Recombinant Proteins
SUMMARY

The cloning of a gene into an expression vector often yields sufficient quantities of proteins for further study. Insulin, interferon, and erythropoietin are all examples of human genes that have been cloned into expression vectors and appropriately expressed into protein in a host organism, termed recombinant proteins. Despite the relative ease of expressing and using recombinant proteins, problems can occur. These problems include issues with stability of the recombinant DNA or vector, inability of bacteria to process eukaryotic gene structures like introns, issues with codon preference differences among various organisms, toxic overproduction and secretion, and appropriate modifications of the proteins while expressed in a surrogate system.

Expression vectors are useful for cloning foreign genes and expressing those into protein within a bacterial host, typically *E. coli*. An antibiotic resistance gene located on the vector and the presence of the antibiotic itself in the culture ensure the stability of the vector. When attempting to express eukaryotic genes in a bacterial system, one must keep in mind that eukaryotic promoters are not recognized by bacterial RNA polymerase. Additionally, introns present in eukaryotic genes are not processed by bacteria. So the cDNA copy of eukaryotic mRNA must be cloned instead of the actual gene. Expression vectors are typically optimized for transcription. However, translational expression vectors possess the consensus sequence for the ribosomal binding site located the exact number of bases away from the ATG start codon for efficient translation initiation.

The first genetically engineered hormone available to humans was insulin. Insulin is produced by the pancreas of animals and helps to regulate blood glucose levels. Individuals with diabetes mellitus have difficulty regulating blood sugar levels and often suffer from high blood sugar. This inability to regulate blood sugar levels stems from lack of insulin and/or defective insulin receptors. Patients who are insulin-dependent require injections of insulin. Prior to genetic engineering, diabetics would inject swine or bovine insulin. An occasional consequence of using nonhuman insulin was allergic reactions.

Insulin is initially made in an inactive form called preproinsulin, which is then processed by several enzymes to yield the final, functional product. Since bacteria do not possess mammalian processing enzymes, expression of the cloned insulin gene in bacteria yields only preproinsulin. In addition, some proteins require disulfide bridges for stability, but these bridges are not easily formed in the bacterial cytoplasm. To circumvent these issues for human insulin expression, researchers generated two artificial mini-genes: one gene for the insulin A-chain and the other for the insulin B-chain. These genes were then fused to a tag that allowed easy purification of the resulting proteins. Two separate bacterial cultures were used to generate the two insulin chains separately. The bacterially expressed insulin chains were purified from the host cells, mixed together, and treated to link the disulfide bridges. The coding sequence of the insulin artificial mini-genes was also altered to introduce amino acid substitutions in the final protein product. This prevented the protein from clumping and also created a fast-acting insulin.

Despite the redundancy in the genetic code, organisms still have preferences for some codons over others for the same amino acid, which means that the tRNA responsible for the rare codons in a host are in lower concentrations within the cell. When one is trying to express a foreign protein in a host system, codon usage differences could affect the efficiency of expression. Changing the codons of the foreign protein is laborious but could help. A quicker approach is to simply supply the cells with extra copies of the tRNAs for the rare codons.

Sometimes, overexpression of foreign proteins in bacterial cells yields harmful effects. Bacteria that are expressing foreign proteins will often form inclusion bodies, which are dense, insoluble, crystalline structures containing nonfunctional and misfolded proteins. Toxic overproduction can be controlled by using expression systems with features that allow expression of the foreign gene only upon addition of a chemical signal to the culture. This
enables a researcher to control when and how much protein is expressed. Two systems are commonly used. The pET system uses a hybrid T7/lac promoter and is transformed into a host cell that has been genetically engineered with the gene for T7 RNA polymerase in the chromosome. Upon addition of the chemical signal, IPTG in this example, the gene for T7 RNA polymerase is expressed and the product binds to the vector’s T7 promoter located upstream of the gene of interest. The pBAD system works using the arabinose promoter, which is induced through the addition of arabinose. The amount of protein produced is controlled by varying the concentration of arabinose to induce the promoter.

Inclusion bodies are dense aggregates of misfolded proteins and often occur as a consequence of overexpression of recombinant proteins. The use of molecular chaperones can aid in proper protein folding. Protein misfolding can also be alleviated by expressing the recombinant protein in the periplasmic space or within the culture medium. Alternatively, attempts can be made to solubilize the inclusion bodies and refold the proteins through the addition of chaotropic agents and other chemicals.

