Rinas U (2004) Stress induced by recombinant protein production in *Escherichia coli*. Adv Biochem Eng Biotechnol 89:73–92
76. Hollenberg C, Gellissen G (1997) Production of recombinant proteins by methylotrophic yeasts: *Pichia pastoris*. Curr Opin Biotechnol 8:554–560
77. Holms WH (1986) The central metabolic pathways of *Escherichia coli*: relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. Curr Top Cell Regul 28:69–105
78. Hossler P, Khattak SF, Li ZJ (2009) Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 19:936–949
79. Hu SY, Wu JL, Huang JH (2004) Production of tilapia insulin-like growth factor-2 in high cell density cultures of recombinant *Escherichia coli*. J Biotechnol 107:161–171
80. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol 394:644–652
81. Jeong KJ, Choi JH, Yoo WM, Keum KC, Yoo NC, Lee SY, Sung MH (2004) Constitutive production of human leptin by fed-batch culture of recombinant rpoS- *Escherichia coli*. Protein Expr Purif 36:150–156
82. Jeong KJ, Lee SY (1999) High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification. Appl Environ Microbiol 65:3027–3032
83. Jeong KJ, Lee SY (2002) Excretion of human beta-endorphin into culture medium by using outer membrane protein F as a fusion partner in recombinant *Escherichia coli*. Appl Environ Microbiol 68:4979–4985
84. Jong WSP, Saurí A, Luirink J (2010) Extracellular production of recombinant proteins using bacterial autotransporters. Curr Opin Biotechnol 21:646–652
85. Kamionka M (2011) Engineering of therapeutic proteins production in *Escherichia coli*. Curr Pharm Biotechnol 12:268–274
86. Kayser A, Weber J, Hecht V, Rinas U (2005) Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. Microbiology 151:693–706
87. Khalilzadeh R, Shojaosadati SA, Maghsoudi N, Mohammadian-Mosaabadi J, Mohammadi MR, Bahrami A, Maleksabet N, Nassiri-Khalilli MA, Ebrahimi M, Naderimanesh H (2004) Process development for production of recombinant human interferon-gamma expressed in *Escherichia coli*. J Ind Microbiol Biotechnol 31:63–69
88. Khushoo A, Pal Y, Mukherjee KJ (2005) Optimization of extracellular production of recombinant asparaginase in *Escherichia coli* in shake-flask and bioreactor. Appl Microbiol Biotechnol 68:189–197
89. Kim M, Elvin C, Brownlee A, Lyons R (2007) High yield expression of recombinant pro-resilin: lactose-induced fermentation in *E. coli* and facile purifcation. Protein Expr Purif 52:230–236
90. Kim YC, Kwon S, Lee SY, Chang HN (1998) Improved production of a bioadhesive precursor protein by fed-batch cultivation of a recombinant *Escherichia coli* with a pLysS vector. Biotechnol Lett 20:799–803
91. Kirk O, Borchert TV, Fuglsang CC (2002) Industrial enzyme applications. Curr Opin Biotechnol 13:345–351
92. Kleist S, Miksch G, Hitzmann B, Arndt M, Friehs K, Flaschel E (2003) Optimization of the extracellular production of a bacterial phytase with *Escherichia coli* by using different fed-batch fermentation strategies. Appl Microbiol Biotechnol 61:456–462
93. Kolaj O, Spada S, Robin S, Wall JG (2009) Use of folding modulators to improve heterologous protein production in *Escherichia coli*. Microb Cell Fact 8:9
94. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer WD (1995) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. J Biotechnol 39:59–65
95. Koyanagi S, Maeda T, Murakami T, Kawatsu K, Sugawara K, Miyatsu Y, Mizokami H (2008) Large-scale production of major house dust mite allergen der f 2mutant (C8/119S) in *Escherichia coli*. J Biosci Bioeng 106:387–392
96. Kumar A, Grover S, Sharma J, Batish VK (2010) Chymosin and other milk coagulants: sources and biotechnological interventions. Crit Rev Biotechnol 30:243–258
