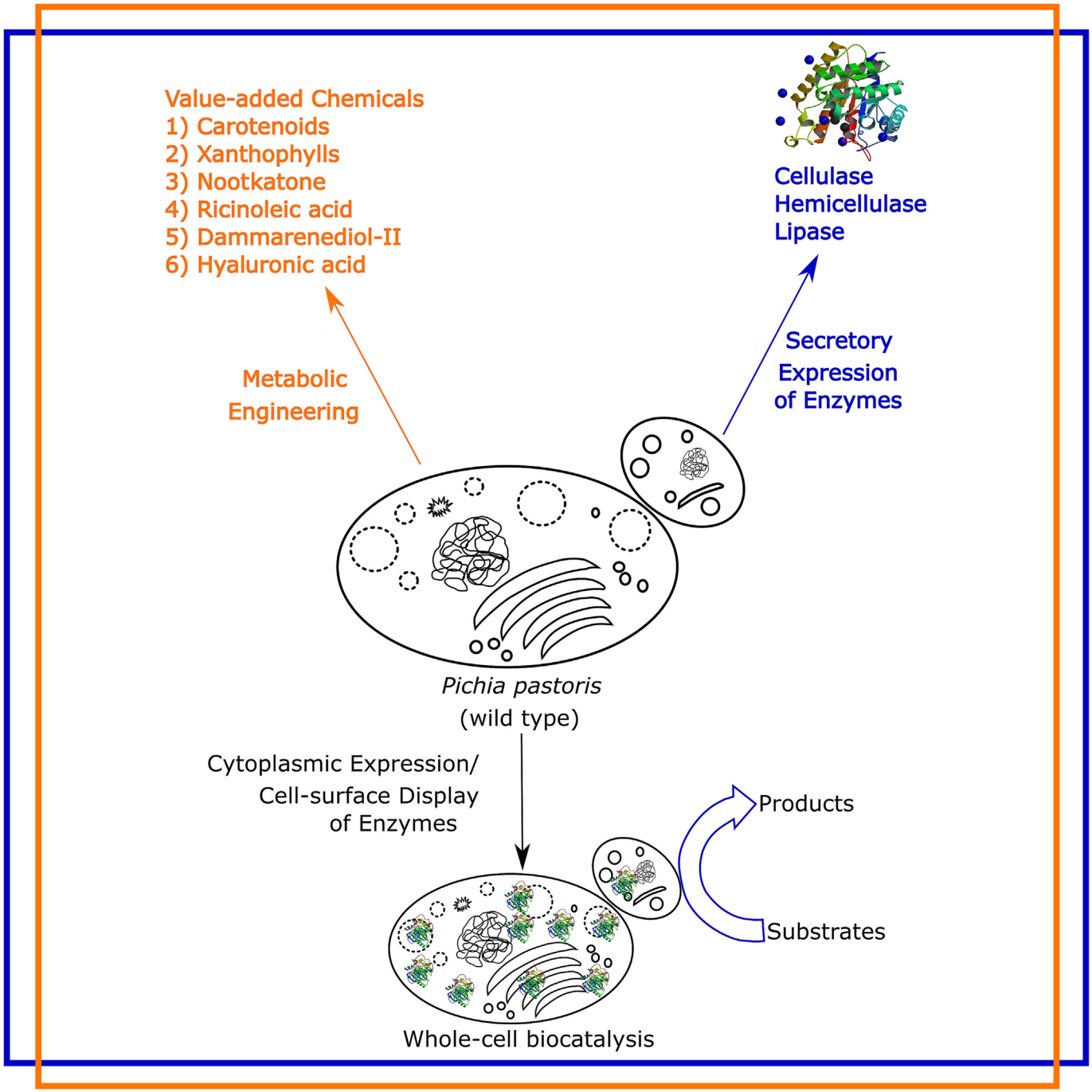
ChemPubSoc Europe
DOI: 10.1002/cbic.201700460
CHEMBIOCHEM Reviews

Heterologous Protein Expression in *Pichia pastoris*: Latest Research Progress and Applications

Veeresh Juturu and Jin Chuan Wu*[a]



Pichia pastoris is a well-known platform strain for heterologous protein expression. Over the past five years, different strategies to improve the efficiency of recombinant protein expression by this yeast strain have been developed; these include a patent-free protein expression kit, construction of the *P. pastoris* CBS7435*Ku70* platform strain with its high efficiency in site-specific recombination of plasmid DNA into the genomic DNA, the design of synthetic promoters and their variants by combining different core promoters with multiple putative transcription factors, the generation of mutant *GAP* promoter variants with various promoter strengths, codon optimization, engineering the α-factor signal sequence by replacing the native glutamic acid at the *Kex2* cleavage site with the other 19 natural amino acids and the addition of mammalian signal sequence to the yeast signal sequence, and the co-expression of single chaperones, multiple chaperones or helper proteins that aid in recombinant protein folding. Publically available high-quality genome data from multiple strains of *P. pastoris* GS115, DSMZ 70382, and CBS7435 and the continuous development of yeast expression kits have successfully promoted the metabolic engineering of this strain to produce carotenoids, xanthophylls, nootkatone, ricinoleic acid, dammarenediol-II, and hyaluronic acid. The cell-surface display of enzymes has obviously increased enzyme stability, and high-level intracellular expression of acyl-CoA and ethanol O-acyltransferase, lipase and D-amino acid oxidase has opened up applications in whole-cell biocatalysis for producing flavor molecules and biodiesel, as well as the deracemization of racemic amino acids. High-level expression of various food-grade enzymes, cellulases, and hemicellulases for applications in the food, feed and biorefinery industries is in its infancy and needs strengthening.

1. Introduction

The development of recombinant plasmids and their transformation into bacterial cells for protein expression is a big revolution in the field of recombinant protein expression.[1,2] In earlier times, different types of bacterial cells were developed as host cells for heterologous protein expression depending on their applications. Progressive increases in knowledge about recombinant protein expression and increasing market demand for recombinant proteins have widened the spectrum of proteins to be expressed heterologously.[3] The expression of eukaryotic proteins in bacterial cells predominantly ends in the formation of inclusion bodies or inactive protein aggregates due to the lack of eukaryotic post-translational modifications such as disulfide bond formation, glycosylation and secretory expression.[4]The yeast *Saccharomyces cerevisiae* has been extensively applied as a platform for protein expression owing to its multiple advantages such as secretory expression, presence of eukaryotic post-translational modifications and glycosylation. Its disadvantages are hyper-glycosylation of proteins and fermentative mode of respiration ending up in low growth hence lower yields.[5]

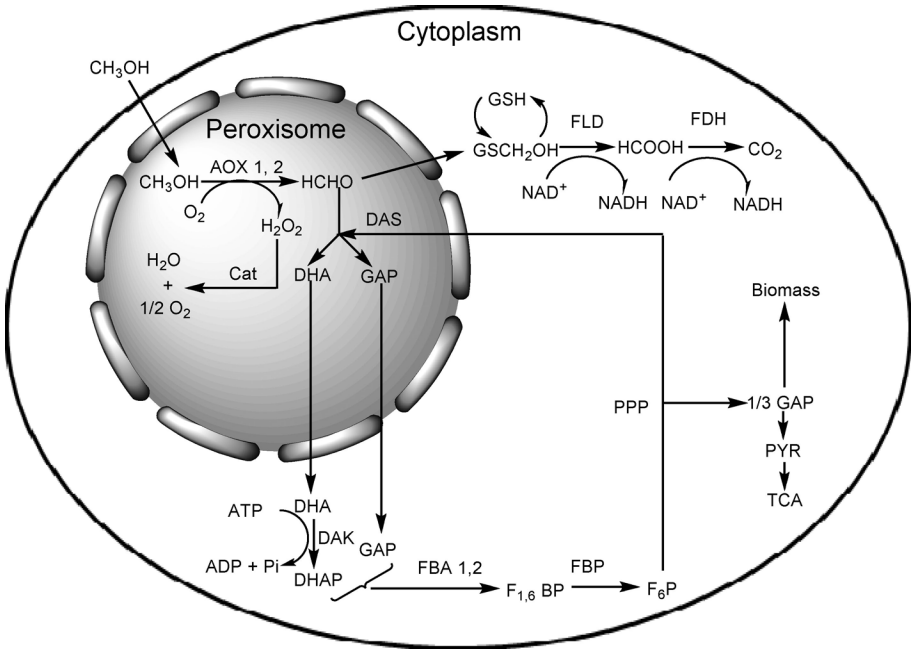
The methylotrophic yeast *Pichia pastoris* or *Komagataella pastoris* was sold by Phillips Petroleum Company as a protein-rich single-cell protein (SCP) source for feeding animals. It was grown on methanol as a carbon source. However, the soaring price of oil in the 1970s affected the economic viability of *P. pastoris* as an SCP. Later Phillips Petroleum contacted Salk Institute Biotechnology/Industrial Associates, Inc. (SIBIA) in order to develop *Pichia* as a host strain for recombinant protein production.[6,7] The advantages of using *Pichia* over *S. cerevisiae* are 1) *Pichia* uses an aerobic mode of respiration that allows it to reach a cell density as high as 130 g L^{-1};[7] 2) Protein expression under methanol-inducible *AOXI* promoter only reached 22 g L^{-1};[8] 3) It is possible to express recombinant proteins both intracellularly and extracellularly by secretory expression;[7] 4) Traditionally yeasts have been used in the food industry with a "generally recognized as safe" (GRAS) status, and 5) Metabolic engineering of the yeast glycosylation pathway so that it mimics the mammalian system has broadened the spectrum of recombinant proteins to be expressed in this host.[9,10]

Phillips Petroleum licensed its patent on the *Pichia* protein expression system to Invitrogen. With the success of this strain as a platform host, several companies and research groups in Europe, the United States and Asia have improved the first-generation expression system for efficient recombinant protein expression. Its applications have now been extended to the areas of synthetic biology and whole-cell biotransformation. Here we address the latest progress in using this methylotrophic yeast for producing fine chemicals as well as biorefinery and food applications based on recent patents and publications.[11–14]

2. Methanol Utilization Pathway (MUT Pathway)

P. pastoris metabolizes methanol as a carbon source for its energy production according to the MUT pathway. The metabolism of methanol inside the yeast cell takes place both in the peroxisome and cytosol (Scheme 1). In the first step, methanol is oxidized to formaldehyde and H_2O_2 by alcohol oxidase (AOX). The genome of *P. pastoris* contains two *AOX* genes, *AOX1* and *AOX2* that encode two alcohol oxidase enzymes. *AOX1* produces the major enzyme, which comprises up to 30% of the total soluble proteins in the extract of yeast grown on methanol as sole carbon source, thus showing the strong promoter strength of *AOX1*. *AOX2* is controlled by a weaker promoter and contributes 15% of the overall AOX activity inside the cell. H_2O_2 is broken down into water and oxygen in the presence of catalase. A portion of the formaldehyde is transported into the cytoplasm where it is oxidized into for-

[a] *Dr. V. Juturu, Dr. J. C. Wu*
Institute of Chemical and Engineering Sciences
*Agency for Science, Technology and Research (A*STAR)*
1 Pesek Road, Jurong Island, Singapore 627833 (Singapore)
E-mail: wu_jinchuan@ices.a-star.edu.sg



Scheme 1. Methanol utilization (MUT) pathway in *P. pastoris.* AOX: alcohol oxidase, Cat: catalase, DAK: dihydroxy acetone kinase, DAS: dihydroxy acetone synthase, FBA: fructose-1,6-bisphosphate aldolase, FBP: fructose-1,6-bisphosphatase, FLD: formaldehyde dehydrogenase, FDH: formate dehydrogenase.[15]

mate and CO_2 by two cytoplasmic dehydrogenases (formaldehyde dehydrogenase, formate dehydrogenase) releasing energy in the form of NADH. The leftover formaldehyde is assimilated into metabolic constituents by condensation with xylulose-5-phosphate in the presence of the peroxisomal enzyme dihydroxyacetone synthase (DAS) to produce dihydroxyacetone (DHA) and glyceraldehyde-3-phosphate (GAP). GAP and DHA enter the pentose phosphate pathway that takes place in the cytoplasm of the cell. Two enzymes, AOX and DAS, are produced when yeast cells are exposed to methanol as carbon source. Promoters of the genes involved in the MUT pathway are efficiently used in the construction of plasmids for recombinant protein expression and have been already been discussed in detail.[7,15,16]

3. Plasmids and Protein Expression Platforms

For recombinant protein expression in *P. pastoris,* it is very important to design a suitable recombinant system that includes plasmid, promoter, selection marker, secretary signal sequence, and host strain. All of these factors change with the type of recombinant protein to be expressed and its final applications. Here we briefly summarize the current status of the various types of promoter and plasmid systems, markers and platform strains that are commercially available.