Once protein is successfully produced in a host system, stability of the protein depends on its susceptibility to proteases. The amino acid present on the N-terminus of the protein affects the half-life of that particular protein within the cell. The most stable amino acids are methionine, glycine, alanine, serine, threonine, and valine. The least stable amino acids are leucine, phenylalanine, aspartate, lysine, and arginine. Protein stability is also influenced by PEST (proline, glutamate, serine, threonine) sequences scattered throughout the polypeptide. These sequences are recognized by proteases and cut the protein into nonfunctioning fragments. Changes to the N-terminal amino acid are relatively simple compared with changes to the PEST sequences, with the goal of eventually obtaining stable but functional protein.

Recombinant proteins produced by bacteria can be excreted into the surrounding medium by one of several mechanisms. A signal sequence is engineered into the cloned gene, which is recognized by a general secretory system to export the protein to the periplasm. Another option is to fuse the protein of interest to a protein that is normally secreted from the bacterial cell, such as the MaLE protein of *E. coli*. This is done at the genetic level, and the two proteins must be in frame to be correctly expressed into protein. The recombinant protein would eventually need to be separated from the carrier protein for further analysis. Gram-positive bacteria do not have an outer membrane. Exporting foreign proteins directly into the surrounding medium in these systems might be easier if the genetics for gram-positives were better understood. Finally, specialized export systems in *E. coli* could be used to secrete the protein of interest across both the inner and outer membranes and eventually into the surrounding medium. Autotransporters of the type V export system of *E. coli* are also useful, particularly for very large proteins.

Eukaryotic expression vectors are made for yeast, insects, and even mammalian cell cultures to express eukaryotic proteins within eukaryotic cells, which have advantages over the expression of the same proteins in prokaryotic systems. Some eukaryotic proteins are post-translationally modified to be functional, such as by the addition of new functional groups to amino acids, or sugars to proteins (glycosylation), and even formation of disulfide bridges for stability. In terms of glycosylation, many eukaryotic proteins must be glycosylated in order to function. Cloning and expressing these genes in bacteria such as *E. coli* usually result either in no glycosylation or in inappropriate glycosylation (O-linked instead of N-linked). Some prokaryotes produce enzymes that can form the N-linked sugar groups. The genes for the enzymes have been cloned and expressed in the same host system as the recombinant proteins to produce N-linked glycosylated proteins. Insect expression systems are advantageous in some cases for expressing mammalian proteins because insect cell cultures are easier to maintain and post-translationally modify proteins in a fashion similar to mammalian cells, particularly in regards to glycosylation of proteins. Insect cell lines expressing glycosylation pathways of mammals are also available for the expression and subsequent N-glycosylation of recombinant proteins.
Sometimes proteins are made as longer polypeptides that are cleaved to form active protein. Yeast expression systems are advantageous because the organisms are easy to grow and manipulate, and they are well studied. Ultimately, mammalian expression vectors are used to express difficult mammalian proteins. They have several features, such as mammalian origins of replication and promoters for efficient production of protein in mammalian cell lines. Since mammalian cells do not respond to antibiotics, these vectors are maintained through the use of enzymes or components of biosynthetic pathways, such as for amino acids. Proteins that function only when individual subunits associate are more problematic. Each subunit must be expressed in the cell line either by expressing them all separately on one vector or as an artificial operon, or on multiple vectors.
Case Study  Microbials for the Production of Monoclonal Antibodies and Antibody Fragments


Antibodies are glycosylated by mammalian enzymes to produce functional proteins. Host cell limitations must be considered when expressing recombinant antibody protein in expression systems. Traditionally, mammalian host cells have been used for production of full-length monoclonal antibodies because these systems possess the N-glycosylation enzymes and so add sugars to appropriate atoms on the antibodies. However, using mammalian systems to produce large amounts of antibody is time-consuming and costly. Since production of full-length monoclonal antibodies represents a large portion of the biopharmaceutical marketplace, more efficient and cost-effective systems have been sought.

Production of antibody fragments in microbial systems yields fragment proteins that are still able to bind antigen but are produced at a lower cost and with no toxic industrial waste from chemical synthesis. The full-length antibody is stabilized by glycosylation, and although the full-length version is not needed for antigen binding, it is needed for cell-mediated cytotoxicity and enhanced clearance rates.

In this article, the authors review three expression systems for monoclonal antibody production: *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Escherichia coli*.

The *S. cerevisiae* expression system is widely used for the production of recombinant proteins. What features of this system make it attractive for antibody production?