97. Kurland CG, Dong H (1996) Bacterial growth inhibition by overproduction of protein. Mol Microbiol 21:1–4
98. Kweon DH, Han NS, Park KM, Seo JH (2001) Overproduction of *Phytolacca insularis* protein in batch and fed-batch culture of recombinant *Escherichia coli*. Proc Biochem 36:537–542
99. Lee SY (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol 14:98–105

100. León-Rodríguez AD, Rivera-Pastrana D, Medina-Rivero E, Flores-Flores JL, Estrada-Baltazar A, Ordóñez-Acevedo LG, de la Rosa APB (2006) Production of penicillin acylase by a recombinant *Escherichia coli* using cheese whey as substrate and inducer. Biomol Eng 23:299–305
101. Lin HY, Mathiszik B, Xu B, Enfors SO, Neubauer P (2001) Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. Biotechnol Bioeng 73:347–357
102. Liu Y, Li XL, Zhang DD, Wang Q, Liu Y, Liu DM, Wu CT, Cui CP (2011) Production of hepatopoietin Cn in high-cell-density cultures of recombinant *Escherichia coli* and detection of its antioxygen activity. Mol Biotechnol 47:111–119
103. Lotti M, Porro D, Srienc F (2004) Recombinant proteins and host cell physiology. J Biotechnol 109:1–2
104. Luo Q, Shen YL, Wei DZ, Cao W (2006) Optimization of culture on the overproduction of TRAIL in high-cell-density culture by recombinant *Escherichia coli*. Appl Microbiol Biotechnol 71:184–191
105. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22:249–270
106. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev 60:512–538
107. March JC, Eiteman MA, Altman E (2002) Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. Appl Environ Microbiol 68:5620–5624
108. Martínez-Alonso M, García-Fruitós E, Ferrer-Miralles N, Rinas U, Villaverde A (2010) Side effects of chaperone gene coexpression in recombinant protein production. Microb Cell Fact 9:64
109. Martínez-Martínez I, Kaiser C, Rohde A, Ellert A, García-Carmona F, Sanchez-Ferrer A, Luttmann R (2007) High-level production of *Bacillus subtilis* glycine oxidase by fed-batch cultivation of recombinant *Escherichia coli* Rosetta (DE3). Biotechnol Prog 23:645–651
110. Meerman HJ, Georgiou G (1994) Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. Nat Biotechnol 12:1107–1110
111. Mergulhão FJM, Summers DK, Monteiro GA (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnol Adv 23:177–202
112. 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
113. Murby M, Uhlén M, Stähl S (1996) Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. Protein Expr Purif 7:129–136
114. Nakano K, Rischke M, Sato S, Märkl H (1997) Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. Appl Microbiol Biotechnol 48:597–601
115. Nanba H, Ikenaka Y, Yamada Y, Yajima K, Takano M, Takahashi S (1999) Production of thermotolerant *N*-carbamyl-D-amino acid amidohydrolase by recombinant *Escherichia coli*. J Biosci Bioeng 87:149–154
116. Narayanan N, Chou CP (2009) Alleviation of proteolytic sensitivity to enhance recombinant lipase production in *Escherichia coli*. Appl Environ Microbiol 75:5424–5427
117. Natale P, Brüser T, Driessen AJM (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. Biochim Biophys Acta 1778:1735–1756
118. Nevalainen KMH, Te'o VSJ, Bergquist PL (2005) Heterologous protein expression in filamentous fungi. Trends Biotechnol 23:468–474
119. Ni Y, Chen R (2009) Extracellular recombinant protein production from *Escherichia coli*. Biotechnol Lett 31:1661–1670
120. Noguchi Y, Satoh S, Yamaguchi M, Watanabe K, Hayashi M, Yamada H, Saito Y, Kobayashi M, Shimomura K (1996) An approach to high-level production of a mecasermin (somatomedin C) fused protein in *Escherichia coli* HB101. J Ferment Bioeng 82:128–133
121. Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 8:765–778
