J Ind Microbiol Biotechnol (2013) 40:257–274
DOI 10.1007/s10295-013-1235-0

REVIEW

Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells

Aleš Berlec · Borut Štrukelj

Received: 5 November 2012 / Accepted: 22 January 2013 / Published online: 6 February 2013

Abstract Almost all of the 200 or so approved biopharmaceuticals have been produced in one of three host systems: the bacterium *Escherichia coli*, yeasts (*Saccharomyces cerevisiae*, *Pichia pastoris*) and mammalian cells. We describe the most widely used methods for the expression of recombinant proteins in the cytoplasm or periplasm of *E. coli*, as well as strategies for secreting the product to the growth medium. Recombinant expression in *E. coli* influences the cell physiology and triggers a stress response, which has to be considered in process development. Increased expression of a functional protein can be achieved by optimizing the gene, plasmid, host cell, and fermentation process. Relevant properties of two yeast expression systems, *S. cerevisiae* and *P. pastoris*, are summarized. Optimization of expression in *S. cerevisiae* has focused mainly on increasing the secretion, which is otherwise limiting. *P. pastoris* was recently approved as a host for biopharmaceutical production for the first time. It enables high-level protein production and secretion. Additionally, genetic engineering has resulted in its ability to produce recombinant proteins with humanized glycosylation patterns. Several mammalian cell lines of either rodent or human origin are also used in biopharmaceutical production. Optimization of their expression has focused on clonal selection, interference with epigenetic factors and genetic engineering. Systemic optimization approaches are applied to all cell expression systems. They feature parallel high-throughput techniques, such as DNA microarray, next-generation sequencing and proteomics, and enable simultaneous monitoring of multiple parameters. Systemic approaches, together with technological advances such as disposable bioreactors and microbioreactors, are expected to lead to increased quality and quantity of biopharmaceuticals, as well as to reduced product development times.

Keywords Biopharmaceutical production · *Escherichia coli* · Yeast · Mammalian cells · Optimization of expression

A. Berlec (✉) · B. Štrukelj
Department of Biotechnology, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: ales.berlec@ijs.si

B. Štrukelj
Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, 1000 Ljubljana, Slovenia

Introduction

The great majority of biopharmaceuticals that have been approved for therapeutic applications by regulatory authorities are proteins that have been produced by means of recombinant DNA technology in various expression systems. They constitute approximately one-sixth of the total pharmaceutical market and are its fastest growing segment [126]. In general, biopharmaceuticals are used to compensate for deficiency or lack of body proteins important for normal functioning of the organism. They can be divided mainly into the following categories: blood factors, thrombolytics and anticoagulants, hormones, enzymes, growth factors, interferons and interleukins, vaccines and monoclonal antibodies [126]. Of the 211 biopharmaceuticals that have gained regulatory approval by the end of 2011, 66 (31 %) were produced in *Escherichia coli*, 31 (15 %) in yeast (of those, 30 in *Saccharomyces cervisiae* and 1 in *Pichia pastoris*) and 91 (43 %) in mammalian cells [126, 127]. *E. coli*, yeast and mammalian cells together account for the production of 89 % of approved biopharmaceuticals and form the topic of this review. As this is a very broad subject, the authors have had to be highly

selective, and it has therefore not been possible to acknowledge many otherwise significant contributions.

The choice of expression system depends on the recombinant protein in question. A few examples of biopharmaceuticals that have been produced in *E. coli*, yeast or mammalian cells are shown in Table 1, together with some of the protein properties that were important in host selection. The least demanding expression system that still produces the protein at the desired quality and quantity, is normally used. *E. coli* is the simplest and cheapest expression system, but its use is limited due to the problems associated with correct folding and lack of post-translational modification. Yeasts are simple eukaryotes, and therefore advantageous in protein folding and post-translational modification, although in many cases with a nonoptimal glycosylation pattern. Mammalian cells are the most complicated and expensive to maintain, but enable the highest degree of protein quality and post-translational modification, such as glycosylation. The other 11 % of approved biopharmaceuticals are produced by hybridoma cells (some monoclonal antibodies), insect cells, transgenic animals, plants, or other hosts.

Characteristics of the *E. coli* expression system

In addition to its use for the expression of recombinant proteins, *Escherichia coli* is possibly the most thoroughly studied organism and its physiology is very well documented. In *E. coli*, recombinant proteins are expressed in the bacterial cytoplasm. They can remain there or be directed to other bacterial compartments, most often to the periplasm. They can be deposited in the inner or outer membrane, or (rarely) secreted to the growth medium. Each compartment has its own special characteristics which can be beneficial for recombinant protein expression and are regularly exploited for that purpose. Other characteristics can be disabling for recombinant expression, and strategies have been devised to overcome or avoid these obstacles. The properties of the cytoplasm, periplasm and

Table 1 Examples of biopharmaceuticals produced in *E. coli*, yeast and mammalian cells (from [126])

Therapeutic group	Recombinant protein	Host	Protein properties	
			Molecular weight (kDa)	Post-translational modifications
Blood factors, thrombolytics, anticoagulants	Factor VIII	Mammalian cells	267.0	Disulphide bonds, glycosylation, sulphation
	Tissue plasminogen activator	Mammalian cells, *E. coli* (fragment)	62.9	Proteolytic cleavage, disulphide bonds, glycosylation
	Hirudin	*S. cerevisiae*	7.0	Disulphide bonds, glycosylation, sulphation
Hormones	Insulin	*E. coli*, *S. cerevisiae*	12.0	Proteolytic cleavage, disulphide bonds
	Human growth hormone	*E. coli*, *S. cerevisiae*	24.8	Disulphide bonds, phosphoprotein
	Follicle-stimulating hormone	Mammalian cells	14.7 (subunit beta)	Disulphide bonds, glycosylation
	Glucagon	*E. coli*, *S. cerevisiae*	20.1	Amidation, proteolytic cleavage
Growth factors	Erythropoietin	Mammalian cells	21.3	Disulphide bonds, glycosylation
	Granulocyte-colony stimulating factor	*E. coli*, mammalian cells	22.3	Disulphide bonds, glycosylation
	Granulocyte–macrophage colony stimulating factor	*E. coli*	16.3	Disulphide bonds, glycosylation
Cytokines	Interferon-alpha	*E. coli*	21.5	Disulphide bonds, glycosylation
	Interferon-beta	*E. coli*	22.3	Disulphide bonds, glycosylation, phosphoprotein
Monoclonal antibodies	Infliximab	Mammalian cells	144.2	Disulphide bonds, glycosylation
Enzymes	Alpha-galactosidase	Mammalian cells	48.8	Disulphide bonds, glycosylation
	Deoxyribonuclease	Mammalian cells	31.4	Disulphide bonds glycosylation
	Uricase	*S. cerevisiae*	34.2	Acetylation

Protein properties that influence host selection include, among others, molecular weight and post-translational modifications, and were obtained from UniProt (http://www.uniprot.org/) and DrugBank (http://www.drugbank.ca/)

secretion in regard to recombinant expression are outlined below.

The cytoplasm of *E. coli*

Expression into the cytoplasm is the most straightforward method, and therefore is most frequently used. It nevertheless has several disadvantages, which include the presence of numerous cytoplasmic proteins that accompany the target protein during downstream processing, the inability to form disulphide bonds, the presence of proteases, and translation of the start codon to formylmethionine. Additionally, a very common situation in cytoplasmic expression is the association of yet-to-be stably folded proteins to form inclusion bodies. Inclusion bodies are electron-dense protein granules that are rod- or spherelike in shape, with diameters ranging from 0.2 to 1.2 μm and observable by optical microscopy [129]. They are largely inactive. Recent studies suggest that, although inclusion bodies appear amorphous, they contain beta strands in cross-beta structures that resemble amyloid-like fibrils [129, 130]. They also contain recombinant protein in native-like, partially folded or unstructured forms. Inclusion bodies are usually homogenous, mainly contain proteins of interest, and have little contaminating host protein or ribosomal components, DNA, or RNA [103]. The high density of the particles (1.3 mg/ml) facilitates their separation from other cellular components by centrifugation [120]. Their formation usually depends on the primary structure of the protein and is connected to the failure to form correct disulphide bonds and to the higher content of hydrophobic regions [83, 136]. It usually occurs at high levels protein production in which, consequently, there are insufficient levels of chaperones which assist in protein folding. The ability to predict the formation of inclusion bodies from the protein sequence has advanced, and computer tools have been developed [57, 58]. The formation of inclusion bodies can be advantageous, since they enable recombinant proteins to be obtained at relatively high yields and rapidly purified. They are appropriate for the production of otherwise toxic proteins and are mostly resistant to protease digestion [21]. Their disadvantage is the need to develop a refolding protocol, which can be tedious and not always effective in yielding native folded protein [21]. Inclusion bodies are solubilized with high concentrations (6–8 M) of chaotropic agents (urea, guanidinium chloride, guanidinium thiocyanate), detergents (sodium dodecyl sulphate, sarkosyl, TritonX 100) and reducing agents (dithiothreitol, dithioerythritol, cysteine) [114]. The simplest and most common method of refolding is the so called rapid dilution method, where the solubilized protein is added directly to the renaturation buffer. This requires large amounts of the latter, with subsequent concentration steps. The presence of pairs of oxidizing and reducing agents (glutathione, cysteine and cystamine) aids in the formation of disulphide bonds. Low molecular weight additives (acetone, acetamide, urea, detergents, sucrose, dimethyl sulfoxide, polyethylene glycol) often improve yields of functional proteins. L-arginine is commonly used since it reduces protein aggregation [114]. Additional refolding methods include dialysis, solid-phase separation, pulse renaturation, size-exclusion chromatography and adsorption chromatography [114]. Size-exclusion chromatography can enable removal of denaturant and folding at the same time.

Where the native protein contains disulphide bonds, their appropriate formation is essential. The *E. coli* cytoplasm is characterized by reducing conditions which prevent disulphide formation. Thioredoxins (TrxA, TrxC) and glutaredoxins (GrxA, GrxB, GrxC) rapidly reduce nascent disulphide bonds, while themselves become oxidized in the process. Thioredoxin reductase TrxB recycles oxidized TrxA and TrxC, while glutathione is responsible for the reduction of glutaredoxins [104]. Mutation in TrxB causes a reversal of TrxA and TrxC from reductases to oxidases, enabling cytoplasmic disulphide formation [118].