The *S. cerevisiae* expression system is well developed. Yeast cells have been engineered for optimal expression and production. Additionally, *Saccharomyces* are easy to grow and transform. Multiple shuttle vectors are available for expression of the recombinant proteins. *S. cerevisiae* undergoes a mixed oxidative/fermentative metabolism that can be shifted to a more aerobic metabolism during protein production.

What types of shuttle vectors are available for expression of recombinant proteins in *S. cerevisiae* expression systems? Are there advantages or disadvantages for each?

Yep, Ycp, and Yip are three main shuttle vectors designed for various types of expression of recombinant proteins. Yep (yeast episomal plasmid) contains the 2 micron origin of replication for *S. cerevisiae* and expresses large amounts of protein because it exists in the cells in high copy numbers. The Ycp (yeast centromeric plasmid) vector contains a replication sequence and exists in lower copy numbers. Yip (yeast integrative plasmid) contains sequences for integration of the vector into the yeast chromosome.

Higher levels of expression, such as those levels observed with the high-copy Yep vectors, have the potential to increase protein yield at the expense of protein stability and quality. Increased protein stability and quality can be achieved using the lower copy number Ycp vector or the integration plasmid Yip. The disadvantage of having fewer copies of the gene, however, is that expression level is decreased. To some extent, this decreased expression potential can be overcome by cloning the gene behind a highly transcribed promoter.

How has genetic engineering of *S. cerevisiae* enhanced the production process of recombinant proteins in this expression system?

Improper folding and trafficking inhibit production of recombinant monoclonal antibodies. To increase the efficacy of production, researchers have engineered strains of *S. cerevisiae* to overexpress molecule chaperones and enzymes necessary for proper folding and secretion. Additionally, several strains have deletions in genes encoding proteases although these genes have not yet been used for production of full-length antibodies or fragments. Several other engineered strains have also been developed that could potentially work well for the production of antibodies. Also, a yeast strain was engineered that undergoes only aerobic metabolisms, thus eliminating the potential for toxic metabolite accumulation due to a fermentative metabolism.

Are there any advantages or disadvantages of using the yeast *P. pastoris* over *S. cerevisiae* for the production of recombinant antibodies?

Expression in *P. pastoris* and *S. cerevisiae* is similar. Both are easily grown and transformed with recombinant DNA. *P. pastoris* is in some ways more challenging due to its tendency to undergo nonhomologous recombination events. However, there are engineered *P. pastoris* strains with inactivated nonhomologous recombination pathways. Strains of *P. pastoris* have been produced to introduce glycosylation enzymes from other yeast, insect, protistan, and mammalian origins to make the yeast better able to express and modify full-length monoclonal antibodies. One advantage for *P. pastoris* is its preference for aerobic metabolism over fermentation. This generates high biomass in a relatively inexpensive media and does not accumulate some of the toxic metabolites from fermentation.

*E. coli* has been the gold standard for working with recombinant proteins and other aspects of molecular biology and biotechnology. What are the challenges of using *E. coli* expression systems for the production of human proteins, specifically antibodies?

Despite the relative ease of growth and genetic manipulation, *E. coli* still is a prokaryotic organism. For it to express mammalian proteins like antibodies, one must make some modifications. The most glaring issue is the lack of N-glycosylation enzymes in *E. coli*. Glycosylation of the antibodies is needed for protein stability. Recently, an N-glycosylation pathway was discovered in the prokaryote *Campylobacter jejuni*. Potentially, this pathway could be coexpressed with recombinant full-length antibodies in *E. coli* and used to produce glycosylated antibodies. Additionally, the reduced environment of the cytoplasm is not conducive for disulfide bridge formation, and often proteins misfold and inclusion bodies form. Coexpression (Continued)
of chaperones with the recombinant protein or targeting the protein to the periplasmic space helps alleviate some of these issues.

Much work has been done to increase the production of full-length, glycosylated monoclonal antibodies in a more cost-effective and efficient manner. Mammalian systems are ideal because the proteins are more human-like; however, the process is lengthy and costly. Microbials are more cost-effective to maintain and can produce recombinant proteins at higher yields but with some drawbacks. Strains of yeast and bacteria are currently being engineered to deal with the various intricacies of producing a functional mammalian protein in a nonmammalian system.
Microbials for the production of monoclonal antibodies and antibody fragments

Oliver Spadiut¹, Simona Capone¹, Florian Krainer², Anton Glieder³, and Christoph Herwig¹

¹Vienna University of Technology, Institute of Chemical Engineering, Research Area Biochemical Engineering, Gumpendorfer Strasse 1a, A-1060 Vienna, Austria
²Graz University of Technology, Institute of Molecular Biotechnology, Graz, Austria
³Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria

Monoclonal antibodies (mAbs) and antibody fragments represent the most important biopharmaceutical products today. Because full length antibodies are glycosylated, mammalian cells, which allow human-like N-glycosylation, are currently used for their production. However, mammalian cells have several drawbacks when it comes to bioprocessing and scale-up, resulting in long processing times and elevated costs. By contrast, antibody fragments, that are not glycosylated but still exhibit antigen binding properties, can be produced in microbial organisms, which are easy to manipulate and cultivate. In this review, we summarize recent advances in the expression systems, strain engineering, and production processes for the three main microbials used in antibody and antibody fragment production, namely Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli.

Introduction

Over the past three decades, the biopharmaceutical market has become a significant component of the global pharmaceutical market accounting for around 40% of its sales. The use of organisms as biopharmaceutical production factories offers several advantages over chemical synthesis. Microorganisms can produce high molecular weight compounds such as proteins [1] and carry out highly enantio- and regio-selective reactions by their native enzymatic machinery – these reactions are hard to achieve by chemical synthesis. The use of microorganisms also enables repeated implementation of immobilized enzymes or cells resulting in the reduction of the overall production costs [2]. Finally, processes employing microorganisms do not generate organic and inorganic pollutants, such as mercury and toluene [3].

The biopharmaceutical market originated in the late 1970s with the establishment of recombinant DNA techniques. The industrial interest materialized almost immediately and in 1982 the US Food and Drug Administration (FDA) approved the commercialization of humulin, the human insulin analog, recombinantly produced in the bacterium E. coli [4]. For a while the FDA only allowed the transformation of bacteria and the expression of small, non-glycosylated proteins, like insulin, due to concern about introducing new toxicities such as contaminating bacterial substances, which raise immunogenic reactions in patients. However, with the development of selectable resistance markers, like antibiotic resistance markers, and the possibility of production in eukaryotic organisms, the FDA began showing increasing flexibility towards biotechnological innovation, leading to a continually increasing number of approved new biological entities (NBEs). In 2012, the biopharmaceutical market turnover was estimated at around 100–120 billion US dollars per year [5], with more than 200 biopharmaceutical proteins already on the market [6], and is expected to reach 170 billion US dollars in 2014. This exceptionally high market turnover is largely derived from the marketing of mAbs and antibody fragments that currently represent the fastest growing class of approved biopharmaceutical products. In fact, production of full length mAbs (Figure 1) is the most important biopharmaceutical venture to date, with several therapeutic products reaching blockbuster status (e.g., Avastin, Herceptin, Remicade, Rituxan, Humira, and Erbitux).

More recently, interest has grown in the production of antibody fragments that can be used not only in therapeutical applications but also in immunodetection, purification, and bioseparation applications [7]. Antibody fragments still exhibit antigen binding properties and can be produced in microbials, which are easy to manipulate and cultivate. In this review, we summarize recent advances in the expression system, strain engineering, and production process for the three main microbials for antibody fragment production, namely S. cerevisiae, P. pastoris, and E. coli, and highlight ongoing research that may allow full
length antibody production in these organisms in the future.

**mAbs and antibody fragments: an overview**

A full length mAb consists of the constant Fc (crystallizable fragment) domain and an antigen binding domain, comprising the Fv (variable fragment) and the Fab region (antibody binding fragment; Figure 1). Native full length mAbs are glycosylated during their synthesis. Although the glycosylated Fc domain does not directly interact with antibodies, it stabilizes the antibody and is important for antibody-dependent, cell-mediated cytotoxicity. Moreover, glycosylation strongly impacts the clearance rate of the recombinant mAb from the body, and incompatible glycoforms can cause severe immunogenic effects in patients. Thus, much current work is focused on optimizing and controlling glycosylation events in mammalian cells [8], which at this time are the most often used cell type for the production of mAbs (Box 1).

Nevertheless, a full length antibody with a glycosylated Fc domain is not necessary for antigen recognition. In fact, both the Fv and the Fab region alone (Figure 1) exhibit antigen binding properties. Furthermore, antibody fragments show increased tissue penetration and a lower retention time in non-target tissues compared to mAbs [9]. Although the lack of the stabilizing Fc domain causes reduced stability [10], the absence of glycosylation on both the Fv and the Fab regions allows their production to be less complex and enables easier engineering and cultivation of microbial host organisms such as bacteria and yeasts.