Promoters used for the recombinant protein expression are of different types depending on the mode of controlled gene expression. In this methylotrophic yeast, the promoters are classified into two groups, induced and synthetic promoters. A synthetic promoter is made by bringing together the primary elements of promoter regions from diverse origins and has been included later in this review.

3.1. *AOX1* promoter

The *AOX1* gene product is the most important protein in initializing the MUT pathway. Given the high strength of this promoter, a promoter sequence from this gene was used to construct expression vectors for heterologous protein expression. The advantages of using the *AOX1* promoter are 1) transcription of foreign proteins under the control of this promoter is tightly regulated and controlled by a repression/de-repression mechanism; 2) high levels of heterologous protein expression can be achieved; 3) the repression of the *AOX1* gene by most carbon sources other than methanol ensures high cell growth before gene expression, and 4) Induction of transcription is easily achieved by the simple addition of methanol as an inducer. The disadvantages of using this promoter system include 1) online monitoring of methanol use during fermentation is difficult; 2) methanol is highly flammable, so the large-scale storage of methanol is undesirable; 3) methanol is a petrochemical product, so its application in protein production for food-grade applications is undesirable and 4) the use of two carbon sources (glucose—growth phase and methanol—induction phase) and the difficulty of switching from one to the other at a precise time point.[17–20]

3.2. *GAP* promoter

The *GAP* gene encodes a NAD-dependent glyceraldehyde 3-phosphate dehydrogenase that forms tetramers. The 500-bp upstream sequence from the translation initiation codon ATG was used to construct the expression vectors. *GAP* is a constitutive promoter expressing proteins continuously on all carbon sources including glucose, glycerol, ethanol, and oleic acid. The *GAP* promoter can be used, instead of the *AOX1* promoter,

to produce heterologous proteins in *P. pastoris*. The advantages of using the *GAP* promoter are methanol-free protein expression and no requirement for induction by shifting culture from one carbon source to another. However, the continuous production of heterologous toxic proteins can be toxic to yeast cells and lead to their death.[21,22]

3.3. *FLD1* promoter

Glutathione-dependent formaldehyde dehydrogenase (FLD) is a key enzyme required as a carbon source for the metabolism of methanol and as a nitrogen source for certain alkylated amines, such as methylamine. Its primary function in the cells is to protect the host cell from the toxic effects of formaldehyde with an additional benefit of yielding a net reducing power in the form of NADH. The *P. pastoris FLD1* promoter sequence was used to construct expression vectors, that is, P*FLD1* is a potential alternative to the routinely employed P*AOX1* expression. P*FLD1* is a highly regulated promoter capable of producing heterologous proteins at the same or higher levels than P*AOX1*. Using P*FLD1* provides a broader choice of different carbon sources for cell induction such as methanol and alkylated amines, that is, methyl amine or choline. Expression of the *FLD1* gene under *GAP* or *AOX1* promoter confers resistance to high concentrations (30 mM) of formaldehyde when compared to the wild-type strain and therefore it can be used as a selectable marker in *Pichia* and other methylotrophic yeasts.[23–25]

3.4. *ICL1* promoter

Isocitratelyase is a key enzyme that catalyzes the catabolism of isocitrate to succinate and glyoxalate in the glyoxalate cycle. The *ICL1* gene is repressed in the presence of glucose and induced in its absence or in the presence of ethanol. Due to this characteristic, use of this promoter has been considered as an attractive alternative to the conventional P*AOX1* promoter for the expression of foreign genes. However, no actual figures for protein yields under this promoter are available and no detailed studies have been conducted on the use of *ICL1* as an efficient promoter for the construction of expression vectors.[26,27]

3.5. *PEX8* and *YPT1* promoters

P*AOX1*, P*GAP* and P*FLD1* are strong promoters that express heterologous genes at high levels. High levels of protein expression can block the post-translational machinery that aids the proper folding of proteins. Thereby a significant portion of the expressed proteins might be improperly folded leading to their degradation. In a few cases, moderately expressing promoters are desirable, such as P*PEX8* directing the formation of Pex8p, which is a peroxisomal matrix protein that is essential for peroxisome biogenesis. It is expressed at a low but significant level on glucose and is induced sufficiently when cells are shifted to methanol.[28] *YPT*1 genes encode a GTPase involved in secretion, and its promoter provides a low but constitutive protein expression in media containing glucose, methanol or mannitol as carbon sources.[22]

3.6. *NPS* promoter

P. pastoris was grown under limited-phosphorus conditions to screen for the genes that are over expressed under phosphorus starvation. During this screening, an auto-inducible sodium-phosphate symporter gene (NPS) promoter and its gene were obtained. A plasmid was constructed by using *NPS* promoter with α-amylase signal sequence. The lipase gene from *Bacillus stereothermophillus* was fused with the cellulose binding protein (CBD) DNA, which was cloned and transformed into the *His4* gene loci of GS115. Recombinant clones of yeast cells grown in 50 g L^{-1} glucose, and an initial phosphorus concentration of 2.0 mM was used as fermentation medium; the lipase–CBD fusion protein was expressed constitutively after depletion of the phosphorus.[29]

3.7. Mining new promoter sequences for heterologous protein expression

To overcome the limited number of promoters, two different strategies were applied to searching for novel promoters that aid in recombinant protein expression. The first strategy, heterologous microarray hybridization, was performed with *Pichia* cDNA isolated after being grown on different carbon sources at two different pH values and hybridized with *S. cerevisiae* cDNA microarrays. Mining of the transcriptome data showed 15 genes with high expression levels, thus indicating the presence of strong promoters. The second strategy, rational selection of promoters from different yeasts based on literature mining, led to nine potential strong promoter sequences. In total, 24 potential promoter sequences were studied for their promoter activities for both intracellular and extracellular protein expression, and 80% of the promoter sequences identified from the transcriptome data showed their promoter activity on all carbon sources normally used for *Pichia* growth. The success rate from the rationally selected promoter sequences was very low. Many of these identified promoters, for example, three ribosomal (RPL1, RPS2 and RPS31) promoters and two chaperone (HSP82 and KAR2) promoters, showed a growth dependent behavior. The promoter of the thiamine biosynthesis gene P*THI11* showed high expression levels at low specific growth rate, but was suppressed by the thiamine availability in growth medium.[30]

The MUT pathway of *P. pastoris* is a complex metabolic network involving multiple genes and multiple branches, and the enzymes involved in these pathways will have various levels of differential expression. Expression levels of MUT pathway genes in *P. pastoris* after exposure to various carbon sources were verified by the microarray studies. Based on the transcriptome studies, the promoter strength was analyzed by using the *GFP* gene as a reporter system. On the basis of their strength and regulation, MUT promoters have been grouped into strongly, intermediately and weakly inducible promoters. All methanol-inducible promoters are tightly repressed on glu-

cose, some are repressed under glucose-depleted conditions, and others show constitutive expression. P*DAS1*, P*DAS2*, P*CAT1* and P*PMP20* showed stronger expression, similar to P*AOXI*, upon exposure to methanol, P*DAS2* has even surpassed P*AOXI*. P*FGH1*, P*DAK1* and P*PEX5* have moderate expression and PADH2, PTPI1 and PFBP1 show constitutive expression. Similarly, a set of 20 terminators suitable for use along with MUT promoters for multigene co-expression in yeast were studied. Based on these results, a diverse tool box for metabolic and biosynthetic pathway engineering can be set up in *P. pastoris* for coordinated and balanced co-expression of multiple genes to optimize the metabolic flux towards the desired end products.[31]

4. Selectable Markers

P. pastoris has a few selectable marker genes for molecular genetic manipulation. The plasmids used for recombinant protein expression contain antibiotic resistance markers such as *Shble*, *bsr* and *npt*II or *npt*III, which confer resistance to zeocin, blasticidin and kanamycin. Whereas plasmids used for applications in the food and pharmaceutical industries have food-grade selection markers to complement the missing genes in the auxotrophic strains, these genes originate from *S. cerevisiae* or *P. pastoris* biosynthetic pathways (*His4*, *Arg4* and *Ura3*), in which the *His4* gene encodes a trifunctional enzyme catalyzing the second (phosphoribosyl-ATP pyrophosphohydrolase), third (phosphoribosyl-AMP cyclohydrolase) and tenth (histidinol dehydrogenase) steps in histidine biosynthesis; *Arg4* encodes argininosuccinate lyase, which catalyzes the last step in arginine biosynthesis; and *Ura3* encodes orotidine 5′-phosphate decarboxylase, a key enzyme in the synthesis of pyrimidine ribonucleotides.[32]

5. Host Strains

The *P. pastoris* yeast strains used for heterologous protein expression were all originally derived from NRRL-Y 11430. Mutations in the auxotrophic genes have given rise to strains GS115 (*his4*) and GS200 (*his4, Arg4*). These auxotrophic mutants grow in complex media and minimal media supplemented with histidine and arginine. Transformation of the recombinant plasmids with the genes supplementing the replacement of mutant genes with the functional genes will help in screening of the transformants. Similarly, mutant variants of *P. pastoris* have been developed by deleting *AOXI* and *AOXII* genes. GS115 (*his4*) has two functional *AOX* genes and is capable of metabolizing methanol at a high rate, similar to that of wild type, and these phenotypes are Mut^+ (methanol utilization phenotype). Strain KM71 (*his4 arg4 aox1Δ::ARG4*) was generated by deleting the chromosomal *AOX1* gene and replacing it with *S. cerevisiaeARG4*, it relies on the weaker *AOXII* gene and grows on methanol at a slow rate; this phenotype is Mut^S (methanol utilization slow). The third type of expression host, MC100-3 (*his4 arg4 aox1Δ::SARG4 aox2Δ::Phis4*), was generated by deleting two *AOX* genes and cannot grow on methanol as carbon source; this phenotype is Mut^-or (methanol utilization minus). Recombinant proteins secreted by *P. pastoris* into the culture medium were rapidly degraded by yeast vacuolar proteases during high-cell-density fermentations. The host strains, SMD1165 (*his4 prb1*) and SMD1168 (*his4 pep4*), were developed by deleting the *PEP4* gene, which encodes proteinase A, and *PRB1*, which encodes proteinase B. The pep4 prb1 double-mutant SMD1163 (*his4 pep4 prb1*) shows a significant decrease in protease activities (www.pichia.com).