Another difficulty is the translation of the bacterial start codon—in 91 % cases AUG—into formylmethonine. Methionine is deformylated during synthesis of the protein, depending on the length of the side chain of the second amino acid residue and is thus not obligatory [85]. In the latter case, the modified N-terminal of the recombinant protein presents a problem in the production of therapeutic proteins, where a completely unaltered protein primary structure is required. Further, the protein's N-terminal may be important for its function.

Recombinant proteins are additionally susceptible to proteolytic degradation prior to their correct folding to the native state. Five ATP-dependent proteases are present in the cytoplasm of *E. coli*: Lon, ClpYQ/HslUV, ClpAP, ClpXP and FtsH. They degrade prematurely terminated polypeptides, sensitive folding intermediates and partially folded proteins. Lon and ClpYQ are thought to be the most active in the degradation in *E. coli* of non-native proteins [7].

Chaperones are folding modulators that help proteins achieve their native conformation, and can also be beneficial in recombinant protein expression. Their expression is often induced under conditions of stress, among them being recombinant protein expression. Chaperones bind to short sequences of consecutive hydrophobic amino acids, which are flanked by basic residues. They can be divided into three classes on the basis of their function. Folding chaperones DnaK and GroEL use ATP to drive conformational changes that cause refolding of their substrate proteins. "Holding" chaperones IbpA and IbpB hold partially folded proteins until folding chaperones become

available. Disaggregating chaperone ClpB promotes solubilization of aggregated proteins, which are then refolded by folding chaperones [7, 52].

The periplasm of *E. coli*

The periplasm is the compartment located between the inner and outer membranes. Its specific properties can be exploited for effective recombinant protein expression. Periplasm contains only about 4 % of the total cell protein, which makes it feasible for downstream processing. Proteins are translocated from cytoplasm to periplasm if they contain the appropriate leader peptide, which usually consists of 18–30 amino acids and contains two or more basic residues at the N-terminus, a central hydrophobic core and a hydrophilic C-terminus. The leader peptide can be derived from naturally secreted proteins such as OmpA, OmpT, PelB, beta-lactamase and alkaline phosphatase, and is cleaved by specific peptidases. This enables the expression of recombinant proteins with the authentic N-terminus, without the addition of methionine [6]. The majority of *E. coli* proteins are secreted by the Sec-dependent pathway in an unfolded form. The secretory chaperone SecB binds the protein and carries it to the membrane protein SecA, and the complex is then translocated through the membrane pore, formed by SecYEG. The alternative pathway includes the signal recognition particle (SRP), which recognizes highly hydrophobic sequences emerging from the ribosome and delivers the complex to FtsY membrane protein. This is followed by translocation of the protein through the SecYEG pore. The twin-arginine (Tat)-dependent secretion pathway can also secrete folded or partially folded proteins. The conserved secretion signal contains two arginine residues. The pathway consists of TatABC proteins [94], where TatC recognizes the substrate protein and TatA forms a pore.

Periplasmic recombinant expression is favourable due to the generally lower proteolytic activity in comparison to the cytoplasm [85]. However, two major proteases, DegP and Prc are present in the periplasm and digest misfolded proteins. Other proteases include DegS, DegQ, Protease III and OmpT [7].

The most prominent advantage of the periplasm over the cytoplasm is its oxidizing environment that facilitates the formation of disulphide bonds. Disulphide bond formation is catalyzed by Dsb proteins, which are thiol-disulphide oxidoreductases. DsbA is a soluble protein that oxidizes cysteines in the target protein and is later recycled by membrane protein DsbB. If incorrect disulphides form, they can be rearranged by the disulphide bond isomerase DsbC [7].

The main periplasmic chaperone is Skp which assists in the folding of proteins that emerge from the Sec translocation machinery. Other folding modulators include peptidyl-prolyl cis–trans isomerases SurA, FkpA, PpiA and PpiD [7].

Secretion from *E. coli* to the growth medium

Secretion is a desirable property in recombinant protein production due to the more straightforward subsequent downstream processing. *E. coli* has been traditionally regarded as an organism that is incapable of secreting proteins. Nevertheless, there are reports of extracellular proteins, secreted by an unknown mechanism, which were not released by cell lysis [93]. Four strategies have been applied to induce secretion of recombinant proteins from *E. coli* [95]. The first involves the use of engineered strains with dedicated secretion mechanisms that originate from pathogenic strains of *E. coli* or other Gram-negative bacteria. Six distinct secretion systems have been reported, Type I being the most popular. The *E. coli* haemolysin system is an example [95]. The second strategy involves the use of carrier proteins that are secreted by an unknown mechanism. The use of the IgG binding domain from *Staphylococcus* protein A [100] and YebF [143] as fusions have been demonstrated. The third strategy involves the use of mutants with defects in the outer membrane structure that cause increased permeability [95]. The potential drawback is the influence on cell growth. The fourth strategy involves the co-expression of lysis promoting proteins such as Kil [105].

The influence of recombinant expression on *E. coli* physiology

High yields are of paramount importance in recombinant protein expression. To achieve them, high-level gene expression and high cell density cultivation are applied. These two are sometimes mutually exclusive, however they both cause stress to producing cells and impact negatively on cellular physiology. Undesired physiological effects can be caused by intracellular and extracellular factors [23]. Intracellular factors include the presence of multicopy plasmids (causing metabolic burden and growth inhibition), potential toxicity of gene products, high-level gene expression (causing metabolic burden, stress and starvation) and protein misfolding (causing stress). Extracellular factors include accumulation of toxic metabolites (particularly acetate, which is toxic to *E. coli*), limitation of nutrients (causing starvation) and limitation of oxygen (which inhibits growth) [23]. Major types of physiological stress include heat shock, starvation and stationary-phase stress; other types include anaerobiosis, oxidative, osmotic and pH stress, or phage infection. Stress responsive proteins are induced under stressful conditions, and are

responsible for protection and repair of vital macromolecules. Transcription of stress responsive genes is regulated by sigma factors, which subsequently bind to their promoters. The most important sigma factors are σ^S (RpoS), σ^H (RpoH) and σ^E (RpoE). σ^S regulates the general stress response as cells enter the stationary phase or face a situation that causes growth arrest, such as starvation, osmotic shock or low pH. The response includes changes in cell morphology, protein content, gene expression and metabolism. σ^H induces cytoplasmic heat-shock response and σ^E extra-cytoplasmic heat-shock. Both result in the production of proteases, that degrade misfolded proteins, and chaperones, which assist in folding. Sigma regulons partially overlap and stress response can be induced via several sigma factors. All three sigma factors can be induced in recombinant protein production [23].

Optimization of recombinant expression in *E. coli*

Optimization of the gene

Variations in the gene nucleotide sequence may result either in amino acid substitutions or in the authentic protein product as a result of genetic code degeneracy. The former can be beneficial if the specific mutants (engineered therapeutic proteins) show improved properties such as increased stability, reduced aggregation or enhanced activity [68]. However, they can be problematic from the regulatory point of view. This can be exploited to optimize the availability of regulatory regions, stability of mRNA and codon usage. Further, the nucleotide sequence defines mRNA secondary structure which can affect the availability of the ribosome binding site or other regulatory regions and, consequentially, the yield of the recombinant protein [79]. The stability of mRNA depends on the resistance of 5′- and 3′-ends to exonucleases, and can also be controlled by the mRNA structure and sequence. Codon usage can differ significantly between different species. Rarely used codons (such as AGA and AGG for arginine) may inhibit translation [85] and decrease recombinant protein yield. A cluster of rarely used codons may lead to mis-incorporation of certain amino acids [116]. A recent systematic study of codon usage optimization revealed highly significant increases in recombinant protein yield. High expression was correlated with the use of codons that are highly charged during amino acid starvation but not with those that are abundant in highly expressed *E. coli* genes. Algorithms have been developed that enable rational design of highly expressed genes [134]. With the advance of custom gene synthesis and lowering of its cost, the optimization of gene nucleotide sequences has become readily available.

Optimization of the expression plasmid

Plasmids are the most common carriers of gene information in recombinant protein production. Insertion into the chromosome of *E. coli* has been reported [4], but not commonly used. Plasmids need to be stable, both structurally (no sequence changes during multiplication) and segregationally (presence in all bacteria during multiplication) [61]. *E. coli* strains can be engineered to improve the production of plasmid DNA [40]. Expression plasmids contain common elements responsible for controlled recombinant protein production. The origin of replication (ori) controls the replication rate of the plasmid and thereby the copy number. The most common are ColE1 and p15A. ColE1 is derived either from pBR322 plasmid (low copy number; 15–20) or from pUC plasmid (high copy number; 500–700) [116]. Plasmid copy number defines the recombinant gene dosage. High gene dosage can be either stimulating or detrimental for recombinant protein expression, depending on the specific effect on cell physiology. Resistance markers confer antibiotic resistance and enable stable plasmid maintenance. Ampicillin, kanamycin, chloramphenicol and tetracycline resistance markers are commonly used. Ampicillin may be problematic due to its susceptibility to degradation in the growth medium and can be substituted by carbenicillin [116]. The promoter controls the transcription of the target gene. It should be strong (capable of initiating a high level of transcription), tightly regulated (low level of basal transcription), transferable to a number of *E. coli* strains, and inducible in a simple and cost-effective manner [85]. Induction may be achieved by thermal change or with the addition of chemicals. Isopropyl-β-D-thiogalactopyranoside (IPTG) induction is the most common, but has certain drawbacks, mainly high price and cell toxicity [85]. Another two important elements are the ribosome binding site (RBS), which initiates ribosome binding, and the transcription terminator, which ends transcription, prevents run through transcription and stabilizes the 3′-end of mRNA. ATG start codon is the most effective in starting translation, and TAA stop codon, followed by T, in stopping translation [85].

The most often used plasmid series is the pET expression system, which is based on T7 promoter [119]. The host strain requires the T7 RNA polymerase-encoding DE3 phage fragment, which is under the control of *lacUV5* promoter. The latter can be induced by the addition of IPTG, which binds and removes LacI repressor. T7 RNA polymerase is faster than native *E. coli* polymerase and transcribes the gene under the control of T7 promoter, which is present on the plasmid. Tight control of transcription is achieved by the presence of T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase. The pBAD expression system, which is based on *araBAD*

promoter, is representative of the new and improved series of plasmids. Induction is triggered by L-arabinose and enables graded expression of the target protein, based on the L-arabinose concentration [44].