**Microbial expression hosts for mAbs and antibody fragments**

**The yeast S. cerevisiae**

*S. cerevisiae* was the first yeast employed in the production of recombinant proteins, and several biopharmaceuticals produced in this yeast have since been successfully marketed [11]. There are several intrinsic characteristics, like the stability of the expression system and the ease of cultivation, as well as advances in host engineering, that make *S. cerevisiae* an attractive host for the production of mAbs and antibody fragments. In fact, the production of Llama heavy chain antibody fragments (Hvv) in *S. cerevisiae* already represents a well-established industrial process, ensuring production titers up to hundreds of mg/l [12].

**Expression system.** *S. cerevisiae* is easy to transform either chemically or by electroporation. There are three main types of shuttle vectors in use: (i) yeast episomal plasmids (Yep), which contain the 2 μ origin of replication, allowing gene expression without genomic integration at high copy numbers; (ii) yeast centromeric plasmids (Yep), which contain an autonomously replicating sequence and replicate with single or very low gene copy number; and (iii) yeast integrative plasmids (Yep), which lack the yeast origin of replication and are integrated into the host genome [13]. Although genomic integration of the target gene leads to a reduced expression level, it is highly desirable in terms of process quality and stability [14]. To overcome the disadvantage of low expression, targeted integration of the heterologous gene at the highly transcribed ribosomal DNA locus was developed recently [15]. In addition, commonly used promoters derived from the native glycolytic pathway, such as the promoters for glyceraldehyde-3-phosphate dehydrogenase (GAP), alcohol dehydrogenase 1 (ADH1), phosphoglycerate kinase (PGK), and phosphoglycerate kinase (PGK1), allow high transcription levels [16]. Finally, new cloning strategies introduced recently allow the concomitant expression of two or more genes located on specially designed self-replicating plasmids [17], which also addresses the issue of low expression levels of heterologous genes caused by genomic integration.

**Strain engineering.** Despite continuing advances in genetic manipulation, efficient production of mAbs and antibody fragments in *S. cerevisiae* can still be impaired by endoplasmic reticulum (ER) misfolding and inefficient trafficking. Although Hvv can be produced successfully in sufficient amounts [12], the expression of the significantly smaller single chain Fv (scFv) region (Figure 1) leads to intracellular accumulation of misfolded proteins in the ER or in vacuolar-like organelles. A possible explanation for this is the higher hydrophobicity of the variable light and heavy chains of scFv compared to Hvv [18]. However, additional overexpression of chaperones and foldases can correct protein folding and allow subsequent scFv secretion [19].

Several strategies have been developed to increase the overall secretory capacity and productivity of *S. cerevisiae*. These approaches include engineering intracellular protein trafficking by over-expression of soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor proteins (SNAREs) [20], reduction of proteolytic degradation by multiple protease gene deletions [21], and engineering of the heat shock response (HSR) pathway by overexpressing the heat shock transcription factor (Hsf) [22]. Although these engineered strains have not yet been used for the production of mAbs and antibody fragments, they demonstrate the ongoing, intensive strain engineering work that is being done with *S. cerevisiae*. 

---

**Figure 1.** Schematic view of a full length antibody (the antigen binding sites are highlighted in yellow).
Box 1. Production of mAbs in mammalian cells: advantages and drawbacks

Mammalian cells are used most often for production of mAbs due to their ability to perform post-translational modifications (PTM), especially human-like N-glycosylation. Their use simplifies subsequent medical applications by eliminating the risk of an immunogenic response in patients due to incompatible N-glycans on the protein. Chinese Hamster Ovary (CHO) cell lines are used most frequently to generate full length mAbs with human-like Fc N-glycosylation and production titers of around 10 g/l [8]. However, the use of mammalian cells for heterologous protein expression holds several drawbacks such as low product yield and growth rate, risk of viral contamination, and requirement for serum. Despite the introduction of serum-free (SF) chemically defined media (CDM) encountering regulatory requirements [56], the addition of chemically undefined hydrolysates is still necessary to support cell growth. This, however, highly contradicts QbD guidelines demanding defined growth media [57]. Furthermore, the current standard production process is cumbersome and time-consuming. Cell transfection leads to high clone heterogeneity, necessitating repeated screening procedures at increasing drug concentrations for the isolation of a positive, highly productive clone [8]. Clone evaluation and culture condition optimization is then performed in shake flasks and lab-scale bioreactors before production processes can be set up. However, scale-up is also very challenging. The catabolism of the main carbon sources, glucose and glutamine, leads to formation of the inhibiting metabolites lactate and ammonium, respectively; hence batch and fed-batch operation modes are used to represent closely the final production system, which is only possible for a restricted timeframe. Because the metabolism of mammalian cells is highly sensitive and responsive to changing culture conditions, bioprocesses are hard to model – in fact only unstructured models are possible – and to control, which again contradicts QbD guidelines [57]. Consequently, chemostat cultivations, which describe open cultivation systems where substrate is constantly fed and cultivation broth is continuously removed, are generally employed to avoid metabolite inhibition. To avoid a critical wash out of mammalian cells, perfusion systems that provide cell retention by employing membranes are mainly used. However, operating a continuous culture with a perfusion system requires more devices and control systems than a batch or fed-batch system and also bears the elevated risk of contamination. Another drawback associated with scaling-up mammalian cell cultures is their sensitivity to shear stress, creating further challenges to efficient aeration in large vessels. Thus, although mammalian cells can produce mAbs with compatible PTMs, several drawbacks in bioprocessing are yet to be overcome.