Strains obtained from Invitrogen and RCT *Pichia* technology are protected by patent/materials ownership policies prohibiting their application for commercial purposes. Alternatively, strains developed from the *P. pastoris* CBS7435 are patent free and are valuable resources for commercial applications. The CBS7435 strain provided by the Graz pool has the additional advantage of being marker free as it was constructed by using the Flp/FRT recombinase system for marker removal. To improve the efficiency of the site-specific recombination during transformation, in the same strain the *Ku70* gene was deleted to generate the CBS7435*Ku70* strain. In this strain *AOX1*, *ARG4* and *HIS4* were precisely knocked out for both extracellular and intracellular heterologous expression of proteins.[7,33] Based on our experience, yeast culture at a low OD_{600} of 0.5–1, a DNA concentration of 1 μg, a linearized plasmid with sticky ends, suitable electroporation conditions (voltage of 1.5 KV, capacitance of 25 μF and resistance of 186 Ω) and complete regeneration of transformation cultures are crucial factors for achieving efficient transformation.

6. Secretory Protein Expression and Signal Sequence

It has been reported that, during the secretory expression of proteins in *S. cerevisiae*, after surpassing endoplasmic reticulum (ER)–Golgi complex, a significant portion of the recombinant protein is retained within the yeast cell wall. To avoid such protein retention, the *GAS1* gene of *Pichia* was disrupted. Gas1p is a glycoprotein anchored to the outer layer of the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. Gas1p functions as a β-1,3-glucanosyltransglycosylase, a cross-linking enzyme that catalyzes a *trans* glycosylation with β-1,3-glucan as a substrate. Disruption of *GAS1* results in several morphological changes, such as cells losing their typical ellipsoidal shape, cell bulging and defective budding. Disruption of *GAS1* improved the expression of a lipase activity by two-fold.[34]

The secretory expression of recombinant proteins in yeast requires the presence of a signal sequence that helps the recombinant protein to enter the ER system, the initial step for secretory expression. Enzymes Kex2 and Ste13 usually chop up this signal peptide after it enters the ER lumen. In *P. pastoris*, the *S. cerevisiae* α-factor signal sequence (89 amino acids) and its truncated versions have been efficiently utilized for successful secretion followed by the *Aspergillus niger* α-amylase (20 aa) signal sequence, the *STA1* signal sequence from *Saccharomyces diastaticus* glucoamylase gene (18 aa), the signal sequence (16 aa) from the *Kluyveromyces marxianus* inulinase gene, the signal sequence (19 aa) from *S. cerevisiae* gene *SUC2*,

Table 1. List of commercially available *P. pastoris* expression kits.

Supplier	Plasmid system	Plasmid system	Internet reference
Mobitec	pPICHOLI shuttle vector system	Five different plasmids are available with *AOX1* promoter system	www.mobitec.com/
ATUM	IP free expression vectors	In total 56 expression vectors are available both for secretory and intracellular expression	www.atum.bio
Research Corporation Technologies	*Pichia* classic *Pichia* glycoswitch	Classic system has traditional plasmid system with *AOX1*, GAP, FLD1 and traditional expression strains. Glycoswitch has glyco-engineered strains for expressing biopharmaceutical proteins	www.pichia.com/welcome/
Thermo Fischer Scientific	Original *Pichia* expression kit *Pichia* glycoswitch, *Pichia* pink	Most routinely used *Pichia* expression system. Restricted by IP/material transfer agreement	www.thermofisher.com
BioGrammatics	Glycoswitch and classical	Different expression system based on *AOX1* and GAP promoter for both extracellular and intracellular protein expression. Glycoswitch has glyco-engineered strains for expressing biopharmaceutical proteins	www.biogrammatics.com/
Bioingenium	Proprietary expression system	Different expression system based on *AOX1* and GAP promoter for both extracellular and intracellular protein expression. High-level protein expression based on coexpression of chaperones	www.bioingenium.net/

the secretion signal (26 aa) from the killer toxin of *S. cerevisiae*, the secretion signal (26 aa) from chicken lysozyme, and the signal sequence (18 aa) from human serum albumin (HAS; www.dna20.com/products/expression-vectors/yeast). Recently, the native α-factor signal sequence of *P. pastoris* CBS7435 was identified and sequenced. It is larger than that of *S. cerevisiae*, the pre-sequence size is identical, but followed by an extended pro-sequence and then nine identical copies of mature α-factor. *P. pastoris* α-mating factor drives the secretion of *Bacillus subtilis* levanase from *P. pastoris* CBS7435 as efficiently as the α-mating factor signal sequence of *S. cerevisiae*.[35] All commercially available *Pichia* expression kits are listed in Table 1.

7. N- and O-linked Glycosylation in *P. pastoris*

For therapeutic proteins such as interferon, erythropoietin and monoclonal antibodies to attain their complete biological activity, glycosylation must play an important role. *P. pastoris* is capable of performing both N- and O-linked glycosylation.[36] In mammals, O-linked oligosaccharides are composed of a variety of sugars including *N*-acetylglucosamine, galactose (Gal), and sialicacid (NeuAc). Very little is known about O-linked glycosylation in *P. pastoris*, but O-linked oligosaccharides are mainly composed of mannose (Man) residues. Basically, it contains α-1,2-mannans made up of di-, tri-, tetra-, and pentameric oligosaccharides; no α-1,3-linkages were detected. It has been found, based on sequence homology, that there are five native protein-O-mannosyltransferases (PMTs) that initiate O-linked glycosylation, and they are grouped into subfamilies *PMT1*, *PMT2* and *PMT4*. The *PMT1* and *PMT2* subfamily genes play a significant role in O-linked glycosylation. Knock out of these two gene subfamilies and the addition of O-linked glycosylation inhibitors to the growth medium significantly reduced the mannosylation, thus paving the way to producing glycosylated proteins.[37]

N-linked glycosylation is an important post-translational modification. In therapeutic proteins, this glycosylation constitutes only a small fraction of the molecular mass but plays a pivotal role to its hydrodynamic volume and therefore to its pharmacodynamics behavior. N-Linked glycosylation begins in the ER with the transfer of an oligosaccharide core unit, Glc3Man9GlcNAc2 (Glc = glucose, GlcNAc = *N*-acetylglucosamine) to the asparagine residue in the recognition sequence Asn-X-Ser/Thr. This oligosaccharide unit is trimmed to Man8-GlcNAc2. From this point, glycosylation patterns start to differ between the lower and higher eukaryotes. The Golgi apparatus performs a series of trimming and addition reactions that generate oligosaccharides composed of Man5-6GlcNAc2. Secreted proteins of *P. pastoris* are decorated by Man8GlcNAc2 or Man9-GlcNAc2. Finally, *P. pastoris* oligosaccharides appear to not have any terminal α-1,3-linked mannosylation; this makes recombinant proteins unsuitable for pharmaceutical use in humans. However, *Pichia* has been engineered to the mammalian N-glycosylation pathway.[9]A significant breakthrough was achieved in engineering the N-linked glycosylation that mimics mammalian glycosylation. All the milestones in glycol engineering steps have been patented and are included in Table 2.[38] Glycoengineering, as explained so far is useful for expressing mammalian proteins, but the expression of fungal cellulases and hemicellulases requires fungal glycosylation.[39] As much knowledge is available on the glycosylation mechanism in *Pichia,* we presume that engineering this yeast to mimic fungal glycosylation would be an added advantage in the large-scale functional expression of these enzymes.

8. Whole-Genome Sequencing

Progress in researching *Pichia* to develop it as an ideal platform strain has been negatively affected by the confidentiality of the known draft sequence of this yeast. To overcome this bottleneck, Mattanovich et al. successfully sequenced and published genome sequences of both GS115 and DSMZ 70382, but with a special emphasis on DSMZ 70382.[40] In total, 9405451 bases were sequenced by combined use of Roche GS-FLX and illumina GA with a total guanine-cytosine (GC) content of 41.34%. The number of open reading frames (ORFs) was predicted to be 7935, but manual correction reduced that to 5450. Among them, 741 ORFs with introns were found, there

Table 2. List of filed and granted patents for engineering N-glycosylation pathway of *P. pastoris*.

S. no	Engineered glycosylation step	Ref(s).
1	Inhibition of high-mannose glycan structures by inactivation of the *OCH1* gene along with heterologous expression of α-1,2-mannosidase, ending in 42 mammalian-type high-mannose oligosaccharide Man5-GlcNAc2-glycans. The gateway strain for successive glycosylation engineering	[96], [97], [98]
2	Introduction of non-native β-1,2 *N*-acetyl glucosamine transferase enzymes from the groups GnTI and GnTII, ending in mammalian GlcNAc-Man5-GlcNAc2 and GlcNAc2-Man3-GlcNAc2 glycans	[99]
3	Yeast cell lines capable of producing mammalian galactosylated *N*-glycans	[100], [101]
4	Yeast cells capable of performing mammalian-type sialylated glycoproteins	[102]
5	The yeast-cell native *Pno1* gene, capable of adding mannose phosphate to the sugar chain of a glycoprotein, is replaced by the mutant version of this gene	[103]
6	The yeast-cells native *Pno1* gene, capable of adding mannose phosphate to a sugar chain of a glycoprotein, is replaced by the mutant version of this gene, thereby decreasing the addition of acidic sugars by 10%	[104]
7	Yeast cells lines capable of producing reduced O-linked glycosylation	[105]

were 194 truncated ORFs, and 4257 ORFs with annotation. Secretome analysis of *Pichia* reveals 88 putative proteins, of which the functions of 55 were clearly annotated. If this yeast was grown on glucose, only 20 different types of proteins were secreted into the fermentation broth at a detectable range, evidence of a lower native host protein contamination with the heterologous proteins.

De Schutter et al. published the genome sequence of the GS115 strain.[41] The whole genome was sequenced by using 454/Roche sequencing (GS-FLX version). This yeast has a 9.7 Mbp total chromosomal DNA condensed into four chromosomes. There are 5313 coding genes in total, the coding percentage is 79.6%, total GC% content is 41.6%, average gene length is 1442 bp and there are 4680 single exon genes. There is a total of 123 tRNA genes, and 21 copies of the 5S RNA genes are distributed over the four chromosomes. Along with the chromosomal genome, the mitochondrial DNA sequence was analyzed. It is 36119 bp long with 22% GC content and contains 16 coding and 31 tRNA genes.

Kuberl et al.[35] have published the whole-genome sequence of *P. pastoris* CBS7435, which is the parent strain for all commercially available strains for heterologous protein expression. By the combined use of pyrosequencing (454 life sciences, Roche) and sequencing-by-synthesis methods (Illumina), a 9.35-Mbp-long sequence has been determined, with special emphasis on the MUT pathway, native α-factor signal sequence and mitochondrial genome. The 9.35 Mbp genomic DNA was organized into four chromosomes, for which 5007 coding DNA sequences (CDSs) were predicted, 83.62% of genes have a single exon and 12.42% of them harbor single introns. The remaining genes contain two (2.56%), three (0.84%), four (0.18%), or even five (0.08%) introns. The 124 tRNAs and 29 rRNAs are distributed over all chromosomes. Manual annotation of the mitochondrial genome showed 15 protein-coding sequences, 25 tRNA and 2rRNA gene loci.[35] The online genome data of all the sequenced strains is publicly available in two genome databases of *P. pastoris*, http://pichiagenome.org and http://bioinformatics.psb.ugent.be/genomes/view/Pichia-pastoris.