Appropriate plasmid design enables expression of recombinant proteins fused with another protein partner. Fusion proteins are usually highly expressed proteins that can increase the overall yield of the recombinant protein, possibly by stabilizing the mRNA [2]. The fusion tag can be inserted in the N-, or C-terminal part of the fused protein. The most commonly used fusion proteins are thioredoxin (TRX), glutathione-S-transferase (GST) and maltose binding protein (MBP). They can also assist in folding and disulphide formation and in protection against proteolysis, thereby increasing the yield of soluble protein. They can be exploited in affinity purification, which facilitates downstream processing. The most popular affinity fusion is the addition of a peptide consisting of at least six histidine residues (His Tag), which can be used in immobilized metal affinity chromatography. Insertion of protease recognition sequences between the recombinant protein and its fusion partner enables cleavage with specific peptidases (e.g., thrombin or factor Xa) and removal of the fusion protein [116].

Optimization of the producer strain

Producer strains should have low levels of proteolytic activity, enable stable maintenance of expression plasmids, and contain genetic elements, which are important for expression systems, or increase protein yield and solubility. BL21 is an *E. coli* B strain widely used in recombinant expression. It lacks Lon and OmpT proteases which often interfere with recombinant protein production. Other modifications include *recA* negative strains, which stabilize expression plasmids, and *trxB/gor* negative strains that enhance disulphide bond formation. LacY mutants enable adjustable levels of protein expression [116]. Supplementation with tRNA genes for rarely transcribed codons (usually on another plasmid) can result in higher protein yield [25]. Expression of chaperones (GroEL-GroES and DnaK-DnaJ-GrpE) can assist in protein folding and prevent inclusion body formation [96]. Manipulation or prevention of stress response can be achieved by interfering with sigma factors. *RpoS* mutation had a positive effect on recombinant protein expression when using fed-batch fermentation [63]. *E. coli* cells have been metabolically engineered to decrease detrimental acetate production during growth, for example by over-expression of phosphoenol pyruvate carboxylase [30]. Synthetic biology was employed to reduce the *E.* coli genome up to 15 % by removing the unnecessary genetic elements [102]. This led to improved electroporation efficiency and plasmid stability, and did not compromise recombinant protein expression [102, 112]. Transfer of *Campylobacter jejuni* glycosylation machinery to *E. coli* enabled the production of glycosylated recombinant protein [125]. This opened up possibilities for glycosylation engineering in *E. coli*.

Optimization of the fermentation procedure

Higher protein production can also be achieved with the use of higher concentrations of producing cells, such as in high-cell density culture systems. This is more cost-effective and environmentally friendly. Different fermentation strategies, such as batch, fed-batch and continuous fermentation, can be applied and yield more than 100 g of dry cell mass per litre of fermentation broth. Fermentation parameters that can be controlled include growth medium composition, temperature, pH and dissolved oxygen. The composition of growth medium has to be controlled in order to sustain high cell growth and, at the same time, prevent unwanted metabolic effects. Glycine is an additive to the growth medium specific for the release of periplasmic proteins. Fermentation parameters can also have an influence on proteolytic activity, secretion and the level of expression.

High cell density culture systems have several disadvantages, including lower availability of dissolved oxygen, higher levels of carbon dioxide (which decrease growth rate and increase acetate production), reduced stirring efficiency and increased heating [81]. This results in generally lower specific protein production than in shake flask cultures [64]. Stirring and heat generation are especially problematic in large fermentors and have to be considered in scale-up design. Nutrient feeding strategy is critical in high cell density cultures, and constant rate feeding, increased feeding or exponential feeding can be selected [21]. Exponential feeding enables growth at a constant specific growth rate by using glucose as a growth-limiting nutrient. Acetate production can be minimized by keeping specific growth rate between 0.2 and 0.35 h^{-1} and by keeping glucose concentration near zero [21]. To prevent overfeeding of nutrients, more sophisticated feedback-controlled feeding systems can be used, in which a certain amount of nutrient is added in response to a change in pH or dissolved oxygen [21]. It is often useful to separate growth and production phases. This can be achieved by delaying induction until the appropriate cell density is reached. High cell density culture systems can cause a stress response as described previously. They can also lead to cell filamentation, which should be avoided [65].

Novel systemic optimization approaches

Numerous approaches to optimizing recombinant expression in *E. coli* have been described and the most common have been included in this review. Several approaches should be tested for each recombinant protein—it should

be borne in mind that this is necessarily a trial-and-error approach. Conditions that work with one protein can be detrimental for another, and no general rule can be set.

Nevertheless, the great advantage in system-based approaches in recent years has enabled precise, target-oriented optimization of expression of a particular protein, mainly by strain engineering. Proteome and transcriptome data, obtained by 2D gel electrophoresis and DNA arrays, respectively, reveal proteins that are either up- or down-regulated during recombinant expression and may serve as a target in the optimization procedure [33, 45]. Differentially expressed genes may be grouped in the following categories: heat shock and stringent response, phage-related, elongation factors, ribosomal proteins, amino acid biosynthesis and tRNA-related, transposon-related, nucleotide biosynthesis and tricarboxylic acid cycle [124]. Successful examples of such approaches include, among others, co-expression of down-regulated *glpF* and *prsA* genes in the expression of insulin-like growth factor I fusion protein [22], co-expression of cysteine synthase A in the expression of leptin [49], and co-expression of phage shock protein A in the expression of humanized antibody fragment [1].

Characteristics and optimization of yeast expression systems

Yeasts are unicellular organisms and are among the simplest eukaryotes. They have a sub-cellular organization similar to that of higher eukaryotes and contain a nucleus, mitochondria, endoplasmic reticulum (ER), Golgi apparatus, secretory vesicles, vacuoles and microbodies. Baker's yeast (*Saccharomyces cerevisiae*) has a special status since it has been used as a component of human diet for centuries. It has been accepted as "generally recognized as safe" (GRAS), which is beneficial in the production of recombinant pharmaceutics from the regulatory point of view. In addition, its molecular and biochemical properties are well known, as well as the manipulation techniques. Other yeasts include *P. pastoris*, *Hanensula polymorpha*, *Yarrowia lypolytica*, *Kluyveromyces lactis* and *Schizosaccharomyces pombe*. *Pichia pastoris* is probably the second most frequently used yeast. It is widely applied for recombinant expression in research labs and on the industrial scale, since it offers higher production and secretion capability than *S. cerevisiae*. The first biopharmaceutical produced in *P. pastoris* was approved in 2009 and several others are undergoing clinical studies.

Characteristics of the *S. cerevisiae* expression system

S. cerevisae embodies some characteristics of prokaryotes (fast growth in cheap media, simple genetic manipulation) and of eukaryotes (proteolytic processing, folding, disulphide bond formation, and post-translational modifications without carboxylation). On the other hand, it is unable to reach high cell densities, and exhibits limited secretion and excessive/irregular glycosylation.

Transformation of *S. cerevisiae* cells with expression vectors can be achieved by treating cells with lithium ions. Expression vectors are based mainly on 2 μ plasmids. They are usually *E. coli*/yeast shuttle vectors, which enable their propagation in *E. coli*. 2 μ plasmid is present at about 100 copies per haploid genome. It contains four genes (*FLP*, *REP1*, *REP2* and *D*), origin of replication, *STB* locus and two inverted repeats. *FLP* encodes a site-specific recombinase that assists in achieving higher copy number and is therefore beneficial in recombinant gene expression [107]. Another important factor in recombinant expression is the wide variation in the productivity of different transformants, which therefore have to be tested.

Auxotrophic selection markers are commonly used to select the plasmid-containing transformants. *LEU2*, *TRP1*, *URA3* and *HIS3* genes are used with complementary mutant strains, being auxotrophic for leucine, tryptophan, uracil and histidine, respectively. Minimal growth media which lack the relevant nutrient are used in selection. Dominant markers can be used for selection in rich medium and include resistance to antibiotics G418, hygromycin B and chloramphenicol. Copper resistance is conferred by *CUP1* gene, which can therefore also be used as a dominant marker [107].

The absence of bacterial DNA is advantageous from the regulatory point of view. Bacterial DNA can be removed by integration of the expression cassette into the native 2 μ plasmid, or by the use of a shuttle plasmid and subsequent removal of the bacterial sequence by in vivo recombination [107]. Additionally, more stable maintenance of foreign DNA can be achieved by the use of integrating vectors YIp. These contain yeast chromosomal DNA to target integration, a selectable marker and a bacterial replicon. In order to achieve integration in multiple copies, integration can be targeted to reiterated chromosomal DNA (tandem repeats) of, e.g., ribosomal DNA cluster [82] or *Ty* transposable element [73].

Promoters and terminators are essential for successful recombinant expression. Yeast promoters consist of at least three elements: upstream activation sequences (UAS), TATA elements and initiator elements. Yeast promoters may be highly complex, with lengths of over 500 bp, and can contain multiple UAS, multiple TATA elements and negative regulatory sites. Glycolytic promoters of alcohol dehydrogenase I (*ADH1*), phosphoglycerate kinase (*PGK*) and glyceraldehyde-3-phosphate dehydrogenase (*GAP*) are among the strongest promoters. They are induced by the addition of glucose; however the induction rate is low. Poor

regulation makes these promoters less suitable for use. The strongest tightly regulated promoters are those of galactose regulated genes *GAL1*, *GAL7* and *GAL10*, which are involved in the metabolism of galactose. These promoters are rapidly induced by more than 1,000-fold upon addition of galactose [117] and are strongly repressed by glucose. Glucose therefore has to be depleted from the growth medium before induction can take place. Promoters can also be regulated by inorganic phosphate concentration, as in the case of acid phosphatase (*PHO5*). Promoters that are independent of culture nutrients are potentially useful. Among those, temperature-regulated systems and a copper ion–inducible system (promoter of *CUP1* gene) are worth mentioning [107].

Optimization of the *S. cerevisiae* expression system

Since recombinant protein expression in *S. cerevisiae* has been well established for some time, very little progress has been made in the optimization of expression vectors and promoters in the last two decades. Optimization has rather focused on one of the weaknesses of *S. cervisiae* as a recombinant expression host—the ability to secrete recombinant protein. Parameters that can be optimized to increase secretion efficiency are: cultivation conditions, vector system, promoter, codon usage, secretion leader sequences, processing and folding [56]. It has been postulated that the secretion efficiency of *S. cerevisiae* is 100-fold to 1,000-fold lower than that predicted theoretically [59].