Production process. Production of antibody fragments in *S. cerevisiae* is generally done in glucose-limited fed-batch cultivations [12]. Yeast shows a mixed oxidative/fermentative metabolism, which can result in the undesired production of toxic metabolites. Fermentative mode shift is triggered by oxygen depletion or by elevated carbon source concentration. Limiting glucose is therefore a valid strategy for preventing fermentation during cultivation processes with this yeast. Recently, a fully aerobically engineered strain, in which glucose uptake was reduced, was developed, allowing a full aerobic respiration even at elevated glucose concentrations [23].

As this discussion indicates, there are ongoing efforts to optimize the yeast *S. cerevisiae* for the production of mAbs and antibody fragments. Because antibody fragments are not glycosylated, they can be produced successfully in this yeast and are not affected by hypermannosylation, which characterizes *S. cerevisiae* [24]. Furthermore, current studies are investigating the possibility of humanizing the glycosylation machinery in *S. cerevisiae* [25], in an attempt to engineer this yeast for the production of full length mAbs.

The yeast *P. pastoris* As an alternative to *S. cerevisiae*, the methyloptic yeast *P. pastoris*, which is closely related to *S. cerevisiae*, can be used for the production of mAbs and antibody fragments as it also holds a generally recognized as safe (GRAS) status [26].

Expression system. Similar to the process in *S. cerevisiae*, the target gene is integrated into the genome of *P. pastoris* to guarantee reproducibility and stability of the expression system. However, a major obstacle in *P. pastoris* is the substantial degree of non-homologous recombination. One solution to this challenge is the use of a recently developed *P. pastoris* strain with an inactivated non-homologous end joining pathway [27].

*P. pastoris* can use methanol as a sole carbon source, as it is a crucial part of its metabolism (e.g., [28]). However, instead of the traditional hard-to-control alcohol oxidase promoter system typically used for *P. pastoris*, alternative adjustable promoters are currently under investigation [29]. Furthermore, the generation of artificial and semi-artificial, tunable promoter variants are the subject of recent synthetic biology approaches [30].

Strain engineering. The genome sequences of the wild type strains NRRL Y-1603 (identical to DSMZ 70382 or CBS704) [7], NRRL Y-11430 (identical to ATCC 7673 or CBS7435), and GS115 are available online [31,32] and a genome-scale metabolic model of *P. pastoris* was published recently [33], allowing straight-forward strain engineering approaches. For example, co-overexpression of helper proteins, such as the protein disulfide isomerase or the transcription factor of the unfolded protein response Hac1 [34], as well as inactivation of endogenous proteases (e.g., [35]) enhances the production and secretion of recombinant proteins. Engineering the protein trafficking pathway represents another successful approach to improve secretion [36]. In addition, intensive glycoengineering work is ongoing to humanize the glycosylation events in *P. pastoris* and allow production of full length mAbs in this yeast (Box 2).

Production process. In contrast to *S. cerevisiae*, *P. pastoris* prefers respiratory over fermentative growth, allowing cultivations to very high cell densities, for example, 160 g/l cell dry weight [37], on inexpensive, defined media without the risk of accumulating ethanol. The very well-studied production processes in *P. pastoris* are most commonly performed as fed-batch processes. The possibility of performing mixed-feed fed-batch cultivations, where two substrates are concomitantly fed facilitating biomass growth due to higher biomass yields on the second substrate and leading to lower oxygen consumption and lower heat production, is a significant advantage of yeasts over mammalian cells and has already been applied successfully for the production of scFvs with *P. pastoris* [38]. In addition, a recent study presented a dynamic approach for determining strain-specific parameters in simple batch
Box 2. Glycoengineering of *Pichia pastoris* allows mAb production