122. Ouellette T, Destrau S, Ouellette T, Zhu J, Roach JM, Coffman JD, Hecht T, Lynch JE, Giardina SL (2003) Production and purification of refolded recombinant human IL-7 from inclusion bodies. Protein Expr Purif 30:156–166
123. Ow D, Nissom P, Philip R, Oh S, Yap M (2006) Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5 α during batch fermentation. Enzyme Microb Tech 39:391–398
124. Palomares LA, Estrada-Mondaca S, Ramírez OT (2004) Production of recombinant proteins: challenges and solutions. Methods Mol Biol 267:15–52
125. Pandhal J, Wright PC (2010) N-Linked glycoengineering for human therapeutic proteins in bacteria. Biotechnol Lett 32:1189–1198
126. Parsell DA, Sauer RT (1989) Induction of a heat shock-like response by unfolded protein in *Escherichia coli*: dependence on protein level not protein degradation. Genes Dev 3:1226–1232
127. Pei X, Wang Q, Qiu X, Ying L, Tao J, Xie T (2010) The fed-batch production of a thermophilic 2-deoxyribose-5-phosphate aldolase (DERA) in *Escherichia coli* by exponential feeding strategy control. Appl Biochem Biotechnol 162:1423–1434
128. Poo H, Song JJ, Hong SP, Choi YH, Yun SW, Kim JH, Lee SC, Lee SG, Sung MH (2002) Novel high-level constitutive expression system, pHCE vector, for a convenient and cost-effective soluble production of human tumor necrosis factor-α. Biotechnol Lett 24:1185–1189
129. Porro D, Gasser B, Fossati T, Maurer M, Bruanduardi P, Sauer M, Mattanovich D (2011) Production of recombinant proteins and metabolites in yeasts: why are these systems better than bacterial production systems? Appl Microbiol Biotechnol 89:939–948
130. Preston A (2003) Choosing a cloning vector. Methods Mol Biol 235:19–26
131. 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 272:15661–15667
132. Qiao CL, Shen BC, Xing JM, Huang J, Zhang JL, Zhao DX, Yang B (2006) Culture and characteristics of recombinant protein production of an *Escherichia coli* strain expressing carboxylesterase Bl. Int Biodeterior Biodegrad 58:77–81
133. Rainczak K, Bajzert J, Galli J, Selera A, Wieliczko A, Borkowski J, Stefaniak T (2011) Optimization of *Salmonella enteritidis* recombinant heat shock protein 60 production. Pol J Vet Sci 14:145–146
134. Rathore AS, Latham P, Kaltenbrunner O, Curling J, Levine H (2004) Costing issues in the production of biopharmaceuticals: manufacturing costs are crucial to overall profit margins. Biopharm Int. http://biopharminternational.findpharma.com/biopharm/article/articleDetail.jsp?id=86832. Accessed 1 Feb 2004
135. Rawsthorne H, Turner KN, Mills DA (2006) Multicopy integration of heterologous genes, using the lactococcal group II

intron targeted to bacterial insertion sequences. Appl Environ Microbiol 72:6088–6093
136. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J (2001) Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. Science 294:158–160
137. Robinson C, Matos CFRO, Beck D, Ren C, Lawrence J, Vasisht N, Mendel S (2011) Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. Biochim Biophys Acta 1808:876–884
138. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307:249–264
139. Schmidt R, Bukau B, Mogk A (2009) Principles of general and regulatory proteolysis by AAA+ proteases in *Escherichia coli*. Res Microbiol 160:629–636
140. Semba H, Ichige E, Imanaka T, Atomi H, Aoyagi H (2008) Efficient production of active form of recombinant cassava hydroxynitrile lyase using *Escherichia coli* in low-temperature culture. Appl Microbiol Biotechnol 79:563–569
141. Seo JH, Kang DG, Cha HJ (2003) Comparison of cellular stress levels and green-fluorescent-protein expression in several *Escherichia coli* strains. Biotechnol Appl Biochem 37:103–107
142. Sevastsyanovich Y, Alfasi S, Overton T, Hall R, Jones J, Hewitt C, Cole J (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