9. Metabolic Engineering of *P. pastoris*

With increased knowledge of *P. pastoris*, the availability of genome data and development of new tools for cloning multiple genes have expanded the applications of this yeast to produce industrially important chemicals by metabolic engineering.

Nootkatone is a sesquiterpenoid. It is naturally present in grapefruit, cedar trees and vetiver grass. This organic compound is a natural insecticide and is also used as a flavoring agent in the food and cosmetic industries; however, large-scale production of this compound by chemical synthesis is not commercially viable. Metabolic engineering of *Pichia* was successfully conducted to produce this terpene from valencene. (+)-valencene is first converted to the intermediate *trans*-nootkatol and finally to (+)-nootkatone. A stable strain of *P. pastoris* CBS7435 *his4ku70* was used to co-express the engineered variant of *Hyoscyamus muticus* premnaspirodiene oxygenase (HPO), cytochrome P450 reductase (CPR) from *Arabidopsis thaliana*, *P. pastoris* native alcohol dehydrogenase (ADH), valencene synthase from the Nootka cypress and the catalytic domain of hydroxymethylglutaryl CoA reductase (tHMG1). HPO and CPR hydroxylate the externally added substrate (+)-valencene to *trans*-nootkatol. The yeast native ADH converts *trans*-nootkatol to (+)-nootkatone. In yeasts, sesquiterpenoids are derived from farnesyl pyrophosphate (FPP) according to the mevalonate pathway, in which *HMG1* and FPP are two key regulation sites. The expression of HMG1 significantly increased the titers of (+)-nootkatone to 208 mg L^{-1} for high-cell-density fermentations in the bioreactor.[42]

Panax ginseng has been used in traditional Chinese medicines for 5000 years. Dammarenediol-II is a tetracyclic triterpenoid available from this plant and has multiple pharmacological properties such as anticancer, antiaging and antioxidative functions. The tiny amounts of this triterpenoid present and the long time required for the plant to grow make large-scale extraction of this pharmacologically important compound from the plant unfeasible. *Pichia* has the native triterpenoid pathway for ergosterol biosynthesis. In the metabolic pathway, the *ERG1* gene encodes the squalene epoxidase responsible for the conversion of squalene to 2,3-oxidosqualene, which is the key branching point for the synthesis of dammarenediol-II. The

dammarenediol-II synthase gene from *P. ginseng* (*PgDDS*) was expressed heterologously in the yeast to cyclize 2,3-oxidosqualene to dammarenediol-II. To increase the flux towards 2,3-oxidosqualene, native *Pichia* squalene epoxidase was overexpressed, thus increasing the yield of dammarenediol-II from 0.030 mg g^{-1} dry cell weight (DCW) to 0.203 mg g^{-1} DCW. To decrease native competition for the conversion of 2,3-oxidosqualene to lanosterol, the gene responsible for the conversion, *ERG7*, was downregulated; this did not affect cell viability, but increased the dammarenediol-II concentration 3.6-fold to 0.736 mg g^{-1} DCW. The addition of 0.5 g L^{-1} squalene significantly increased the dammarenediol-II concentration to 1.073 mg g^{-1} DCW. This engineered strain is a good starting point for successive rounds of metabolic engineering to increase the carbon flux towards 2,3-oxidosqualene, and optimization of the fermentation conditions would further increase the triterpenoid yield.[43]

Carotenoids are generally produced by plants, photosynthetic bacteria and algae. Carotene performs multiple functions in living cells such as coloration, protecting the cells from sunlight damage and light harvesting for conducting photosynthesis. The global demand for carotenoids is increasing, as it is used as a natural coloring agent in foods, in animal feeds and as a nutraceutical.[44a] The increase in demand for lycopene and β-carotene has led to development of bioprocesses for the large-scale production of these compounds by using fermentation technology.[44b] *P. pastoris* is noncarotenogenic, but it has the native ergosterol biosynthetic pathway, transfarnesyl diphosphate (FPP) is the starting point for the synthesis of these carotenoids. Heterologous expression of genes *crtE*, *crtB* and *crtI*, which encoding GGPP synthase (EC: 2.5.1.29), phytoene synthase (EC: 2.5.1.32) and phytoenedesaturase (EC: 1.3.99.30), respectively, should convert FPP to lycopene. In yeasts, FPP is predominantly localized in peroxisomes, in *Pichia* these organelles proliferate in large numbers during growth. Three genes *crtE*, *crtB*, and *crtI* were cloned under GAP promoter of the plasmid pGAPZB (pGAPZB-EpBpI*P) targeting these enzymes to peroxisomes. After transformation, a stable mutant was selected for production studies on the various carbon sources such as oleic acid, methanol and glucose. In the presences of oleic acid and methanol, lycopene concentrations were 1.69 and 1.96 mg L^{-1}, respectively, after 69 h. Growth in yeast peptone dextrose (YPD) medium for 37 h produced 93.2 g L^{-1} wet cell weight (WCW; 19.8 g DCW) giving 51.7 mg L^{-1} lycopene, equivalent to 2.6 mg lycopene per g DCW. In contrast, when the same strain was grown in minimal dextrose medium (MD) for 39.5 h, the WCW reached 75.3 g L^{-1} (16.0 g DCW) producing 73.9 mg L^{-1} lycopene, equivalent to 4.6 mg lycopene per g DCW. An engineered yeast strain was reported to be constructed with the plasmid pGAPZB-EBI* expressing the genes intracellularly throughout the cell to produce 12–14 mg L^{-1} lycopene in shake flask cultivations, few other details were reported.[45]

Similarly, Araya-Garay et al. constructed the plasmid pGAPZB-EBI*, which expresses the genes *crtE*, *crtB*, and *crtI* from the bacterium *Erwinia uredovora* under the GAP promoter.[46] The plasmid was integrated into the genome DNA by transformation. Stable clones were grown in shake flask cultures in YPD broth for 72 h. Lyophilized cells were treated with DMSO to extract the lycopene, yielding 1.141 μg lycopene per g DCW. In additional, the *crtL* gene from the fig fruit *Ficus carica* was cloned to the same plasmid to get the plasmid pGAPZB-EBI*L. This *crtL* gene encodes β-lycopene cyclase (EC: 5.5.1.19), which catalyzes the cyclization of lycopene to β-carotene. Shake flask fermentation of the stable mutant Pp-EBIL yielded 339 μg β-carotene per g DCW.

Xanthophylls are predominantly used as a food ingredient in the aquaculture and cosmetic industries. Astaxanthin is a type of xanthophyll and can be derived from β-carotene; two enzymes β-carotene ketolase (*crtW*) and β-carotene hydroxylase (*crtZ*) catalyze the conversion of β-carotene to astaxanthin. Araya-Garay et al. constructed β-carotene-producing *Pichia* strain Pp-EBIL as described above and used it as a platform strain for further metabolic engineering modifications.[47] Two genes, *crtW* and *crtZ*, from *Agrobacterium aurantiacum* were cloned under the GAP promoter to generate the plasmid pGAPZA-WZ. Pp-EBIL was transformed with the plasmid to generate the mutant strains Pp-EBILWZ, which produces xanthophyll. Pp-EBILWZ was grown in shake flask cultures for 72 h in YPD broth yielding 3.7 μg g^{-1} DCW. The lower yields of xanthophyll were ascribed to the significant accumulation of reaction intermediates carotene and β-lycopene; which was caused by the differences between synonymous codon usage of the heterologous genes in *Pichia* and the native host. It expected that this can be overcome by codon optimization.

Hyaluronic acid (HA) is a naturally available polysaccharide consisting of repeating units of (1,4)-β-linked glucuronic acid and (1,3)-β-linked *N*-acetylglucosamine. HA has wide application in the biomedicine, cosmetic and food industries. The biosynthesis of HA requires polymerization of two nucleotide sugar precursors UDP-glucuronic acid and UDP-*N*-acetylglucosamine. This reaction is catalyzed by five different enzymes, hyaluronan synthase (HasA), UDP-glucose dehydrogenase (HasB), UDP-glucose pyrophosphorylase (HasC), pyrophosphorylase (HasD), and phosphoglucose isomerase (HasE). HA with a high molecular weight of >2 MDa was produced in *P. pastoris* GS115. *Pichia* lacks native *hasA* and *hasB* genes. *Xenopus laevis xhasA2* and *xhasB* genes and the yeast native *hasC*, *hasD and hasE* genes were cloned in two plasmids, pAO815 and pGAPZB, in various combinations for overexpression. The linearized plasmids were transformed into the same yeast cells for genomic integration. Fermentation at 30 °C produced 1.2 MDa HA polymers. Replacement of the strong *AOXI* promoter with the weak *AOXII* promoter for *xhasA2* increased the MW to 2.1 MDa. Further decreasing the fermentation temperature to 26 °C improved the polymerization rate of HA giving 2.5 MDa HA. The final yield of HA by the recombinant strains in a 1 L fermenter was 0.8–1.7 g L^{-1}.[48]

Ricinoleic acid is *cis*-12-hydroxyoctadec-9-enoic acid that is generally obtained from the plant *Ricinus communis* and has wide applications in the chemical, pharma and cosmetic industries. The presence of the natural potent toxin ricin inhibits production of the fatty acid from cotton seeds. *Pichia* has been metabolically engineered by a "push (synthesis)" and "pull (as-

sembly)" strategy to produce this hydroxyl fatty acid. A fatty acid hydroxylase (*Cp*FAH) for synthesis of ricinoleic acid and a diacylglycerol acyl transferase (*Cp*DGAT1) for synthesis of triacylglycerol from *Claviceps purpurea* were co-expressed under the *AOXI* promoter of pPICZ-B both in a diploid wild-type strain and a haploid mutant strain defective in Δ12 desaturase activity. During ten days of induction in the wild-type strain with co-expressing the two enzymes, the percentage of ricinoleic acid reached 40% on day 4, then started to decline from day 5 reaching 13% on day 10. For a diploid control strain expressing *Cp*FAH, the percentage of ricinoleic acid reached 40% on day 3, then started to decline, reaching 4% on day 10. Co-expression of the two enzymes in the haploid mutant strain produced a maximum of 56% ricinoleic acid on day 3 and reached a steady concentration of 36% from day 8. A haploid control strain with *Cp*FAH alone produced 51% ricinoleic acid on day 3 and reached a steady concentration of 13% from day 8. The co-expression mutant strain at the highest time point had a maximum of 495 $\mu g\,mL^{-1}$ ricinoleic acid. Inside yeast cell, the hydroxyl fatty acid was predominantly localized in the neutral lipid fractions in the free fatty acid form.[49]

10. Strategies for Improving Recombinant Protein Production in *P. pastoris*

After transformation, the yeast clones usually produce recombinant proteins on the milligram scale. Various strategies have been applied to increase the level of protein production; they have been discussed in detail in a review by Ahmad et al.[50]