Significant increase in secretion levels has been achieved by optimizing the fermentation process. Control of pH and temperature increased cell density and reduced protein degradation [59]. Advanced approaches of secretion optimization have focused on strain engineering, in which crucial steps in protein secretion are modified on the genetic level. These strategies also apply to other yeasts besides *S. cerevisiae*, and encompass protein folding in the ER, intracellular vesicular protein trafficking and proteolytic degradation of the protein [56, 59].

The ER lumen contains numerous proteins responsible for correct folding of the protein, processing of signal sequences, formation of disulphide bonds, etc. Protein folding is under stringent control, and several mechanisms are employed in preventing misfolding or for removal of misfolded proteins. Over-expression of chaperones is a logical strategy in increasing correct protein folding and, consequently, protein secretion. Chaperone BiP assists in protein folding; however, prolonged binding of BiP results in an unfolded protein response and, finally, in the translocation of the misfolded protein to the cytosol, where it is degraded. Thus, over-expression of BiP has ambiguous effects, depending on the protein which is to be expressed. BiP over-expression resulted in a 26-fold increase in bovine prochymosin secretion [51] and a 5-fold increase in erythropoietin secretion [106], but decreased the expression of glucose oxidase in *H. polymorpha* [122]. Single or multiple over-expression of chaperones Jem1p, Sil1p, Lhs1p and Scy1p resulted in increased secretion of several pharmaceutically important proteins [101]. Over-expression of protein disulphide isomerase, either alone or in combination with BiP, also increased secretion [113, 115]. Another strategy involves manipulation of the unfolded protein response and its regulator Hac1p which activates expression of chaperones [59].

The intermediary step in protein secretion involves trafficking of protein in membrane-enclosed transport vesicles from endoplasmic reticulum (ER) to Golgi, inside Golgi, and out of Golgi. Ineffective trafficking results in intracellular accumulation of the protein and impairs secretion. A protein has to contain a signal sequence (cleaved in later steps) to enter the ER. MFα1 signal peptide from α-factor mating pheromone [12] and, recently, signal peptide from viral K28 preprotoxin [29] have been used for ER targeting. Vacuolar mis-sorting results in intracellular retention and is mediated by vacuolar protein sorting receptor Vps10p. Deletion of the *vps10* gene results in improvement of protein secretion in some cases [54] while, in others, deletions of *vps4*, *vps8*, *vps13*, *vps35* and *vps36* were beneficial [142]. Deletion of *MON2* gene, that encodes scaffold protein for vesicle formation, increased secretion of recombinant luciferase [69]. Over-expression of syntaxins Sso1p and Sso2p, which function as fusion targeting proteins for vesicles on the plasma membrane, also increased expression [108]. This shows the importance of regulating trafficking from Golgi to plasma membrane. In general, the regulation of vesicular transport is difficult due to the plethora of genes that are involved—further studies are therefore needed.

After secretion, proteins can be degraded by proteases that are secreted by host cells. Protease-deficient strains are therefore expected to yield more secreted protein. Deletion of the major vacuolar protease genes *PEP4* and *PRB1* impairs maturation and activation of other vacuolar proteases [66], resulting in significantly lower proteolytic activity. Other protease genes which were also deleted are *CPY1*, *YPS1* and *KEX2* [59]. Deletion of the mitochondrial metalloendoprotease gene CYM1 also decreases proteolysis and improves secretion [67]. Generation of strains deficient in multiple proteases is expected to further improve secretion efficacy, as in the case of *Schizosaccharomyces pombe* [60].

As in *E. coli*, future studies will not concentrate on the study of a single protein or group of proteins, but rather involve genome-, proteome-, transcriptome- and/or metabolome-wide information. cDNA over-expression libraries have already been analyzed using high throughput screening

approaches, including flow cytometry and cell sorting, as well as DNA microarrays, to identify factors which may contribute to secretion [37, 88].

Characteristics of the *P. pastoris* expression system

Pichia pastoris is a methylotrophic yeast, which indicates its ability to utilize methanol as a sole carbon and energy source. Initial metabolic reactions take place in peroxisomes and involve alcohol oxidase, catalase and dihydroxyacetone synthase [19]. Genetic manipulation techniques (high-frequency transformation by electroporation, gene targeting, stable integration into the genome, cloning by functional complementation) have been well established in *P. pastoris*, making it an attractive host for recombinant expression. Proteins can be expressed at high levels and successfully secreted. This makes purification relatively straightforward due to low amounts of host secreted proteins [132]. *P. pastoris* can perform eukaryotic protein modifications, including glycosylation, formation of disulphide bonds and proteolytic processing [84]. Unlike *S. cerevisiae*, *P. pastoris* can grow to very high cell densities (more than 130 g/l of dry cell weight) on minimal media [132].

Expression vectors, such as *E. coli/P. pastoris* shuttle plasmids, have been designed. They contain origins of replication and selection markers that are functional in both organisms. Selection markers are similar to those in *S. cerevisiae*, and are either auxotrophic (*HIS4*, *ARG4*, *ADE1*, *URA3*) or dominant (zeocin) [19]. To increase the stability of expression strains, vectors can be integrated into the genome. This can be performed by digesting the vector with the restriction enzyme in the marker gene or promoter region and transforming it into an appropriate auxotrophic strain, resulting in high-frequency, single cross-over integration.

Alcohol oxidase (AOX) catalyzes the first step of the utilization of methanol. Its expression is highly induced (from practically undetectable to more than 30 % of total soluble protein) on the addition of methanol. *AOX1* promoter has therefore been the most widely applied promoter in recombinant protein expression in *P. pastoris*. Alternative promoters, that do not require methanol for induction, include *GAP* promoter (which enables strong constitutive expression on glucose), *FLD1* promoter (which can be induced by either methanol or methylamine) and *PEX8* promoter (when less intensive transcription is required). *AOX2* promoter is approximately 10–20 times less active than *AOX1* promoter, but has nevertheless been used successfully in certain cases of recombinant protein production [19, 84].

Optimization of the *P. pastoris* expression system

Higher yields of recombinant protein in *P. pastoris* can be achieved by inhibiting proteolytic degradation. Similarly to *S. cerevisiae*, strains of *P. pastoris* have been created that lack protease genes *PEP4* and *PRB1* [19]. Another strain, deficient in Kex1 protease, was applied in the expression of endostatin [11]. The downside of protease deficient strains is their lower viability, slower growth and resistance to transformation [19].

Much effort has been put into optimizing fermentor conditions. Growth in a fermentor is necessary for the control of pH, temperature, aeration and feeding. The pH has to be optimized for each protein. Growth at lower temperatures has been shown to be desirable for certain proteins [84]. During fermentation, the methanol feeding process is usually used. It consists of three stages [18]. In the first, the expression strain is grown in batch fashion in a repressing carbon source, usually glycerol. The second stage includes transition to fed-batch culture. Glycerol is fed in a growth-limiting way to increase the biomass. The third stage is induced with low-rate methanol feeding. The methanol feed rate is increased after the culture becomes adapted to methanol [18]. An alternative fermentation approach is mixed feeding, where glycerol and methanol are fed simultaneously. This can result in improved viability, shorter induction and higher production rate; however, repression of AOX1 promoter by glycerol may be problematic [18]. Therefore, the feed rate has to be carefully optimized and monitored.

Another characteristic that has recently received much research attention is the ability to glycosylate expressed recombinant proteins. This is especially relevant in the production of pharmaceuticals, many of which require authentic glycosylation in order to be biologically active. Yeasts can glycosylate produced proteins; however the glycosylation pattern consists of many mannose residues and differs significantly from that in humans. High mannose content can result in faster degradation of the administered protein [38], resulting in lower activity. A largely successful effort has therefore been made to genetically engineer a strain with the human *N*-glycosylation pattern. The initial glycosylation, which occurs in the ER, is the same in yeast and humans. In both, proteins that arrive at the Golgi carry $Man_8GlcNAc_2$ glycan. In humans, this structure is first trimmed to $Man_3GlcNAc_2$ then converted to $GlcNAc_2Man_3GlcNAc_2$, galactosylated and finally sialylated to $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$. In yeast, however the $Man_8GlcNAc_2$ intermediate is not trimmed, but further mannosylated, resulting in the generation of glycans with more than 30 mannose residues [48]. The first step in humanization therefore involves the deletion of *och1* gene (α-1,6-mannosyltransferase), which is responsible for the transfer of another mannosyl residue to $Man_8GlcNAc_2$ intermediate, creating a substrate for further mannosyltransferases [48]. Simultaneous deletion of *och1* and *mnn1* genes (the latter encoding α-1,

3-mannosyltransferase) produced a strain with a single glycan, $Man_8GlcNAc_2$, that acts as a starting point for further humanization [92]. The subsequent steps included introduction of glycosyltransferase and glycosidase genes (*MnsI*, *GnTI*, *GnTII*, *SiaT*) to *P. pastoris* by screening combinatorial libraries of transmembrane domains of known Golgi- and ER-localized proteins, and various glycosyltransferases and glycosidases. This high-throughput approach resulted in a strain capable of producing $Gal_2GlcNAc_2Man_3GlcNAc_2$ intermediate [10, 20, 46]. The final step of transferring the sialic acid was complicated by the fact that this sugar is not produced in yeast cells. Enzymes responsible for sialic acid biosynthesis were therefore introduced to *P. pastoris*, together with sialyl-transferase. The use of this strain resulted in the production of erythropoietin with a humanized glycosylation profile [47]. O-glycosylation was also tackled in *S. cerevisiae* by introducing mucin-type glycosylation and O-fucosylation. Humanization of the glycosylation machinery in yeasts opens up new possibilities for the production of biopharmaceuticals.