143. Sharma R, Katoch M, Srivastava P, Qazi G (2009) Approaches for refining heterologous protein production in filamentous fungi. World J Microbiol Biotechnol 25:2083–2094
144. Shiloach J, Kaufman J, Guillard AS, Fass R (1996) Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by *Escherichia coli* BL21 (*λ*DE3) and *Escherichia coli* JM109. Biotechnol Bioeng 49:421–428
145. Shin CS, Hong MS, Bae CS, Lee J (1997) Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2]. Biotechnol Prog 13:249–257
146. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci U S A 89:8794–8797
147. Shuo-shuo C, Xue-zheng L, Ji-hong S (2011) Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948. Protein Expr Purif 77:166–172
148. Siguenza R, Flores N, Hernandez G, Martínez A, Bolivar F, Valle F (1999) Kinetic characterization in batch and continuous culture of *Escherichia coli* mutants affected in phosphoenol-pyruvate metabolism: differences in acetic acid production. World J Microbiol Biotechnol 15:587–592
149. Sinclair AM, Elliott S (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci 94:1626–1635
150. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99:303–310
151. Sivakesavaa S, Xua Z, Chena Y, Hacketta J, Huanga R, Lama E, Lama T, Siua K, Wonga R, Wong W (1999) Production of excreted human epidermal growth factor (hEGF) by an efficient recombinant *Escherichia coli* system. Proc Biochem 34: 893–900
152. Sleight SC, Bartley BA, Lieviant JA, Sauro HM (2010) Designing and engineering evolutionary robust genetic circuits. J Biol Eng 4:12
153. Solá RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. J Pharm Sci 98:1223–1245
154. Sonoda H, Kumada Y, Katsuda T, Yamaji H (2010) Functional expression of single-chain Fv antibody in the cytoplasm of *Escherichia coli* by thioredoxin fusion and co-expression of molecular chaperones. Protein Expr Purif 70:248–253
155. Srinivasan S, Barnard GC, Gerngross TU (2003) Production of recombinant proteins using multiple-copy gene integration in high-cell-density fermentations of *Ralstonia eutropha*. Biotechnol Bioeng 84:114–120
156. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol 32:1022–1030
157. Sørensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115:113–128
158. Takahashi M, Nordén B (1994) Structure of RecA-DNA complex and mechanism of DNA strand exchange reaction in homologous recombination. Adv Biophys 30:1–35
159. Tegel H, Ottosson J, Hober S (2011) Enhancing the protein production levels in *Escherichia coli* with a strong promoter. FEBS J 278:729–739
160. Tegel H, Tourle S, Ottosson J, Persson A (2010) Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta(DE3). Protein Expr Purif 69:159–167
161. 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
162. Thomas JG, Baneyx F (1996) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells over-expressing heat-shock proteins. J Biol Chem 271:11141–11147
163. Tseng TT, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the gene ontology. BMC Microbiol 9:S2
164. Vallejo LF, Brokelmann M, Marten S, Trappe S, Cabrera-Crespo J, Hoffmann A, Gross G, Weich HA, Rinas U (2002) Renaturation and purification of bone morphogenetic protein-2 produced as inclusion bodies in high-cell-density cultures of recombinant *Escherichia coli*. J Biotechnol 94:185–194
165. Vallejo LF, Rinas U (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact 3:11
166. van Beilen JB, Li Z (2002) Enzyme technology: an overview. Curr Opin Biotechnol 13:338–344
167. Varma A, Boesch BW, Palsson BO (1993) Biochemical production capabilities of *Escherichia coli*. Biotechnol Bioeng 42:59–73
168. Varma A, Boesch BW, Palsson BO (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59: 2465–2473
169. Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24:179–185
170. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, Vusser KD, Contreras R (2004) In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. Appl Environ Microbiol 70:2639–2646
171. Vethanayagam JG, Flower AM (2005) Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase. Microb Cell Fact 4:3
172. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-Linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298:1790–1793

173. Waegeman H, Beauprez J, Moens H, Maertens J, De Mey M, Foulquié-Moreno MR, Heijnen JJ, Charlier D, Soetaert W (2011) Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). BMC Microbiol 11:70
174. Walsh G (2010) Post-translational modifications of protein biopharmaceuticals. Drug Discov Today 15:773–780
175. Wardenga R, Hollmann F, Thum O, Bornscheuer U (2008) Functional expression of porcine aminoacylase 1 in *E. coli* using a codon optimized synthetic gene and molecular chaperones. Appl Microbiol Biotechnol 81:721–729
176. Warren CE (1993) Glycosylation. Curr Opin Biotechnol 4:596–602
177. Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, Siemann M, Altenbuchner J (2001) High-cell-density fermentation for production of L-*N*-carbamoylase using an expression system based on the *Escherichia coli rhaBAD* promoter. Biotechnol Bioeng 73:95–103
178. Wong MS, Wu S, Causey TB, Bennett GN, San KY (2008) Reduction of acetate accumulation in *Escherichia coli* cultures for increased recombinant protein production. Metab Eng 10:97–108
179. Wu J, Zhang L, Lei J, Cai G, Zhu W, Lu D, Jin J (2009) Enhancement of recombinant human ADAM15 disintegrin domain expression level by releasing the rare codons and amino acids restriction. Appl Biochem Biotechnol 157:299–310
180. Wu PH, Nair GR, Chu IM, Wu WT (2008) High cell density cultivation of *Escherichia coli* with surface anchored transglucosidase for use as whole-cell biocatalyst for alpha-arbutin synthesis. J Ind Microbiol Biotechnol 35:95–101
181. Xie L, Hall D, Eiteman MA, Altman E (2003) Optimization of recombinant aminolevulinate synthase production in *Escherichia coli* using factorial design. Appl Microbiol Biotechnol 63: 267–273
182. Xu R, Du P, Fan JJ, Zhang Q, Li TP, Gan RB (2002) High-level expression and secretion of recombinant mouse endostatin by *Escherichia coli*. Protein Expr Purif 24:453–459
183. Yan SS, Yan J, Shi G, Xu Q, Chen SC, Tian YW (2005) Production of native protein by using *Synechocystis* sp. pcc6803 DnaB mini-intein in *Escherichia coli*. Protein Expr Purif 40:340–345
184. Yang YT, Aristidou AA, San KY, Bennett GN (1999) Metabolic flux analysis of *Escherichia coli* deficient in the acetate production pathway and expressing the *Bacillus subtilis* acetolactate synthase. Metab Eng 1:26–34
185. Yildirim S, Konrad D, Calvez S, Drider D, Prévost H, Lacroix C (2007) Production of recombinant bacteriocin divercin V41 by high cell density *Escherichia coli* batch and fed-batch cultures. Appl Microbiol Biotechnol 77:525–531
186. Yoon S, Han MJ, Lee S, Jeong K, Yoo JS (2003) Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnol Bioeng 81:753–767
187. Yoon S, Kim S, Kim JF (2010) Secretory production of recombinant proteins in *Escherichia coli*. Recent Pat Biotechnol 4:23–29
188. Zhang D, Weib P, Fanc L, Liana J, Huanga L, Caia J, Xu Z (2010) High-level soluble expression of hIGF-1 fusion protein in recombinant *Escherichia coli*. Proc Biochem 45:1401–1405
189. Zhao JB, Wei DZ, Tong WY (2007) Identification of *Escherichia coli* host cell for high plasmid stability and improved production of antihuman ovarian carcinoma × antihuman CD3 single-chain bispecific antibody. Appl Microbiol Biotechnol 76:795–800