10.1. Codon optimization

High-yield functional expression of recombinant proteins has been achieved by codon optimization of the coding DNA sequence of interest. Significant progress has been made in determining the role of codon bias and its impact on the overall stability of mRNA and protein expression. It has been demonstrated that there is a significant difference in codon utilization of a gene in native and expression hosts leading to low levels of recombinant protein expression. Codon usage of a gene in an expression host can be analyzed by using free web tools such as graphic codon usage analyzer.[51] Codon optimization of the gene can be done by replacing the rarely using codons with frequently used ones.[52,53] A version of *xynB* from *Thermotoga maritime* was optimized by replacing the codons TCG (Ser), CTC (Leu), AGC (Ser) and GCG (Ala), which have less than 10% usage, with frequently used ones; 77.8% of the native gene sequence was maintained, and the G+C content changed from 42.7 to 43.1%. The codon-optimized sequence was cloned under *AOXI* promoter of pPIC9K plasmid and transformed into the GS115 strain. After 228 h of induction, the xylanase concentration in the supernatant reached 40020 $U\,mL^{-1}$ with a total protein concentration of 10.1 $g\,L^{-1}$ in a 5 L fermenter. In comparison to native gene protein expression, protein expression in the optimized gene increased 2.8-fold.[54]

10.2. Introduction of artificial glycosylation sites

Glycosylation is one of the most important post-translational modifications to play a role in proper protein folding and secretion. N- and O-glycosylations are the two major modes that affect the turnover number of the recombinant proteins. Protein glycosylation occurs in the lumen of the ER. O-Mannosylation is the major form of O-glycosylation; it plays an important role in protein maturation but its role in improving the recombinant proteins in yeasts is not yet well elucidated.[55] On the other hand, *Pichia* N-glycosylation has been well studied, and it has been engineered to produce tailor-made glycosylated proteins. Asn-Xaa-Ser/Thr is the amino acid motif for N-glycosylation, in which Xaa is any amino acid except proline. Interestingly, in *Pichia*, the motif Asn-Xaa-Thr is more efficiently glycosylated than Asn-Xaa-Ser. It has been postulated that glycosylation on nascent proteins strongly influences the structure and folding kinetics due to the chaperone-like activity of glycans. The absence of N-glycosylation on the nascent polypeptide leads to misfolding of the protein, and it is degraded by the ERAD pathway, thus decreasing the productivity of glycosylated proteins.[56] β-Glucosidase from *Aspergillus terreus* was expressed in *Pichia*. The enzyme has four N-glycosylation sites (Asn224, -295, -363, and -429). To know the role of glycosylation, Asn was replaced with Gln, and the mutant enzymes were expressed in this yeast. It was found that none of the glycosylation sites except Asn224 affected the expression and activity. In the case of mutant Asn224Gln, after methanol induction, the intracellular Asn224Gln protein concentration increased within 16 h. After 20 h, the Asn224Gln protein detection level dropped to zero, and there was no protein secretion even after 24 h. These results reveal that *N*-glycan at position 224 is important for proper folding, and its lack causes protein degradation before secretion. Co-expression of chaperone *HAC*1 did not improve protein secretion, and the amount of protein was not even enough for purification; this indicates the important role of N-glycosylation in the secretory expression of glycosylated proteins.[57] At the same time, inclusion of a glycosylation site at the wrong position would significantly reduce the expression of the protein in this methylotrophic yeast,[58] thus indicating the importance of N-glycosylation *at the right sites* in heterologous protein expression. However, detailed studies on the role of glycosylation for improving the expression of cellulases and hemicellulases need to be continued.

10.3. Promoter engineering

Pichia p*GAP* is a constitutively expressing promoter and is smaller in size than the *AOXI* promoter, for these two reasons it has been used for promoter engineering. The *GAP* promoter was cloned into pGHg plasmid with yeast-enhanced green fluorescence protein (yEGFP) as a reporter protein. The *GAP* promoter was subjected to error-prone PCR for three rounds, with a final mutation rate of 4%. Sequencing results confirmed the uniform distribution of the mutations throughout the promoter. Mutant plasmids were transformed into GS115 for ge-

nomic integration. 30000 mutants with their yEGFP fluorescence varying by three orders of magnitude were screened by a plate-reader assay. From this pool, 33 mutants showing 296-fold higher fluorescence were selected to form a functional promoter library. Seven random PGAP mutants (G1–G7) were selected and sequenced. Mutations were uniformly distributed throughout the promoter, and the mutation rate was in the range of 1 (G1) to 3.4% (G5). The increase in the gene expression was confirmed by mRNA isolation and was normalized for the increase in the fluorescence activity by using wild-type *GAP* as a control. Additionally, the reporter gene *yEGFP* was replaced with *lacZ* and *ds56*, which express β-galactosidase and methionine adenyltransferase (MAT), respectively. Similar to yEGFP, the expression of β-galactosidase and MAT was increased 370- and 57-fold, respectively. The relative expression level of the reporter genes was normalized against that of wild-type P*GAP*, and the enzyme activity results correlated well with the yEGFP fluorescence.[59]

P. pastoris AOXI promoter is a very strong promoter that favors high levels of recombinant protein expression. However, a high level of expression might not be required for every protein, for example, high-level expression of toxic proteins could lead to the death of yeast cells, the and expression of some biopharmaceutical proteins requires moderate protein expression with low traffic of recombinant proteins passing through the secretory pathway so as to produce mature, functional proteins to meet the pharmaceutical grade requirement. For fine-tuned protein expression, promoter engineering was achieved by classical mutagenesis.[60,61] In contrast, Vogl et al.[14] developed variants of synthetic promoters. Four differentially regulated natural core yeast promoters (p*GAP*, p*AOXI*, p*HIS4* and p*ScADH2*) were aligned by using MultAlin to identify a general minimal consensus and generating pCore1. These four natural core promoters were scanned by Mat Inspector to identify several different putative transcription factor binding sites (TFBS). TFBS predicted in these natural promoters were incorporated into pCore1 by removing unwanted sequences of pCore1 to generate pCore2. pCore1 and Core2 were fused to the upstream region of p*AOX1*. These fusions were tested for protein expression by using GFP as reporter. pCore1 slightly surpassed the background fluorescence; the re-engineered pCore2 showed tight repression on glucose. Upon methanol induction, the fluorescence reached about 10% of that of wild-type pCore *AOX1*. These results show that functional synthetic yeast core promoters can, in principle, be obtained by complementation of a core sequence elements with additional nucleotides.[14,62,63]

P. pastoris AOXII promoter is a weakly expressing promoter. It was engineered to increase the expression level of heterologous genes for recombinant proteins by three strategies. 1) Deletion of the 845–960 (≈115) base pairs upstream from the nucleotide sequence 1187 improved recombinant HSA production 16- to 26.8-fold. (The deleted region was predicted to contain a transcription-repressing sequence). 2) During subculturing, the *AOX1*$^-$ mutant strain accidentally showed two mutant strains that can grow actively in the presence of methanol. This growth is presumed to be due to the presence of the natural mutation in the *AOXII* promoter. Sequencing of these mutant promoters showed a point mutation, replacing the thymine (T) nucleotide at position 1274 with cytosine (C), and duplication of the nucleotide sequences between bases 1296–1314. These changes increased the expression of HSA 60- and 40-fold. The increase in promoter activity caused by these mutations is presumed to be due to the presence of changes in the potential transcription inhibitor sequence at 1274–1314; 3) A mutant *AOXII* promoter was constructed by including the point (T to C1274) and duplication mutations (1296–1314) and transformed into yeast strains. Recombinant HSA was increased twofold compared to the wild type.[64,65]

10.4. Signal peptide engineering

The *S. cerevisiae* α-factor signal sequence is routinely used for the secretory expression of proteins. This signal peptide sequence has a single *Kex2* cleavage site between Arg and Glu (nucleotide bases 1195–1196). Yang et al. replaced the glutamic acid at the *kex2* cleavage site with each of the other 19 natural amino acids in plasmid pPICZαA, and the effect on protein expression was studied by using Venus fluorescent protein and luciferase as reporter genes. All the recombinant plasmids were transformed into the X-33 strain. Single copy strains were selected based on their ability to resist lower concentrations of zeocin, and clones were grown and induced with methanol. Fluorescence and luciferase activities were recorded by using a plate reader. The fluorescence measurements showed that replacement of Glu with Ser gave 13-fold higher expression than the mutant with Tyr. In the case of luciferase activity, the mutants with Ala, Asp, Lys, Asn, and Ser showed higher luciferase activity than those with Cys, Leu, Gln, and Trp. Application of the replacement of amino acids was verified with biopharmaceutical proteins stem cell factor (SCF) and fibroblast growth factors 16 and 20 (FGF-16; FGF-20) as reporter genes. Substitution with Val, Phe and Trp significantly improved the secretory expression of SCF, FGF-16 and 20, respectively. Artificial addition of the *Kex2* protease gene into the host genomic DNA expressing Venus fluorescence protein doubled the productivity of the recombinant protein.[66] The mammalian salivary α-factor signal sequence (SS) was engineered to increase the secretory expression of *Rhizopus oryzae* glucoamylase (GA). The signal peptide sequence was engineered by replacing Ser8 with Leu as well as addition of five amino acids (Ala, Pro, Val, Asn, and Thr) from the yeast α-factor signal peptide region. The yeast clone with a single copy of mutant signal sequence plasmid increased the protein expression level 3.6-fold compared to the clone with one copy of the wild-type plasmid. To further increase the secretory expression, the *Sec4p* gene, which is involved in the vesicular transport of recombinant proteins, was overexpressed. The yeast clone with seven copies of recombinant plasmids and co-expression of Sec4p increased secretion 100-fold compared to the mutant containing the single copy of GA under wild-type SS, corresponding to *R. oryzae* GA secretion in *P. pastoris* increased from 3.085 to 308.986 μg in 1 mL of culture medium.[67] Promoter and signal

peptide engineering look like powerful tools for improving recombinant protein expression.

10.5. Cell-surface display

Yeast surface display (YSD) is a strategy that has been employed to increase enzyme stability and recyclability in order to reduce production costs. Lipase B (LiB) from *Candida antarctica* was successfully expressed by YSD over two different native anchor proteins Flo9 (FLO9LIPB) and Pir1 (PIRLIPB) in *P. pastoris*. Both enzymes showed high activity (>100 and $>80\ U\,mg^{-1}$ DCW, respectively) towards tributyrin at pH 7 and 45 °C in organic solvents. However, the PIRLIPB construct showed 17% higher specific activity than the FLO9LIPB. Lipase immobilized on the yeast cell surface showed a conservation of 85% of its initial activity at 40 °C and 40% at 45 °C after 3 h, whereas free enzyme lost 60% of its activity within 10 min at 45 °C.[68]

P. pastoris was genetically modified to display the minicellulosomes on its surface. The *S. cerevisiae* anchoring protein Flo1P CDS was fused to the truncated version of the scaffoldin CDS subunit of CipA from *Clostridium acetobutyltyticum* and expressed under α-factor signal sequence in *P. pastoris* X33-FSCipA. Endoglucanase (NtEG) from the wood termite *Nasutitermes takasagoensis* cds was fused with the *C. acetobutyltyticum* dockerin cds (NtEGD) and expressed under α-factor signal sequence in *P. pastoris* X33-ND. Both strains were grown in shaking flask cultures. Expression and displaying of FSCipA were confirmed by immune fluorescence microscopy and western blot. Expression of the endoglucanase dockerin fusion NtEGD and its assembly over FsCipA was confirmed by western blotting. Free NtEGD hydrolysis efficiencies over carboxymethyl cellulose (CMC), microcrystalline cellulose and filter paper were 1.2-fold, 1.7-fold, and 2.4-fold, respectively, and those anchored on the cell surface were enhanced up to 1.4-fold, 2.0-fold, and 3.2-fold, respectively. These results indicate that the cell-surface-anchored minicellulosomes are capable of enhancing the hydrolytic abilities of cellulases, which are incorporated into minicellulosomes by cohesion–dockerin interactions.[69]