Characteristics of mammalian cell expression systems

Mammalian cells nowadays constitute the most often used expression system for biopharmaceuticals, two-thirds of the revenue being from those produced in mammalian cells [145]. The majority of approved proteins have been produced in Chinese hamster ovary (CHO) cells. Other rodent cells include baby hamster kidney (BHK) cells and mouse myeloma (NS0) cells. Cells of human descent have also been applied, including human embryonic kidney (HEK-293) cells and human-retina-derived cells (PER-C6). All these cells can be grown in suspension culture in bioreactors and are amenable to scale-up [15]. The reason for the popularity of mammalian cells is their superior ability to perform post-translational modifications, which are sometimes a prerequisite for the biological activity of recombinant proteins. Post-translational modifications performed by mammalian cells include proteolytic processing, disulphide bond formation, glycosylation, γ-carboxylation, β-hydroxylation, *O*-sulphation and amidation [128]. Of these, proteolytic processing and disulphide bond formation can also be achieved in other expression systems. Glycosylation has received the most attention and its advances will be described in more detail below. The sugar moiety of the glycoprotein was shown to be important in protein folding, protein targeting and trafficking, ligand recognition and binding, biological activity, stability, protein half-life, and in evoking an immune response [128].

Recombinant genes can be delivered to mammalian cells (transfection) in several ways. Viral delivery is generally not desirable due to regulatory restraints. Non-viral delivery methods are therefore favoured, and include treatment of cells with calcium phosphate, electroporation, lipofection, or polymer-mediated gene transfer [15]. Upon delivery, plasmid DNA is linearized by endo- or exonucleases and inserted randomly into the host cell genome by the action of recombinases. The efficacy of insertion can be improved by linearizing the plasmid prior to transfection. Ligation of plasmids prior to integration can cause integration of several plasmid molecules to the genome [138]. Transient gene expression enables gene expression for a shorter period of time, in contrast to stable gene expression. Transient expression involves shorter process development and faster access to recombinant protein. Such development of appropriate vectors and processes for transient expression has enabled a considerable increase in the yield of recombinant proteins. In the future, transient gene expression could be considered for smaller scale biopharmaceutical production or in personalized medicine [5].

Stable gene expression is achieved by the use of appropriate selection systems. This may include resistance to certain antibiotics such as neomycin, hygromycin or puromycin. However, the most common selection systems are based on dihydrofolate reductase (DHFR) gene in combination with CHO cells, which lack DHFR activity, or glutamine synthase (GS) in NS0 cells. Selection is achieved by the absence of the appropriate metabolite in the medium composition (hypoxanthine and thymidine for DHFR, or glutamine for GS) [138]. The selection gene may be present on the same plasmid as the gene for recombinant protein, or on a separate plasmid. Promoters that drive gene expression are usually strong promoters of viral origin [138].

The composition of cell culture media is of utmost importance in enabling sustained culture growth and effective recombinant protein production. Fetal bovine serum used to be an essential component of mammalian cell growth media. It has since been disregarded by regulatory authorities due to its uncharacterized nature and the fear of transfer of adventitious agents [138]. Instead, chemically defined media are used that contain a pool of peptides, growth factors, proteins, lipids, carbohydrates and small molecules. Biopharmaceutical producers use their own proprietary mixtures that are carefully optimized for a particular cell type or cell mutant. Even for a single cell line, different media can be used during the production process and can be adapted specifically to the phases of cell subcultivation or recombinant protein production [138]. Cell lines can have specific nutrient requirements; for example, NS0 cells require the presence of cholesterol since they lack the cholesterol synthesis pathway [41].

Cells can be grown in adherent culture or in the much more common suspension culture. Adherent cultures can be grown in partially filled roller bottles, which ensure the

required availability of nutrients and oxygen. Another option is adherence to spherical microcarriers maintained as a suspension in stirred tank bioreactors [138]. Adherent cell growth is the default cell growth mode and the transition to suspension growth has to be achieved through the use of appropriate growth medium and clonal selection. Suspension growth can result in much higher cell densities and is the preferred culturing mode in biopharmaceutical production today. Suspension cultures can be grown in batch, fed-batch or perfusion modes. Batch and fed-batch modes are relatively simple. They are performed in stirred tank bioreactors or airlift bioreactors. The reactors are designed in such a way as to minimize the shear forces that inflict damage to the cells. Batch culture is characterized by the successive dilution of a cell culture in the growth medium, which usually occurs over three phases: seed, inoculum and production. This enables a gradual increase in cell density. Fed-batch culture is an improved batch culture that involves feeding nutrients according to cell requirements [9]. Slow feeding of nutrients results in their maintenance at low concentrations. This in turn increases the rate of metabolism and decreases the production of detrimental by-products such as ammonia or lactate. The production phase of fed-bed culture can be further divided into two phases (biphasic strategy). In the first phase, the cells grow to high cell density; in the second, cell growth is arrested by lowering the temperature, which increases specific productivity [141]. Perfusion cultures demand a different, more complex and expensive bioreactor set-up, which has to concomitantly feed fresh medium and remove used culture medium with the product. Perfusion cultures can achieve very high cell densities and can be sustained for prolonged period of time, even for several weeks or months [90]. Due to these reasons the bioreactor volumes can be smaller [109].

Improvement of mammalian cell expression

The expression of recombinant proteins in mammalian cells has been improved with regard to quality and quantity. The importance of higher protein yield is obvious, while protein quality is usually related to post-translational modifications, most commonly to ensure correct and uniform glycosylation. Improvements have been made in the selection of clones with the highest specific productivity, in understanding and influencing epigenetic factors on gene expression, and in genetic engineering of producer cells. All this has been assisted by genome scale technologies such as genomics, transcriptomics and proteomics. Improvements specific for the protein (e.g., signal peptides [75]), as well as improvements in medium composition and process parameters have also been made, but will be mentioned only briefly in this review.

Improvements in clonal selection

Transformed cell lines are significantly heterogeneous, so a single cell has to be isolated to produce a clonal isolate for biopharmaceutical production in order to comply with regulatory requirements. At the same time, several hundred cell lines have to be screened for maximal specific productivity and reasonable viability. Clonal selection of adherent cells is relatively easy, since individual colonies can be observed after growth on the surface of, e.g., multiwell plates. For suspension cultures, the limiting dilution method (LDC) is still among the most commonly used. It involves dilution of the cell suspension to a density that ensures less than one cell per well upon transfer to a microtiter plate. Wells containing only a single cell are identified by observation under a microscope. Those that are able to proliferate are tested for recombinant protein yield [14]. The method has been automated [135], but still has many drawbacks, including the relatively low number of tested clones. Several rounds of LDC are needed to confirm the single clone isolation.

Screening capacity has been increased by introducing flow cytometry and cell sorting, which enable monitoring of protein expression on a single cell level. To monitor the protein expression, the protein has to be labelled by co-expression with a fluorescent reporter protein, such as green fluorescent protein [89]. Alternatively, the required protein can be labelled with a specific antibody; however, in that case the protein has to be displayed on the cell surface. Surface association has been reported to correlate with level of expression and secretion [13], but this does not apply to all proteins. The protein therefore has to be retained on the surface in another manner, either by gel microdrop technology or matrix-based secretion [53, 133]. In the former, the cell is captured in a gel matrix droplet, in which antibody against the recombinant product is incorporated [131]. The latter approach is similar, only the matrix is formed directly on the cell surface to which the antibody molecules are bound via biotin. Diffusion is reduced by the use of high-viscosity medium [53].

Some novel screening methods also use fluorescence and are highly automated. These include laser-enabled analysis and processing, where unwanted cells are removed by laser treatment [50], and automated colony picking; promising cells are picked and transferred to new wells. Completely automated and robotically manipulated systems are commercially available [14].

Improvements in overcoming epigenetic factors

The specific productivity of a given cell depends on the number of gene copies inserted in the genome, as well as on the site of integration. The strategy to increase the gene

copy number involves the use of inhibitors of DHFR (methotrexate) and GS (methionine sulphoximine). The insertion of several hundred or thousand gene copies can be achieved by increasing the inhibitor concentration [139]. The inhibitors should, however, only be used during the selection of appropriate producer clone, because they promote cellular heterogeneity [138].

As noted above, the site of gene integration is of equal importance, since gene copy number depends on the so-called position effect. The gene may be inserted in the transcriptionally inactive region of genomic DNA (heterochromatin), which results in epigenetic gene silencing. The state of transcriptional activity depends largely on histone modifications, most notably acetylation, methylation, phosphorylation and ubiquitination of lysine and serine residues in the N-termini of histones. In general, acetylation of histone residues correlates with increased gene expression and methylation with decreased expression [86].

The histone modification status can be modified in two ways to increase expression. First, the histone deacetylase inhibitor, butyrate, can be added to the growth medium [99]. This is a nonspecific approach, and can also result in growth inhibition and apoptosis. Successful use of this approach has nevertheless been demonstrated for several proteins in CHO cells [77]. Secondly, histone acetyltransferase can be targeted to the promoter of the recombinant gene, which ensures a more targeted action. This can be achieved by fusing histone acetyltransferase to DNA-binding protein LexA, and insertion of LexA binding sites upstream of the promoter [78]. Thirdly, the transcriptional activity of a DNA region can also be improved by adding cis-acting DNA regions to the recombinant gene construct. These include a β-globin locus control region (only active in erythropoietic cell lines), insulators (such as cHS4, which blocks the action of the enhancer region on the promoter), ubiquitous chromatin opening elements (promoter regions of ubiquitously expressed housekeeping genes such as *HNRPA2B1* and *CBX3*), matrix associated regions (which bind to nuclear matrix and influence chromatin structuring) and anti-repressor elements (which block heterochromatin repression) [77].

Another approach to overcome epigenetic regulation involves specific targeting to the sites of the genome where high expression occurs (hot spots), rather than random integration. The Flp/Frt system has been successfully used for that purpose [144], while the Cre/LoxP system resulted in a limited copy number of inserted genes [34, 74]. Another recombination system is ΦC31 system which catalyzes irreversible recombination between attP and attB sites [16].

As an alternative to chromosome integration, an artificial chromosome expression system, based on mammalian artificial chromosomes, has been considered [71]. Artificial chromosomes contain multiple acceptor sites, removing the need for genome integration, and improve the transfection process [71].

Improvements in genetic engineering of mammalian cells

The productivity of mammalian cells, in terms of quantity and quality, can be increased by genetic engineering. Production of lactate can be reduced by lowering the expression of lactate dehydrogenase [62]. The viability of cells can be improved by introducing proto-oncogenes and growth factor genes [138]. Over-expression of cell cycle control genes, such as cyclin-dependent kinase inhibitors, arrests cells in G1-phase and increases productivity [36]. Similar results can be obtained by inhibiting or preventing apoptosis. This is achieved by introducing anti-apoptotic genes *bcl-2* [87] or *bcl-xl* [32] to the host cells, or by inhibiting caspase-3 by antisense RNA [72]. A simpler approach involves adding components of the growth medium that protect against, or inhibit, apoptosis (e.g., suramin, insulin growth factor, caspase inhibitors) [15]. RNAi technology has become an important tool in mammalian cell genetic engineering, and the down-regulated targets include lactate dehydrogenase, dihydrofolate reductase, pro-apoptotic genes and glycosylation-related genes [137].