*P. pastoris* can be used for the production of both antibody fragments and mAbs (e.g., [58]). For mAbs, the correct human-type glycosylation is not only essential for proper folding and biological activity, but also for targeting and stability in circulation. *P. pastoris* lacks the Golgi-resident α-1,3-mannosyltransferase, but harbors four additional β-mannosyltransferases instead [59,60]. The absence of terminal α-1,3-mannoses on *P. pastoris*-derived glycoproteins is of importance because this glycan structure causes high antigenicity in humans [61]. Thus, the humanization of the N-glycosylation pathway in *P. pastoris* has been an important goal. The Outer Chain elongation 1 gene (*OCH1*) coding for an α-1,6-mannosyltransferase was knocked out [62], and an α-1,2-mannosidase, β-N-acetylglucosaminyltransferase I (*GntI*) and an UDP-GlcNAc transporter were introduced [63]. The *Klyuyveromyces lactis* lactis UDP-GlcNAc transporter, mouse α-1,2-mannosidase IA, *Drosophila melanogaster* mannosidase II, human GntI and, rat GntI were introduced into an och1 knockout strain, resulting in the homogeneous formation of the complex human GlcNAc3Man3GlcNAc2 glycan [64]. In other studies, *OCH1* was inactivated via a knock-in strategy [65], an ER-targeted HDEL (His-Asp-Glu-Leu; C-terminal tetrapeptide involved in the lumen sorting of soluble proteins)-tagged α-1,2-mannosidase from *Trichoderma reesei* was introduced, and a chimeric human GntI was fused to the N-terminal part of *Saccharomyces cerevisiae* Kre2 for Golgi localization [66]. A further approach included the construction of a strain expressing mouse mannosidase IA, the *K. lactis* UDP-GlcNAc transporter, human GntI, and rat GntI, in which the ALG3 gene, encoding an α-1,3-mannosyltransferase of the ER lumen, was knocked out [67], leading to the formation of GlcNAc2Man3GlcNAc2. Additional coexpression of a fusion protein consisting of the *S. cerevisiae* Mnn2 Golgi localization domain and the activities of *Schizosaccharomyces pombe* UDP-Gal 4-epimerase and human β-1,4-galactosyl transferase allowed the production of Galα1GlcNAc2Man3GlcNAc2 glycans. An alternative protocol allowed production of Galβ1GlcNAc2Man3GlcNAc2 N-glycans using the GlycoSwitch vector technology [68], where specially designed vectors are used to replace genes of the native glycosylation pathway. Further humanization was achieved by additional biosynthesis of cytidine monophosphate-linked Sia, its transport and the transfer of Sia onto the N-glycans of nascent polypeptides, leading to complex human Sia2Galβ1GlcNAc2Man3GlcNAc2 glycan [69]. Additional glycoengineering studies included the elimination of α-1,2-mannosidase-resistant high Man glycans [70] and overexpression of *Leishmania major* ST3D to increase N-glycan site occupancy [71]. These steps make it possible to use glycoengineered *P. pastoris* strains for the production of full length mAbs (e.g., [72]).

Cultivations. This approach enables the design of efficient mixed-feed strategies for this yeast [39].

In conclusion, *P. pastoris* is a well-established host system for the production of antibody fragments. In fact, two recombinant therapeutic antibody fragments are already on the market: Nanobody ALX0061, which is a recombinant anti-IL6 receptor single domain antibody fragment used for rheumatoid arthritis treatment, and Nanobody ALX00171, a recombinant anti-RSV single domain antibody fragment used for respiratory syncytial virus (RSV) infection treatment. Given recent and ongoing advances in glycoengineering, *P. pastoris* is of increasing interest for the production of glycosylated full length mAbs (Box 2).

The bacterium *E. coli*

Due to rapid growth on inexpensive substrates, the ability to reach high cell densities, well-understood genetics, and easy genetic manipulation, prokaryotic expression systems are widely used for the production of recombinant proteins.

The gram-negative bacterium *E. coli* was the first microbial organism employed for the production of recombinant biopharmaceuticals and still accounts for nearly 40% of all the marketed biopharmaceutical compounds produced today. After the approval of humulin in 1982, several different therapeutic proteins, such as antibody fragments (e.g., the antitumor necrosis factor (TNF)α Fab), have been successfully produced in this prokaryotic organism [11].