11. Expression of Fusion Proteins and Co-expression of Multiple Enzymes

Trichoderma reesei endoglucanase II and cellobiohydrolase II genes were fused in frame and separated by the self-processing 2A peptide sequence from the foot-and-mouth disease virus. The fusion construct was cloned under *AOXI* promoter for secretory expression. The recombinant plasmid was linearized and transformed into the yeast. Transformants producing a high level of enzymes were screened over 0.5% CMC plates stained with Congo Red after 24 h of induction on the plates. Clones that showed a clear halo with a wide diameter were selected for expression in shaking flask cultures. The purified enzyme was able to cleave filter paper to produce $0.37\ g\,L^{-1}$ glucose along with cellobiose and cellotriose.[70]

Enzymes like phytase and xylanases are routinely added to commercial animal feeds to improve their nutritional value. Phytase and endoxylanase were co-expressed in yeast by using the 2A-peptide system. The two enzyme coding sequences were linked by the 2A-peptide coding sequence, fused with α-factor signal sequence and expressed under *AOXI* promoter for independent secretory expression of the two enzymes. After transformation and induction of the clones in shake flasks for three days, the individual expression of the two enzymes was at a 1:1 ratio. The specific activities of the expressed phytase and endoxylanases were 9.3 and $97.3\ U\,mg^{-1}$, respectively, which are 4.4- and 1.3-fold less than those the individually expressed enzymes. The enzymatic properties of co-expressed enzymes were similar to individually expressed recombinant enzymes.[71]

11.1. Co-expression of chaperones and helper proteins

A platform strain of *P. pastoris* expressing β-mannase was generated by transforming the linearized recombinant plasmid encoding β-mannase cds from *A. niger* GIM3.452. After methanol induction, the β-mannase activity reached $26.6\ U\,mL^{-1}$. To increase the enzyme production level, a second plasmid expressing the native chaperone protein disulfide isomerase (PDI) was co-expressed in the platform strain. Co-expression of the chaperone increased the protein expression level to $40\ U\,mL^{-1}$, which is 1.5 times higher than that of the platform strain. Optimization of the induction conditions and continuous induction for seven days gave β-mannase activity up to $222.8\ U\,mL^{-1}$ in the supernatant.[72,73]

A stable platform strain *Pichia* H43 for secretory expressing *Aspergillus fumigatus* endoxylanase was constructed under the control of GAP promoter to give an endoxylanase activity of $247\ U\,mL^{-1}$. The stable mutant strain was then transformed with the recombinant plasmids to express native ER resident chaperones *HAC1* and *ERO1* generating PHE strains. Co-expression of the two native chaperones increased protein expression by 15–25% when compared to *Pichia* H43. Similarly, strain PHEB was developed by co-expressing three native chaperones *HAC1*, *ERO1* and *BIP*; endoxylanase expression in the PHEB strains increased by 45–57% compared to *Pichia* H43.[74]

P. pastoris strain SMD1168 was transformed with pGAPαA plasmid to express 2F5Fab for secretory expression. Fourteen different helper proteins (PDI, CUPS, SSA4, BMH2, KINZ, KARZ, HACl, ERO1, SSE1, BFR2, COG6, SS02, COYl, and IMHl) that aid the proper folding of the recombinant protein from *S. cerevisiae* were co-expressed individually, and the increase in the production of recombinant protein was verified by shaking flask fermentation. The increase in recombinant protein expression was in the range 1.5–2-fold.[75]

12. Intracellular Expression and Whole-Cell Catalysis

The flavor molecules of medium-chain esters have a lot of applications in the food, fragrance, cosmetic, and paint industries. Biosynthesis of these compounds is possible by alcohol lysis catalyzed by acyl-CoA and ethanol O-acyltransferases. In *S. cerevisiae, Eht1* and *Eeb1* code these enzymes. Their cds were am-

plified from wild-type *Saccharomyces* genomic DNA and cloned into plasmid pPIC9K for individual extracellular expression in *P. pastoris* GS115. Successful functional expression of *Eht1* and *Eeb1* produced (100.3 ± 10.3) and (113.6 ± 9.7) mg L^{-1} proteins with the esterase activity of (74.5 ± 10.5) and (81.6 ± 12.7) U mg^{-1}, respectively. After 12 h of induction, the volatile compounds in the culture supernatant produced were confirmed to be ethyl hexanoate (C_6), ethyl octanoate (C_8) and ethyl decanoate (C_{10}). Production by *P. pastoris* EHT1 was at 2, 10, and 11%, respectively, and by EEB1 at 1, 7, and 13%, respectively.[76]

A multicopy strain of *P. pastoris* intracellularly expressing *Thermomyces lanuginosus* lipase (Tll) was constructed. The selected strain showed a specific activity in degrading olive oil of 0.73 U mg^{-1} DCW, which is twice as high as that of the commercially available immobilized Tll-lipozyme TLIM. The whole-cell catalyst (WCC) was permeabilized with 25% isopropanol and lyophilized. The lyophilized and permeabilized WCC was able to convert waste cooking oil to biodiesel with 82% yield within 84 h at 6% loading of whole cells. The thermal stability and short-chain-alcohol tolerance of the WCC were higher than those of the commercial enzyme. The WCC can be recycled three times and still retain 78% of its initial activity.[77]

Trigonopsis variabilis D-amino acid oxidase (*Tv*DAO) is an industrially important enzyme with diverse applications. This enzyme catalyzes the racemic conversion of D-amino acids to L-amino acids in the presence of a catalase, which immediately splits the hydrogen peroxide generated by the racemic reaction into water and oxygen. The stability of *Tv*DAO is affected by both catalase and reaction conditions. Intracellular expression of the enzyme and WCC is an ideal solution. *Tv*DAO was codon optimized and expressed under *AOXI I* promoter, enzyme was targeted into the peroxisomes by replacing the native peroxisomal targeting sequence (-Pro-Asn-Leu) with the sequence -Ser-Lys-Leu. The advantage of targeting the enzyme into peroxisomes is the co-existence of a native catalase that produces O_2 for the oxidase reaction through the hydrolysis of H_2O_2. A transformant (*Tv*DAO-SKL) showing a fivefold improvement in enzyme activity was obtained. A multicopy strain Tv1_mc (16–17 copies) with an intracellular activity of 1283 U g^{-1} WCW and a biomass of 117 g WCW/batch was obtained. The cells with highest enzyme activity were obtained from methanol feeding (3 mL h^{-1}) and freeze dried. Lyophilized cells were permeabilized by using isopropanol; permeabilized cells showed 49% intracellular activity and were notably stable during the conversion of D-methionine using vigorous aerations.[78]

Intracellular expression of enzymes in *P. pastoris* for application as whole-cell biocatalysts in biotransformation is a new and intelligent application. Whole-cell biocatalysis has many advantages such as hosting multiple enzymes for cascade reactions, intercellular recycling of cofactors, effective protection of enzymes from the outside environment, easy recycling of cells for multiple rounds of reactions, and easy separation of cells from the products. The glycosylation function of *P. pastoris* gives it an additional advantage for expressing enzymes that need glycosylation to show their activity, stability or substrate specificity for the biotransformations.

13. Production of Cellulases and Hemicellulases

A wide range of proteins including biopharmaceutical proteins, membrane proteins, and antimicrobial peptides have been produced in *P. pastoris* as host.[50] Here we emphasize recently expressed cellulases, hemicellulases and lipases that have applications in the biorefinery and food industries.

A gene encoding α-L-arabinofuranoside from *A. niger* ATCC120120 was successfully cloned and expressed under the *AOXI* promoter of *P. pastoris* X-33. Recombinant protein expression reached 23 U mL^{-1} after five days of induction. Purified recombinant enzyme had a molecular mass of 83 kDa. After deglycosylation, the molecular mass reduced to 66 kDa. The recombinant enzyme showed its maximal activity at pH 4 and 50 °C. The K_m and V_{max} of the enzyme on *p*-NPA were 0.93 mM and 17.86 μmol mL^{-1} min^{-1}, respectively.[79]

The endo-β-1,4-xylanase gene *xyn2* from *T. reesei* Rut C-30 was cloned for secretory expression by using plasmid pPICZαA in *P. pastoris* X-33. After three days of methanol induction, the recombinant protein expression reached 4350 nkat mL^{-1}. The expressed protein had a molecular mass of 21 kDa and showed its maximal activity at pH 6 and 50 °C. The enzyme was stable at 50 °C, retaining 94% of its activity after 30 min of incubation. The specific activity, K_m and K_{cat} of this enzyme over birchwood xylan were 4687 nkat mL^{-1}, 2.1 mg mL^{-1}, and 219.2 s^{-1}, respectively. This enzyme degrades xylan, predominantly to xylotriose, but lacks cellulolytic activity.[80]

A β-xylosidase gene *xylA* from *A. niger* was cloned into plasmid pPpHis4_GAP_BglII under GAP promoter for secretory expression. After methanol induction, 100 mg L^{-1} of protein was recovered from the supernatant. Purified enzyme showed its maximal activity at 55 °C and pH 4, with K_m and V_{max} of 1.0 mM and 250 μmol min^{-1} mg^{-1}, respectively, over 4-nitrophenyl β-xylopyranoside. The molecular mass was between 153 and 165 kDa. Xylose is a competitive and fructose a noncompetitive inhibitor of this enzyme.[81]

The *xynA* gene from *Thermobifida fusca* YX was successfully cloned and expressed in *P. pastoris* X-33. The recombinant enzyme was expressed without (rXyn11A) and with a C-terminal His_6 tag (rXyn11A-His_6) for protein purification. The specific activities of rXyn11A and rXyn11A-His_6 in the supernatants were 149.4 and 133.4 U mg^{-1}, respectively. These specific activities were approximately four and 3.5 times higher than that of the native enzyme Xyn11A (29.3 U mg^{-1}). The purified proteins rXyn11A and rXyn11A-His_6 showed specific activities of 557.35 and 515.84 U mg^{-1}, respectively. The specific activity of the untagged enzyme was 8% higher than that of the His-tagged one. Recombinant xylanases were most active at 80 °C and pH 8.0, and exhibited more than 60% activity between pH 6–9 and 60–80 °C. N-glycosylation enhanced the thermostability of the recombinant xylanases. These xylanases hydrolyzed beechwood xylan, predominantly to xylobiose, xylotriose, xylotetraose and xylopentaose.[82]

The *Ustilagomaydis* 521 xylanase gene (*rxynUMB*) was His_6 tagged and cloned into plasmid pYPX88 under *AOX1* promoter for secretory expression in *P. pastoris* GS115. The enzyme activity from the culture supernatant of the clone rxynUMB#6 reached 110 U mL^{-1}, and the protein concentration was 0.042 mg mL^{-1}. The purified protein had a molecular mass of 24 kDa, an optimal pH of 4.3 and optimal temperature of 50 °C, with K_m and V_{max} of 1.8 mg mL^{-1} and 119 μmol min^{-1} mg^{-1}, respectively, over birchwood xylan.[83]