The hallmark of recombinant protein quality is its correct glycosylation profile. The glycosylation pattern can vary between cell lines, most notably between rodent and human. CHO and NS0 cell lines can incorporate sialic acid *N*-glycolylneuraminic acid, which is not present in humans and can be immunogenic in larger amounts [97]. Various attempts have been made to modify glycosylation by genetic engineering. Over-expression of β-1,4-galactosyltransferase and α-2,3-sialyltransferase caused reduction in terminal *N*-acetylglucosamine residues and an increase in terminal sialyl residues in recombinant proteins produced in CHO cells, which resulted in increased plasma retention time [133]. Over-expression of 2-β-*N*-acetylglucosaminyltransferase resulted in incorporation of bisecting *N*-acetylglucosamine in recombinant antibody, thereby increasing its antibody-dependant cytotoxicity [24]. Similar results can be achieved by the removal of genes involved in the transfer of α-1,6-fucose to nascent glycan chain, which results in fucose-free recombinant antibodies [55]. Besides over-expression of glycosylation-related enzymes, their localization is also important. By changing the localization of 2-β-*N*-acetylglucosaminyltransferase inside the Golgi, the incorporation of bisecting *N*-acetylglucosamine was increased [31]. Nutrients and medium composition are also well known to influence glycosylation pattern and have to be precisely controlled to ensure a consistent glycosylation profile. Glutamine and glucose have to be supplied in sufficient quantity. Ammonia accumulation has to be

controlled, as well as pH, oxygen supply, growth rate, temperature and shear stress [15, 55].

Systemic optimization approaches

Omics technologies (genomics, transcriptomics, proteomics, metabolomics) currently show the greatest potential for identifying novel variables that could be altered for optimizing recombinant protein expression. Omics technologies offer the possibility to understand the cellular functions at the global level [27], and lead biopharmaceutical manufacturing towards systems biotechnology [17]. Multi-gene changes could be applied and tailored to a particular protein. DNA microarray technology has enabled discoveries in the fields of lower lactate production [76], cholesterol production in NS0 cells [111] and adaptation to growth at lower temperatures [3]. Proteomics studies have the advantage of giving insight into post-translational modifications and sub-cellular localization. The majority of such research has been carried out with two-dimensional gene electrophoresis combined with mass spectrometry. Studies have involved the influence of the treatment of CHO cells with butyrate [123] and the influence of osmotic stress [80]. Recently, a multi-omics approach (transcriptomics, metabolomics and fluxomics) was used to compare a recombinant protein producing HEK293 cell line with its parental cell line [26].

Several limitations of omics technologies, described in a recent review [27], hamper their wider introduction in cell engineering. They include technical (e.g., measurement bias and limited coverage), biological (e.g., high variability and complexity), experimental (e.g., small contrast, use of a single technology), and interpretational (e.g., lack of computational methods) [27].

Further advances are constantly being made in omics technologies. These include transcriptomic and genomic technologies, such as suppression subtractive hybridization, serial analysis of gene expression, comparative genomic hybridization and direct sequencing of expressed sequence tags. Next-generation sequencing technologies have enabled the acquisition of a vast amount of sequence data in a short period of time at relatively low cost, and have already resulted in the determination of a CHO transcriptome [8] and a whole genome [140]. New proteomic approaches include multidimensional peptide separation (instead of two dimensional gel electrophoresis) and stable isotope labelling of proteins in cell culture (enabling relative quantification) [43].

Technological advances in cell cultivation

In general, large-scale recombinant biopharmaceutical production takes place in stainless steel, sterilizable bioreactors. The scale-up process is usually difficult and time-consuming due to the ill-defined process parameters in laboratory-scale production. Novel technological approaches, such as disposable bioreactors [28] and microscale bioprocesses [91], could change this and become an integral part of biopharmaceutical production.

Disposable bioreactors, intended for single use, are made from a sterile material, usually plastic, that is approved by a regulatory authority (polyethylene, polystyrene, polypropylene, etc.). They can be used in smaller volumes (up to 1 l) for the preparation of seed cultures. Recently, disposable bioreactors with a capacity of several thousand litres have become available, and can be used for production cultures. Their complexity varies from simple containers that depend on an external device for maintaining suitable growth conditions, to complex devices equipped with disposable sensors able to control process parameters. The container is either rigid (tube, flask) or flexible (bag). The quality and quantity of produced proteins is comparable to that achieved with traditional stainless steel bioreactors. The advantages include great flexibility, ease of handling and, potentially, lower costs due to pre-sterilization [28]. The disadvantages include lower material strength, limited experience with the technology, limited availability of disposable sensors, controversial safety due to extractables and leachables, and higher running costs. In general, disposable bioreactors are cost effective only for high-value products [28]. However, the cost-efficiency could change in their favour, according to some calculations [42]. Mass transfer in disposable bioreactors is achieved through wave mixing, orbital shaking or stirring. Stirred bag systems are the most established. They consist of a cylindrical bag with preinstalled axial flow impellers, aeration devices, gas filters and sensor probe ports. Disposable bioreactors require sensors for monitoring physical, chemical and biological parameters. Sensors that come into direct contact with the cell culture medium must be disposable, and therefore inexpensive. They have to be reliable, but not necessarily durable. Semiconductor devices can be used or, alternatively, disposable inexpensive sensing elements coupled to reusable analytical equipment. Optical sensors have the greatest applicability, since they enable noninvasive monitoring through a transparent window. Different parts of the electromagnetic spectrum can be used to monitor optical density, biomass and nutrients. Optical chemosensors rely on indicators that are in physical contact with the medium and must therefore be disposable. They enable sensing of dissolved oxygen and pH. Besides optical and conductivity sensors, ultrasonic sensors could also be applied in a noninvasive manner [39].

The miniaturization of bioprocess volume can lead to higher throughput and faster process development.

Microscale bioprocesses can be performed in either microwell or microfluidic format, or in a combination of both. Microwell systems are widely applied in everyday research; however the transfer of bioprocess data to larger systems is not trivial and is rarely established. Homogeneity, following the addition or removal of liquid, can be achieved by jet mixing or shaking. In the latter, the critical shaking speed has to be achieved. Oxygen transfer rate in microwells can be calculated. Fluid dynamics are more difficult to monitor; however, advances have been made with microparticle image velocity measurements and direct numerical simulations [91]. Microfluidic systems are characterized by continuous flow of the liquid and could be particularly useful in designing continuous and perfusion cultures. The scale-up is, however, even more difficult, and the large surface area to volume ratio is a problem. Recently, several microbioreactors have been developed which feature properties of both microwell and microfluidic systems, and a great deal of attention is focused on instrumentation and automation [110]. Good agreement of process parameters of *E. coli* growth were observed between a microbioreactor and a laboratory-scale fermentor [35]. Monoclonal antibody purification was established with automated liquid handling in a miniaturized multi-column format [121]. Microbioreactors may significantly shorten the time for scale-up process development.

Conclusions and perspectives

Production in mammalian cells accounts for the largest part of the biopharmaceutical repertoire. *E. coli* and yeasts remain the second and third most commonly used production hosts. Relatively low production costs and well established regulatory compliance appear to be crucial for their popularity, despite their associated problems. The adoption of a new expression host is slow, mainly due to regulatory constraints. It is nevertheless occurring. Recent approval of biopharmaceuticals produced in *P. pastoris*, insect cells and transgenic animals [126], as well as plant cells [98], could open the door more widely for future candidates, and also for other expression hosts. Considerable progress has been made in the expression of biopharmaceuticals in plants, which was not considered in this review. The major limiting issue, immunogenicity of plant-specific glycosylation, has been addressed successfully [70]. A lot of effort has been invested in metabolic engineering of various producer cells, with special focus on glycosylation engineering. Engineered cells have to face regulatory hurdles; however, their impact is likely to increase in the future, especially due to high-throughput approaches, with which new improvements can be rapidly identified. Systematic optimization approaches, together with new fermentation technological solutions will be the driving forces of future advances and may result in improved biopharmaceuticals and their faster delivery to the market.

Acknowledgments This study was supported by the Slovenian Research Agency Grant No. P4-0127. We are grateful to Prof. Roger Pain for critical reading of the manuscript.

References

1. Aldor IS, Krawitz DC, Forrest W, Chen C, Nishihara JC, Joly JC, Champion KM (2005) Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical. Appl Environ Microbiol 71(4):1717–1728
2. Arechaga I, Miroux B, Runswick MJ, Walker JE (2003) Over-expression of *Escherichia coli* F1F(o)-ATPase subunit a is inhibited by instability of the uncB gene transcript. FEBS Lett 547(1–3):97–100
3. Baik JY, Lee MS, An SR, Yoon SK, Joo EJ, Kim YH, Park HW, Lee GM (2006) Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. Biotechnol Bioeng 93(2):361–371
4. Balbas P, Gosset G (2001) Chromosomal editing in *Escherichia coli*—vectors for DNA integration and excision. Mol Biotechnol 19(1):1–12
5. Baldi L, Hacker DL, Adam M, Wurm FM (2007) Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. Biotechnol Lett 29(5):677–684
6. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Curr Opin Biotech 10(5):411–421
7. Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol 22(11):1399–1408
8. Becker J, Hackl M, Rupp O, Jakobi T, Schneider J, Szczepanowski R, Bekel T, Borth N, Goesmann A, Grillari J, Kaltschmidt C, Noll T, Puhler A, Tauch A, Brinkrolf K (2011) Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. J Biotechnol 156(3):227–235
9. Bibila TA, Robinson DK (1995) In pursuit of the optimal fed-batch process for monoclonal antibody production. Biotechnol Prog 11(1):1–13
10. 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(9):757–766
11. Boehm T, Pirie-Shepherd S, Trinh LB, Shiloach J, Folkman J (1999) Disruption of the KEX1 gene in *Pichia pastoris* allows expression of full-length murine and human endostatin. Yeast 15(7):563–572
12. Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P, Barr PJ (1984) Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 81(15):4642–4646
13. Brezinsky SC, Chiang GG, Szilvasi A, Mohan S, Shapiro RI, MacLean A, Sisk W, Thill G (2003) A simple method for enriching populations of transfected CHO cells for cells of