**Expression system.** Recombinant protein expression in bacterial hosts is generally driven by self-replicating multicy copy plasmids carrying a strong promoter, like the bacteriophage T7, the *E. coli* lactose operon (*lac*) or the synthetic tryptophan operon (*trp*) promoter, and a ribosome binding site allowing high gene dosages [40]. Although greater volumetric productivity can be reached by implementing self-replicating multicopy plasmids, these cause a severe metabolic burden for *E. coli*, including cell growth inhibition and cell death. Thus, new plasmid-free expression systems, based on site-directed chromosomal integration of the heterologous DNA, have been developed [41]. In order to eliminate the metabolic burden associated with the selection marker, a novel marker-free plasmid selection system using a genomics modified *E. coli* strain was also engineered [42].

**Strain engineering.** Although cytoplasmic production in *E. coli* allows high intracellular product yields, it is often associated with inclusion body formation (e.g., [43]). This *E. coli* characteristic phenomenon arises from unbalanced expression of folding helper elements and the fact that disulfide bridges cannot be formed correctly in the reducing environment of the cytoplasm. This problem can be overcome by the co-expression of chaperones [44] or by the transport of the target protein to the periplasmic space by fusion to a leader peptide at the N terminus [45]. Secretion into the periplasm has already been successfully performed for antibody fragments [46]. However, if efficient refolding is possible, recombinant protein production in inclusion bodies also describes a valuable production strategy, as already described for Fc-fusion proteins [47].

**Production process.** Due to the intrinsic high growth rate of *E. coli*, high cell density cultures are currently used for the production of antibody fragments [48]. Production processes with *E. coli* are commonly conducted in stirred tank reactors (STR) as limited glucose fed-batch processes because glucose excess induces overflow metabolism and causes the production of the inhibiting metabolite acetate. As an alternative to a carbon source-limited feeding strategy, different metabolic engineering approaches have been designed to prevent or at least reduce acetate formation. These approaches include manipulating the native acetate formation pathway [49] and engineering the endogenous glucose uptake system [50]. Another recent advance is to improve the bioprocess via the identification and characterization of key strain-specific physiological parameters instead of excessive strain engineering. The knowledge of the strain characteristic parameters specific substrate uptake rate (*q_s*) and maximum specific substrate uptake rate (*q_{s_{max}}*), for example, allows the design of tailored
bioprocesses avoiding overflow metabolism. Soft-sensor tools, which are virtual sensors processing different signals measured online that give real-time information on a non-measurable process parameter, are powerful tools for that purpose [51]. Besides, strain-specific physiological parameters can easily be determined by applying dynamic changes of process parameters during cultivation [52]. The availability of detailed physiological data enables design that follows the Quality by Design (QbD) guidelines [53].

In summary, antibody fragments, which are not glycosylated, can be produced in E. coli and the required tools are already in place (e.g., [11]). Remarkably, successful production of full length mAbs in E. coli was achieved recently, although the mAbs were not glycosylated [47]. The identification of the N-glycosylation pathway in Campylobacter jejuni and the possibility of introducing it into E. coli [54] may pave the way for the production of the glycosylated Fc domain [55] and the successful expression of full length mAbs in E. coli. For this reason, pharmaceutical companies are now investing effort and capital in reintroducing E. coli to their production facilities.

## Concluding remarks and future perspectives

Full length mAbs as well as antibody fragments represent the most important and valuable class of biopharmaceuticals today. Due to the requirement for surface glycosylation, mAbs are still predominantly produced in mammalian cells, which possess several drawbacks relating to bioprocessing and scale-up. By contrast, antibody fragments, which are not glycosylated but retain antigen binding properties, can also be produced in microbial organisms. Recent advances in the production of full length mAbs and antibody fragments with both mammalian cells and microbials are summarized in Table 1.

As shown in Table 1, current efforts are directed towards optimizing the production of mAbs and antibody fragments in microbial organisms, because they outpace mammalian cells in several aspects, such as the ease of genetic manipulation, greater productivity, and high cell density cultivation processes on inexpensive and defined substrates. Although mAbs are still most frequently produced in mammalian cells, ongoing glycoengineering studies with yeasts (Box 2) and E. coli [54,55] are paving the way for the successful production of glycosylated full length mAbs in microbial host organisms.

## Acknowledgments

The authors are very grateful to the Austrian Science Fund (FWF) project P24861-B19 for financial support.

## References

40 Tegel, H. et al. (2011) Enhancing the protein production levels in Escherichia coli with a strong promoter. FEBS J. 278, 729–739
61 Gregg, J.M. et al. (1993) Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology 11, 905–910