A xylanase B from *A. niger* IA-001 was cloned into pPICZαA and pPICZαmA for secretory expression in *P. pastoris* X-33. Plasmid pPICZαmA has a mutation in the α-factor signal sequence for increasing recombinant protein expression. After methanol induction in shake flasks, the clone from pPICZαmAxynB showed xylanase activity of 1280 U mL^{-1}, which is 12.5 times higher than that of the one from pPICZαAxynB and 19.4 times higher than in the native strain. After 114 h of induction in a fermenter, the enzyme activity reached 10.0 U mL^{-1}. The purified protein had a molecular mass of 24 kDa with an optimal pH of 5 and an optimal temperature of 50 °C. The specific activity, K_m and V_{max} over beechwood xylan were 1916 U mg^{-1}, 4.4 mg L^{-1}, and 1429 U mg L^{-1}, respectively.[84]

The β-xylosidase gene from the fungus *Paecilomyces thermophila* was cloned into plasmid pPICZαA and functionally expressed in *P. pastoris* KM71 with a C-terminal His_6 tag under *AOX1* promoter. After three days of induction, the protein expression reached 0.22 mg L^{-1}. Purified protein had a molecular mass of 52.3 kDa based on protein gel electrophoresis, which is 1.3 times higher than the predicted 39.31 kDa based on its composition, but there are no potential N- or O-glycosylation sites predicted from its amino acid sequence. The increase in molecular mass was presumed to be caused by some unpredictable post-translational modifications based on MS analysis of the recombinant protein. The enzyme was most active at 60 °C and pH 7. It not only showed a β-xylosidase activity with K_m and V_{max} of 8 mM and 54 μmol min^{-1} mg^{-1}, respectively, for the hydrolysis of *p*-nitrophenyl-β-D-xylopyranoside but also showed arabinofuranosidase activity of 6.2 U mg^{-1} on *p*-nitrophenyl-β-D-arabinofuranoside.[85]

Acetyl xylan esterase (AXE) from basidiomycete *Coprinopsis cinerea* Okayama 7 (#130) was cloned into plasmid pPICZαA and functionally expressed in *P. pastoris* KM71with a C-terminal His_6 tag under the *AOX1* promoter and secreted into the medium at 1.5 mg L^{-1}. The molecular mass of pure protein was 65.5 kDa, which is higher than the calculated molecular mass of 40 kDa due to N-glycosylation. The enzyme was most active at 40 °C and pH 8. It showed not only acetyl esterase activity with a K_m of 4.3 mM and a V_{max} of 2.15 U mg L^{-1} for hydrolysis of 4-nitrophenyl acetate, but also a butyl esterase activity for the hydrolysis of 4-nitrophenyl butyrate with a K_m of 0.11 mM and V_{max} of 0.78 U mg L^{-1}. The presence of two additional amino acid residues at its native N terminus was found to help stabilize the enzyme against protease cleavage in culture supernatant without affecting its activity.[86]

Codon-optimized endoglucanase II from *T. reesei* was cloned into pPinkα-HC for secretory expression in *P. pastoris*. After 72 h of fermentation, the enzyme activity reached 2358 U mL^{-1}. The purified enzyme had a molecular mass of 44 kDa and showed optimal activity at pH 4.8 and 75 °C. The enzyme had a K_m of 2.83 g L^{-1}, k_{cat} of 2.87 s^{-1} and a specific activity of 2620.9 U mg^{-1} on low viscosity CMC.[87]

Thermoascus aurantiacus TaCel5A and *A. fumigatus* AfCel12A were cloned for secretory expression with *AOXI* and *GAP* promoters of plasmid pPink. All four recombinant plasmids were transformed into double protease knock-out *Pichia* pink strain 4. Initial fermentation experiments showed that the levels of AfCel12A expression under the *GAP* promoter were similar or higher than those under the *AOX1* promoter, whereas the levels of TaCel5Acellulase were somewhat lower. Under optimized conditions in a bioreactor, the recombinant *P. pastoris* strains utilizing the *GAP* promoter produced 3–5 g L^{-1} of total secreted proteins, with CMCase activities equivalent to 1200 nkat mL^{-1} for AfCel12A and 170 nkat mL^{-1} for TaCel5A.[88]

Endoglucanase (*EglA*) from *A. niger* ATCC 10574 was cloned into plasmid pPICZαC and transformed into *P. pastoris* X-33. After three days of induction, the recombinant protein expression reached 40 mg L^{-1}. The purified protein had a molecular mass of ≈30 kDa and showed an optimal activity at pH 4 and 50 °C. This enzyme showed high affinity towards β-glucan and CMC with a specific activities of 63.83 and 9.47 U mg^{-1}, respectively.[89]

Thermostable β-glucosidase from *A. niger* Z5 was cloned into plasmid pPICZαA and transformed into *P. pastoris* X-33. After three days of induction, the recombinant protein expression reached 0.86 mg mL^{-1} with a highest β-glucosidase activity of 4.95 U mL^{-1}. The purified protein had a molecular mass of 130 kDa and showed optimal activity at pH 6 and 60 °C. K_m, V_{max}, k_{cat} on pNPG were 131.40 μmol min^{-1} mg^{-1}, 1.76 mM and 284.70 s^{-1}, respectively. The properties and kinetic parameters of both native and recombinant enzymes showed no significant differences.[90]

T. reesei cellobiohydrolase 2 (*Tr*CBH2) and β-mannase (*Tr*β-Man) as well as xylanase A (*Tl*XynA) from *T. lanuginosus* genes were codon-optimized and cloned under synthetic *AOX1* promoter variants. The recombinant plasmids were transformed into *P. pastoris* CBS 7435 Mut^S. The best screened clones were used for fed batch fermentation. After 91.5 h of induction, *Tr*CBH2, *Tr*β-Man and *Tl*XynA reached 6.55, 1.14 and 1.2 g L^{-1}, respectively.[91]

Type C phospholipid-specific lipase is generally applied to degum vegetable oils. BD16449 (E.C. 3.1.4.3) is a microbial phospholipid lipase obtained from *Bacillus cereus*. The enzyme-coding gene was cloned into the *AOX1* locus of the protease-deficient *P. pastoris* strain SMD1168 by homologous site-specific recombination. Individual clones were screened for maximal expression of the enzyme, and the strain was designated DVSA-PLC-004. Fed batch fermentation in a 500 L fermenter was conducted according to the cGMP standards for foods. The lyophilized enzyme with zero microbial count had an activity of 315 U PLC mg^{-1} dry weight with a molecular mass of 34 kDa, optimal pH of 7.5 and optimal temperature of 60 °C. The purified enzyme was tested for its toxicity and found to meet the food-grade standards for application in degumming edible vegetable oils.[92]

14. Conclusion and Perspectives

A wide spectrum of important enzymes and proteins has been successfully expressed in *Pichia pastoris*. Some of them have been applied in the chemical industries for producing chemicals. Some therapeutic and biopharmaceutical proteins have obtained approval for human use from the US Food and Drug Administration (FDA).

Although significant progress has been made in improving recombinant protein expression, the expression level for most proteins is still too low to meet the requirements for industrial applications. Whole-cell mutagenesis might be an alternative way of improving recombinant protein expression but has not yet received much attention. It has been reported that random mutagenesis with ethyl methyl sulfate developed a strain, ECCR72, that can simultaneously metabolize glucose and methanol and grow to a high cell densities in fermenters, secreting a large amount of recombinant proteins.[93] Techniques such as error-prone PCR of the whole genome with random primers[94] and random mutagenesis of whole cells by various physical mutation methods, such as γ-ray and plasma radiation,[95] are recommended for obtaining wide array of mutant strains for various applications.

P. pastoris has been thought to be a good platform strain for expressing hydrolytic enzymes for use in biorefining due to its post-translational modification function. However, no significant progress has been made in using this yeast to produce enzymes for lignocellulose biorefining. Currently, their expression in *P. pastoris* is at the milligram per liter level, which needs to be raised to grams per liter for industrial applications. For lignocellulose biorefining, the most important prerequisite is the depolymerization of cellulose and hemicellulose polysaccharides, which requires the synergistic action of a group of cellulases and hemicellulases. Successful coexpression of these hydrolytic enzymes and optimization of their ratios is very challenging, but worth investigating.

Acknowledgements

*The authors acknowledge the Science and Engineering Research Council (SERC) of the Agency for Science, Technology and Research (A*STAR) of Singapore (SERC grant number: 1326004117).*

Conflict of Interest

The authors declare no conflict of interest.

Keywords: biocatalysis · biorefinery · metabolic engineering · *Pichia pastoris* · protein expression