higher specific productivity. J Immunol Methods 277(1–2): 141–155
14. Browne SM, Al-Rubeai M (2007) Selection methods for high-producing mammalian cell lines. Trends Biotechnol 25(9):425–432
15. Butler M (2005) Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. Appl Microbiol Biotechnol 68(3):283–291
16. Cacciatore JJ, Chasin LA, Leonard EF (2010) Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the DHFR-based CHO cell selection system. Biotechnol Adv 28(6):673–681
17. Carinhas N, Oliveira R, Alves PM, Carrondo MJ, Teixeira AP (2012) Systems biotechnology of animal cells: the road to prediction. Trends Biotechnol 30(7):377–385
18. Cereghino GP, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13(4):329–332
19. Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24(1):45–66
20. 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 USA 100(9):5022–5027
21. Choi JH, Keum KC, Lee SY (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem Eng Sci 61(3):876–885
22. 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(8):4737–4742
23. Chou CP (2007) Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Appl Microbiol Biot 76(3):521–532
24. Davies J, Jiang L, Pan LZ, LaBarre MJ, Anderson D, Reff M (2001) Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for FC gamma RIII. Biotechnol Bioeng 74(4):288–294
25. Dieci G, Bottarelli L, Ballabeni A, Ottonello S (2000) tRNA-assisted overproduction of eukaryotic ribosomal proteins. Protein Expr Purif 18(3):346–354
26. Dietmair S, Hodson MP, Quek LE, Timmins NE, Gray P, Nielsen LK (2012) A multi-omics analysis of recombinant protein production in hek293 cells. PLoS ONE 7(8):e43394
27. Dietmair S, Nielsen LK, Timmins NE (2012) Mammalian cells as biopharmaceutical production hosts in the age of omics. Biotechnol J 7(1):75–89
28. Eibl R, Kaiser S, Lombriser R, Eibl D (2010) Disposable bioreactors: the current state-of-the-art and recommended applications in biotechnology. Appl Microbiol Biotechnol 86(1):41–49
29. Eiden-Plach A, Zagorc T, Heintel T, Carius Y, Breinig F, Schmitt MJ (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. Appl Environ Microbiol 70(2):961–966
30. Farmer WR, Liao JC (1997) Reduction of aerobic acetate production by *Escherichia coli*. Appl Environ Microbiol 63(8):3205–3210
31. Ferrara C, Brunker P, Suter T, Moser S, Puntener U, Umana P (2006) Modulation of therapeutic antibody effector functions by glycosylation engineering: influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. Biotechnol Bioeng 93(5):851–861
32. Figueroa B Jr, Sauerwald TM, Oyler GA, Hardwick JM, Betenbaugh MJ (2003) A comparison of the properties of a Bcl-xL variant to the wild-type anti-apoptosis inhibitor in mammalian cell cultures. Metab Eng 5(4):230–245
33. Franchini AG, Egli T (2006) Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology 152(Pt 7):2111–2127
34. Fukushige S, Sauer B (1992) Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. Proc Natl Acad Sci USA 89(17):7905–7909
35. Funke M, Buchenauer A, Mokwa W, Kluge S, Hein L, Muller C, Kensy F, Buchs J (2010) Bioprocess control in microscale: scalable fermentations in disposable and user-friendly microfluidic systems. Microb Cell Fact 9:86
36. Fussenegger M, Schlatter S, Datwyler D, Mazur X, Bailey JE (1998) Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells. Nat Biotechnol 16(5):468–472
37. Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D (2007) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl Environ Microbiol 73(20):6499–6507
38. Gerngross TU (2004) Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol 22(11):1409–1414
39. Glindkamp A, Riechers D, Rehbock C, Hitzmann B, Scheper T, Reardon KF (2009) Sensors in disposable bioreactors status and trends. Adv Biochem Eng Biotechnol 115:145–169
40. Goncalves GA, Bower DM, Prazeres DM, Monteiro GA, Prather KL (2012) Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. Biotechnol J 7(2): 251–261
41. Gorfien S, Paul B, Walowitz J, Keem R, Biddle W, Jayme D (2000) Growth of NS0 cells in protein-free, chemically defined medium. Biotechnol Prog 16(5):682–687
42. Gottschalk U, Brorson K, Shukla AA (2012) The need for innovation in biomanufacturing. Nat Biotechnol 30(6):489–492
43. Griffin TJ, Seth G, Xie H, Bandhakavi S, Hu WS (2007) Advancing mammalian cell culture engineering using genome-scale technologies. Trends Biotechnol 25(9):401–408
44. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177(14): 4121–4130
45. Haddadin FT, Harcum SW (2005) Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. Biotechnol Bioeng 90(2):127–153
46. 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(5637):1244–1246
47. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313(5792): 1441–1443
48. Hamilton SR, Gerngross TU (2007) Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr Opin Biotechnol 18(5):387–392
49. Han MJ, Jeong KJ, Yoo JS, Lee SY (2003) Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling. Appl Environ Microbiol 69(10): 5772–5781

50. Hanania EG, Fieck A, Stevens J, Bodzin LJ, Palsson BO, Koller MR (2005) Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers. Biotechnol Bioeng 91(7):872–876
51. Harmsen MM, Bruyne MI, Raue HA, Maat J (1996) Overexpression of binding protein and disruption of the PMR1 gene synergistically stimulate secretion of bovine prochymosin but not plant thaumatin in yeast. Appl Microbiol Biotechnol 46(4):365–370
52. Hartl FU, Hayer-Hartl M (2002) Protein folding—molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295(5561):1852–1858
53. Holmes P, Al-Rubeai M (1999) Improved cell line development by a high throughput affinity capture surface display technique to select for high secretors. J Immunol Methods 230(1–2): 141–147
54. Hong E, Davidson AR, Kaiser CA (1996) A pathway for targeting soluble misfolded proteins to the yeast vacuole. J Cell Biol 135(3):623–633
55. Hossler P, Khattak SF, Li ZJ (2009) Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 19(9):936–949
56. Hou J, Tyo KE, Liu Z, Petranovic D, Nielsen J (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. FEMS Yeast Res 12(5):491–510
57. Idicula-Thomas S, Balaji PV (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in *Escherichia coli*. Protein Sci 14(3):582–592
58. Idicula-Thomas S, Kulkarni AJ, Jayaraman VK, Balaji PV (2006) A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in Escherichia coli. Bioinformatics 22(3):278–284
59. Idiris A, Tohda H, Kumagai H, Takegawa K (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. Appl Microbiol Biotechnol 86(2):403–417
60. Idiris A, Tohda H, Sasaki M, Okada K, Kumagai H, Giga-Hama Y, Takegawa K (2010) Enhanced protein secretion from multiprotease-deficient fission yeast by modification of its vacuolar protein sorting pathway. Appl Microbiol Biotechnol 85(3): 667–677
61. Jana S, Deb JK (2005) Strategies for efficient production of heterologous proteins in *Escherichia coli*. Appl Microbiol Biot 67(3):289–298
62. Jeong D, Kim TS, Lee JW, Kim KT, Kim HJ, Kim IH, Kim IY (2001) Blocking of acidosis-mediated apoptosis by a reduction of lactate dehydrogenase activity through antisense mRNA expression. Biochem Biophys Res Commun 289(5):1141–1149
63. 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(1):150–156
64. 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(7):3027–3032
65. Jeong KJ, Lee SY (2003) Enhanced production of recombinant proteins in *Escherichia coli* by filamentation suppression. Appl Environ Microbiol 69(2):1295–1298
66. Jones EW (1991) Tackling the protease problem in *Saccharomyces cerevisiae*. Methods Enzymol 194:428–453
67. Jonson L, Rehfeld JF, Johnsen AH (2004) Enhanced peptide secretion by gene disruption of CYM1, a novel protease in *Saccharomyces cerevisiae*. Eur J Biochem 271(23–24):4788–4797
68. Kamionka M (2011) Engineering of therapeutic proteins production in *Escherichia coli*. Curr Pharm Biotechnol 12(2): 268–274
69. Kanjou N, Nagao A, Ohmiya Y, Ohgiya S (2007) Yeast mutant with efficient secretion identified by a novel secretory reporter, Cluc. Biochem Biophys Res Commun 358(2):429–434
70. Karg SR, Kallio PT (2009) The production of biopharmaceuticals in plant systems. Biotechnol Adv 27(6):879–894
71. Kennard ML (2011) Engineered mammalian chromosomes in cellular protein production: future prospects. Methods Mol Biol 738:217–238
72. Kim NS, Lee GM (2002) Inhibition of sodium butyrate-induced apoptosis in recombinant Chinese hamster ovary cells by constitutively expressing antisense RNA of caspase-3. Biotechnol Bioeng 78(2):217–228
73. Kingsman SM, Kingsman AJ, Dobson MJ, Mellor J, Roberts NA (1985) Heterologous gene expression in *Saccharomyces cerevisiae*. Biotechnol Genet Eng Rev 3:377–416
74. Kito M, Itami S, Fukano Y, Yamana K, Shibui T (2002) Construction of engineered CHO strains for high-level production of recombinant proteins. Appl Microbiol Biotechnol 60(4): 442–448
75. Knappskog S, Ravneberg H, Gjerdrum C, Trosse C, Stern B, Pryme IF (2007) The level of synthesis and secretion of *Gaussia princeps* luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide. J Biotechnol 128(4): 705–715
76. Korke R, Gatti Mde L, Lau AL, Lim JW, Seow TK, Chung MC, Hu WS (2004) Large scale gene expression profiling of metabolic shift of mammalian cells in culture. J Biotechnol 107(1):1–17
77. Kwaks TH, Otte AP (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. Trends Biotechnol 24(3):137–142
78. Kwaks TH, Sewalt RG, van Blokland R, Siersma TJ, Kasiem M, Kelder A, Otte AP (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. J Biotechnol 115(1):35–46
79. Laursen BS, Steffensen SAD, Hedegaard J, Moreno JMP, Mortensen KK, Sperling-Petersen HU (2002) Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial infB gene. Genes Cells 7(9):901–910
80. Lee MS, Kim KW, Kim YH, Lee GM (2003) Proteome analysis of antibody-expressing CHO cells in response to hyperosmotic pressure. Biotechnol Prog 19(6):1734–1741
81. Lee SY (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol 14(3):98–105
82. Lopes TS, Klootwijk J, Veenstra AE, van der Aar PC, van Heerikhuizen H, Raue HA, Planta RJ (1989) High-copy-number integration into the ribosomal DNA of *Saccharomyces cerevisiae*: a new vector for high-level expression. Gene 79(2):199–206
83. Luan CH, Qiu SH, Finley JB, Carson M, Gray RJ, Huang WY, Johnson D, Tsao J, Reboul J, Vaglio P, Hill DE, Vidal M, Delucas LJ, Luo M (2004) High-throughput expression of *C. elegans* proteins. Genome Res 14(10B):2102–2110
84. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22(4):249–270
85. Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol Rev 60(3): 512
86. Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? Curr Opin Genet Dev 15(2):163–176
87. Mastrangelo AJ, Hardwick JM, Zou S, Betenbaugh MJ (2000) Part II. Overexpression of bcl-2 family members enhances

survival of mammalian cells in response to various culture insults. Biotechnol Bioeng 67(5):555–564
88. Mattanovich D, Borth N (2006) Applications of cell sorting in biotechnology. Microb Cell Fact 5:12
89. Meng YG, Liang J, Wong WL, Chisholm V (2000) Green fluorescent protein as a second selectable marker for selection of high producing clones from transfected CHO cells. Gene 242(1–2):201–207
90. Mercille S, Johnson M, Lanthier S, Kamen AA, Massie B (2000) Understanding factors that limit the productivity of suspension-based perfusion cultures operated at high medium renewal rates. Biotechnol Bioeng 67(4):435–450
91. Micheletti M, Lye GJ (2006) Microscale bioprocess optimisation. Curr Opin Biotechnol 17(6):611–618
92. Nakanishi-Shindo Y, Nakayama K, Tanaka A, Toda Y, Jigami Y (1993) Structure of the N-linked oligosaccharides that show the complete loss of alpha-1,6-polymannose outer chain from och1, och1 mnn1, and och1 mnn1 alg3 mutants of *Saccharomyces cerevisiae*. J Biol Chem 268(35):26338–26345
93. Nandakumar MP, Cheung A, Marten MR (2006) Proteomic analysis of extracellular proteins from *Escherichia coli* W3110. J Proteome Res 5(5):1155–1161
94. Natale P, Bruser T, Driessen AJM (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. BBA-Biomembranes 1778(9):1735–1756
95. Ni Y, Chen R (2009) Extracellular recombinant protein production from *Escherichia coli*. Biotechnol Lett 31(11):1661–1670
96. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, *Escherichia coli*. Appl Environ Microbiol 64(5):1694–1699
97. Noguchi A, Mukuria CJ, Suzuki E, Naiki M (1995) Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. J Biochem 117(1):59–62
98. Opar A (2011) 'Pharmers' hope for first plant drug harvest. Nat Rev Drug Discov 10(2):81–82
99. Palermo DP, DeGraaf ME, Marotti KR, Rehberg E, Post LE (1991) Production of analytical quantities of recombinant proteins in Chinese hamster ovary cells using sodium butyrate to elevate gene expression. J Biotechnol 19(1):35–47
100. Parente D, Raucci G, D'Alatri L, d'Estais G, Novelli S, Pacilli A, Saccinto MP, Mele A, De Santis R (1998) Overproduction of soluble, extracellular cytotoxin alpha-sarcin in *Escherichia coli*. Mol Biotechnol 9(2):99–106
101. Payne T, Finnis C, Evans LR, Mead DJ, Avery SV, Archer DB, Sleep D (2008) Modulation of chaperone gene expression in mutagenized *Saccharomyces cerevisiae* strains developed for recombinant human albumin production results in increased production of multiple heterologous proteins. Appl Environ Microbiol 74(24):7759–7766
102. Posfai G, Plunkett G 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. Science 312(5776):1044–1046
103. Rinas U, Bailey JE (1992) Protein Compositional Analysis of Inclusion-Bodies Produced in Recombinant *Escherichia coli*. Appl Microbiol Biot 37(5):609–614
104. Ritz D, Beckwith J (2001) Roles of thiol-redox pathways in bacteria. Annu Rev Microbiol 55:21–48
105. Robbens J, Raeymaekers A, Steidler L, Fiers W, Remaut E (1995) Production of soluble and active recombinant murine interleukin-2 in *Escherichia coli*: high level expression, Kil-induced release, and purification. Protein Expr Purif 6(4):481–486
106. Robinson AS, Hines V, Wittrup KD (1994) Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*. Biotechnology 12(4):381–384
107. Romanos MA, Scorer CA, Clare JJ (1992) Foreign gene expression in yeast: a review. Yeast 8(6):423–488
108. Ruohonen L, Toikkanen J, Tieaho V, Outola M, Soderlund H, Keranen S (1997) Enhancement of protein secretion in *Saccharomyces cerevisiae* by overproduction of Sso protein, a late-acting component of the secretory machinery. Yeast 13(4):337–351
109. Ryll T, Dutina G, Reyes A, Gunson J, Krummen L, Etcheverry T (2000) Performance of small-scale CHO perfusion cultures using an acoustic cell filtration device for cell retention: characterization of separation efficiency and impact of perfusion on product quality. Biotechnol Bioeng 69(4):440–449
110. Schapper D, Alam MN, Szita N, Eliasson Lantz A, Gernaey KV (2009) Application of microbioreactors in fermentation process development: a review. Anal Bioanal Chem 395(3):679–695
111. Seth G, Philp RJ, Denoya CD, McGrath K, Stutzman-Engwall KJ, Yap M, Hu WS (2005) Large-scale gene expression analysis of cholesterol dependence in NS0 cells. Biotechnol Bioeng 90(5):552–567
112. Sharma SS, Blattner FR, Harcum SW (2007) Recombinant protein production in an *Escherichia coli* reduced genome strain. Metab Eng 9(2):133–141
113. Shusta EV, Raines RT, Pluckthun A, Wittrup KD (1998) Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. Nat Biotechnol 16(8):773–777
114. Singh SM, Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99(4):303–310
115. Smith JD, Tang BC, Robinson AS (2004) Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast. Biotechnol Bioeng 85(3):340–350
116. Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115(2):113–128
117. St John TP, Davis RW (1981) The organization and transcription of the galactose gene cluster of *Saccharomyces*. J Mol Biol 152(2):285–315
118. Stewart EJ, Aslund F, Beckwith J (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J 17(19):5543–5550
119. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185:60–89
120. Taylor G, Hoare M, Gray DR, Marston FAO (1986) Size and density of protein inclusion-bodies. Bio-Technology 4(6):553–557
121. Treier K, Hansen S, Richter C, Diederich P, Hubbuch J, Lester P (2012) High-throughput methods for miniaturization and automation of monoclonal antibody purification processes. Biotechnol Prog 28(3):723–732
122. van der Heide M, Hollenberg CP, van der Klei IJ, Veenhuis M (2002) Overproduction of BiP negatively affects the secretion of *Aspergillus niger* glucose oxidase by the yeast *Hansenula polymorpha*. Appl Microbiol Biotechnol 58(4):487–494
123. Van Dyk DD, Misztal DR, Wilkins MR, Mackintosh JA, Poljak A, Varnai JC, Teber E, Walsh BJ, Gray PP (2003) Identification of cellular changes associated with increased production of human growth hormone in a recombinant Chinese hamster ovary cell line. Proteomics 3(2):147–156

124. Vijayendran C, Flaschel E (2010) Impact of profiling technologies in the understanding of recombinant protein production. Adv Biochem Eng Biotechnol 121:45–70
125. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 298(5599):1790–1793
126. Walsh G (2010) Biopharmaceutical benchmarks 2010. Nat Biotechnol 28(9):917–924
127. Walsh G (2012) New biopharmaceuticals. Biopharm Int 25(6):34–36
128. Walsh G, Jefferis R (2006) Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol 24(10):1241–1252
129. Wang L (2009) Towards revealing the structure of bacterial inclusion bodies. Prion 3(3):139–145
130. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R (2008) Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol 6(8):e195
131. Weaver JC, McGrath P, Adams S (1997) Gel microdrop technology for rapid isolation of rare and high producer cells. Nat Med 3(5):583–585
132. Wegner GH (1990) Emerging applications of the methylotrophic yeasts. FEMS Microbiol Rev 7(3–4):279–283
133. Weikert S, Papac D, Briggs J, Cowfer D, Tom S, Gawlitzek M, Lofgren J, Mehta S, Chisholm V, Modi N, Eppler S, Carroll K, Chamow S, Peers D, Berman P, Krummen L (1999) Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. Nat Biotechnol 17(11):1116–1121
134. Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C (2009) Design parameters to control synthetic gene expression in *Escherichia coli*. Plos One 4(9):e7002
135. Wewetzer K, Seilheimer B (1995) Establishment of a single-step hybridoma cloning protocol using an automated cell transfer system: comparison with limiting dilution. J Immunol Methods 179(1):71–76
136. Wilkinson DL, Harrison RG (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*. Bio-Technol 9(5):443–448
137. Wu SC (2009) RNA interference technology to improve recombinant protein production in Chinese hamster ovary cells. Biotechnol Adv 27(4):417–422
138. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22(11):1393–1398
139. Wurm FM, Gwinn KA, Kingston RE (1986) Inducible overproduction of the mouse c-myc protein in mammalian cells. Proc Natl Acad Sci USA 83(15):5414–5418
140. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang J (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29(8):735–741
141. Yoon SK, Kim SH, Lee GM (2003) Effect of low culture temperature on specific productivity and transcription level of anti-4-1BB antibody in recombinant Chinese hamster ovary cells. Biotechnol Prog 19(4):1383–1386
142. Zhang B, Chang A, Kjeldsen TB, Arvan P (2001) Intracellular retention of newly synthesized insulin in yeast is caused by endoproteolytic processing in the Golgi complex. J Cell Biol 153(6):1187–1198
143. Zhang G, Brokx S, Weiner JH (2006) Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. Nat Biotechnol 24(1):100–104
144. Zhou H, Liu ZG, Sun ZW, Huang Y, Yu WY (2010) Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. J Biotechnol 147(2):122–129
145. Zhu J (2012) Mammalian cell protein expression for biopharmaceutical production. Biotechnol Adv 30(5):1158–1170