[1] S. N. Cohen, A. C. Y. Chang, H. Leslie, *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2110 – 2114.
[2] S. N. Cohen, A. C. Chang, H. W. Boyer, R. B. Helling, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3240 – 3244.
[3] a) S. R. Aggarwal, *Nat. Biotechnol.* **2014**, *32*, 32 – 39; b) www.marketsandmarkets.com/Market-Reports/protein-expression-market-180323924.html.
[4] V. Juturu, J. C. Wu, *Biotechnol. Adv.* **2012**, *30*, 1219 – 1227.
[5] Z. Liu, K. E. Tyo, J. L. Martínez, D. J. Petranovic, *BioEngineering* **2012**, *109*, 1259 – 1268.
[6] E. H. Wegner (Phillips Petroleum Company), US 4617274 A, **1986**.
[7] J. L. Cereghino, J. M. Cregg, *FEMS. Microbiol. Rev.* **2000**, *24*, 45 – 66.
[8] M. Hasslacher, M. Schall, M. Hayn, R. Bona, K. Rumbold, J. Lückl, H. Griengl, S. D. Kohlwein, H. Schwab, *Protein Expression Purif.* **1997**, *11*, 61 – 71.
[9] S. Wildt, T. U. Gerngross, *Nat. Rev. Microbiol.* **2005**, *3*, 119 – 128.
[10] B. Laukens, C. De Wachter, N. Callewaert, *Methods Mol. Biol.* **2015**, *1321*, 103 – 122.
[11] J. M. Cregg, K. J. Barringer, A. Y. Hessler, K. R. Madden, *Mol. Cell. Biol.* **1985**, *5*, 3376 – 3385.
[12] D. Mattanovich, A. Graf, J. Stadlmann, M. Dragosits, A. Redl, M. Maurer, M. Kleinheinz, M. Sauer, F. Altmann, B. Gasser, *Microb. Cell Fact.* **2009**, *8*, 29.
[13] B. K. S. Chung, S. Selvarasu, A. Camattari, J. Ryu, H. Lee, J. Ahn, H. Lee, D.-Y. Lee, *Microb. Cell Fact.* **2010**, *9*, 50.
[14] T. Vogl, C. Ruth, J. Pitzer, T. Kickenweiz, A. Glieder, *ACS Synth. Biol.* **2014**, *3*, 188 – 191.
[15] F. S. Hartner, A. Glieder, *Microb. Cell Fact.* **2006**, *5*, 39.
[16] F. W. Krainer, C. Dietzsch, T. Hajek, C. Herwig, O. Spadiut, A. Glieder, *Microb. Cell Fact.* **2012**, *11*, 22.
[17] J. F. Tschopp, P. F. Brust, J. M. Cregg, C. A. Stillman, T. R. Gingeras, *Nucleic Acids Res.* **1987**, *15*, 3859 – 3876.
[18] J. M. Cregg, K. R. Madden, K. J. Barringer, G. P. Thill, C. A. Stillman, *Mol. Cell Biol.* **1989**, *9*, 1316 – 1323.
[19] J. M. Cregg (Phillips Petroleum Company), US 5032516 A, **1991**.
[20] M. Inan, M. M. Meagher, *J. Biosci. Bioeng.* **2001**, *92*, 585 – 589.
[21] H. R. Waterham, M. E. Digan, P. J. Koutz, S. V. Lair, J. M. Cregg, *Gene* **1997**, *186*, 37 – 44.
[22] I. B. Sears, J. O'Connor, O. W. Rossanese, B. S. Glick, *Yeast* **1998**, *14*, 783 – 790.
[23] S. Shen, G. Sulter, T. W. Jeffries, J. M. Cregg, *Gene* **1998**, *216*, 93 – 102.
[24] J. M. Cregg (Research Corporation Technologies, Inc.), US 6730499 B1, **2004**.
[25] J. M. Cregg (Research Corporation Technologies, Inc.), US 7745200 B2, **2010**.
[26] J. Menendez, I. Valdes, N. Cabrera, *Yeast* **2003**, *20*, 1097 – 1108.
[27] J. Menendez Diaz, I. Valdes Prado, N. Cabrera Leon, US 20040106783 A1, **2004**.
[28] H. Liu, X. Tan, K. A. Russell, M. Veenhuis, J. M. Cregg, *J. Biol. Chem.* **1995**, *270*, 10940 – 10951.
[29] H. W. Lee, J. O. Ahn, J. K. Jung, E. S. Choi, C. S. Kim, E. G. Lee, J. Y. Hong, H. W. Lee, M. S. Park (Korea Research Institute of Bioscience and Biotechnology), US 8080389 B2, **2011**.
[30] G. Stadlmayr, A. Mecklenbrauker, M. Rothmuller, M. Maurer, M. Sauer, D. Mattanovich, B. Gasser, *J. Biotechnol.* **2010**, *150*, 519 – 529.
[31] T. Vogl, L. Sturmberger, T. Kickenweiz, R. Wasmayer, C. Schmid, A. M. Hatzl, M. A. Gerstmann, J. Pitzer, M. Wagner, G. G. Thallinger, M. Geier, A. Glieder, *ACS Synth. Biol.* **2016**, *5*, 172 – 186.
[32] G. P. Lin Cereghino, J. Lin Cereghino, A. J. Sunga, M. A. Johnson, M. Lim, M. A. Gleeson, J. M. Cregg, *Gene* **2001**, *263*, 159 – 169.
[33] L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner, A. Glieder, *PLoS One* **2012**, *7*, e39720.
[34] H. Marx, M. Sauer, D. Resina, M. Vai, D. Porro, F. Valero, P. Ferrer, D. Mattanovich, *FEMS Microbiol. Lett.* **2006**, *264*, 40 – 47.
[35] A. Küberl, J. Schneider, G. G. Thallinger, I. Anderl, D. Wibberg, T. Hajek, S. Jaenicke, K. Brinkrolf, A. Goesmann, R. Szczepanowski, A. Pühler, H. Schwab, A. Glieder, H. Pichler, *J. Biotechnol.* **2011**, *154*, 312 – 320.
[36] R. Weis, R. Luiten, W. Skranc, H. Schwab, M. Wubbolts, A. Glieder, *FEMS. Yeast. Res.* **2004**, *5*, 179 – 189.
[37] J. H. Nett, W. J. Cook, M. T. Chen, R. C. Davidson, P. Bobrowicz, W. Kett, E. Brevnova, T. I. Potgieter, M. T. Mellon, B. Prinz, B. K. Choi, D. Zha, I. Burnina, J. T. Bukowski, M. Du, S. Wildt, S. R. Hamilton, *PLoS One* **2013**, *8*, e68325.
[38] M. Bollok, D. Resina, F. Valero, P. Ferrer, *Recent Pat. Biotechnol.* **2009**, *3*, 192 – 201.
[39] N. Deshpande, M. R. Wilkins, N. Packer, H. Nevalainen, *Glycobiology* **2008**, *18*, 626 – 637.

[40] D. Mattanovich, N. Callewaert, P. Rouzé, Y. C. Lin, A. Graf, A. Redl, P. Tiels, B. Gasser, K. DeSchutter, *Microb. Cell Fact.* **2009**, *8*, 53.
[41] K. De Schutter, Y. C. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouzé, Y. Van de Peer, N. Callewaert, *Nat. Biotechnol.* **2009**, *27*, 561 – 566.
[42] T. Wriessnegger, P. Augustin, M. Engleder, E. Leitner, M. Müller, I. Kaluzna, M. Schürmann, D. Mink, G. Zellnig, H. Schwab, H. Pichler, *Metab. Eng.* **2014**, *24*, 18 – 29.
[43] X. B. Liu, M. Liu, X. Y. Tao, Z. X. Zhang, F. Q. Wang, D. Z. Wei, *J. Biotechnol.* **2015**, *216*, 47 – 55.
[44] a) www.marketsandmarkets.com/Market-Reports/carotenoid-market-158421566.html; b) P. C. Lee, A. Z. R. Momen, B. N. Mijts, C. Schmidt-Dannert, *Chem. Biol.* **2003**, *10*, 453 – 462.
[45] A. Bhataya, C. Schmidt-Dannert, P. C. Lee, *Process Biochem.* **2009**, *44*, 1095 – 1102.
[46] J. M. Araya-Garay, L. Feijoo-Siota, F. Rosa-dos-Santos, P. Veiga-Crespo, T. G. Villa, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 2483 – 2492.
[47] J. M. Araya-Garay, J. M. Ageitos, J. A. Vallejo, P. Veiga-Crespo, A. Sánchez-Pérez, T. G. Villa, *AMB Express* **2012**, *2*, 24.
[48] E. Jeong, W. Y. Shim, J. H. Kim, *J. Biotechnol.* **2014**, *185*, 28 – 36.
[49] D. Meesapyodsuk, Y. Chen, S.-H. Ng, J. Chen, X. Qiu, *J. Lipid Res.* **2015**, *56*, 2102 – 2109.
[50] M. Ahmad, M. Hirz, H. Pichler, H. Schwab, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5301 – 5317.
[51] http://gcua.schoedl.de/.
[52] E. Angov, *Biotechnol. J.* **2011**, *6*, 650 – 659.
[53] J. R. Wang, Y. Y. Li, D. N. Liu, J. S. Liu, P. Li, L. Z. Chen, S. D. Xu, *Biomed. Res. Int.* **2015**, 248680.
[54] H. Jia, G. Fan, Q. Yan, Y. Liu, Y. Yan, Z. Jiang, *J. Mol. Catal. B* **2012**, *78*, 72 – 77.
[55] C. Xu, D. T. Ng, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 742 – 752.
[56] M. Han, X. Yu, *Bioengineered* **2015**, *6*, 115 – 118.
[57] W. Wei, L. Chen, G. Zou, Q. Wang, X. Yan, J. Zhang, C. Wang, Z. Zhou, *Biotechnol. Bioeng.* **2013**, *110*, 3075 – 3084.
[58] M. Han, W. Wang, G. Jiang, X. Wang, X. Liu, H. Cao, Y. Tan, X. Yu, *Biotechnol. Lett.* **2014**, *36*, 2467 – 2471.
[59] X. Qin, J. Qian, G. Yao, S. Zhuang, J. Chu, *Appl. Environ. Microbiol.* **2011**, *77*, 3600 – 3608.
[60] B. A. Blount, T. Weenink, S. Vasylechko, T. Ellis, *PLoS One* **2012**, *7*, e33279.
[61] L. Berg, T. A. Strand, S. Valla, T. Brautaset, *BioMed Res. Int.* **2013**, 926985.
[62] F. Hartner, A. Glieder (Technische Universität Graz, VTU Holding GmbH), US 9012175 B2, **2008**.
[63] F. S. Hartner, C. Ruth, D. Langenegger, S. N. Johnson, P. Hyka, G. P. Lin-Cereghino, J. Lin-Cereghino, K. Kovar, J. M. Cregg, A. Glieder, *Nucleic Acids Res.* **2008**, *36*, e76.
[64] H. Ohi, M. Miura, S. Uno, M. Chuganji, R. Hiramatsu, T. Ohmura (The Green Cross Corporation), US 5683893, **1997**.
[65] H. Ohi, M. Miura, S. Uno, R. Hiramatsu, T. Ohmura (The Green Cross Corporation), US 5707827, **1998**.
[66] S. Yang, Y. Kuang, H. Li, Y. Liu, X. Hui, P. Li, Z. Jiang, Y. Zhou, Y. Wang, A. Xu, S. Li, P. Liu, D. Wu, *PLoS One* **2013**, *8*, e75347.
[67] M. D.-T. Chang, C. C. Sheu (Simpson Biotech Co., Ltd.), US 7785872 B2, **2010**.
[68] M. V. Moura, G. P. da Silva, A. C. Machado, F. A. Torres, D. M. Freire, R. V. Almeida, *PLoS One* **2015**, *10*, e0141454.
[69] J. Ou, Y. Cao, *J. Microbiol. Biotechnol.* **2014**, *24*, 1178 – 1188.
[70] J. De AmorimAraújo, T. C. Ferreira, M. R. Rubini, A. G. Duran, J. L. De Marco, L. M. de Moraes, F. A. Torres, *AMB Express.* **2015**, *5*, 84.
[71] N. Roongsawang, P. Promdonkoy, M. Wongwanichpokhin, W. Sornlake, A. Puseenam, L. Eurwilaichitr, S. Tanapongpipat, *FEMS Yeast Res.* **2010**, *10*, 909 – 916.
[72] X. Chen, Z. Huang, B. Zhou, M. Xu, CN 104046645 A, **2014**.
[73] X. Chen, B. Zhou, M. Xu, Z. Huang, G. Jia, Q. Jiayun, L. Guangmang, *Turk. J. Biol.* **2015**, *39*, 312 – 319.
[74] X. Xu, H. Wang, D. Li, Wanghai, CN 104152484 A, **2014**.
[75] B. Gasser, D. Mattanovich, M. Sauer, G. Stadlmayr (Polymun Scientific Immunbiologische Forschung GmbH), US 20100297738 A1, **2010**.
[76] S. Zhuang, J. Fu, C. Powell, J. Huang, Y. Xia, R. Yan, *SpringerPlus* **2015**, *4*, 467.
[77] J. Yan, X. Zheng, S. Li, *Bioresour. Technol.* **2014**, *151*, 43 – 48.
[78] S. Abad, J. Nahalka, G. Bergler, S. A. Arnold, R. Speight, I. Fotheringham, B. Nidetzky, A. Glieder, *Microb. Cell Fact.* **2010**, *9*, 24.
[79] N. I. Alias, N. M. Mahadi, A. M. A. Murad, F. D. A. Bakar, A. Rabu, R. M. Illias, *African. J. Biotechnol.* **2011**, *35*, 6700 – 6710.
[80] J. He, B. Yu, K. Zhang, X. Ding, D. Chen, *BMC Biotechnol.* **2009**, *9*, 56.
[81] N. Kirikyali, J. Wood, I. F. Connerton, *AMB Express.* **2014**, *4*, 1 – 7.
[82] L. Zhao, J. Geng, Y. Guo, X. Liao, X. Liu, R. Wu, Z. Zheng, R. Zhang, *BMC Biotechnol.* **2015**, *15*, 18.
[83] H. Han, S. You, B. Zhu, X. Fu, B. Sun, J. Qiu, C. Yu, L. Chen, R. Peng, Q. Yao, *Biotechnol. Lett.* **2015**, *37*, 697 – 703.
[84] W. Fang, H. Gao, Y. Cao, A. Shan, *J. Basic Microbiol.* **2014**, *54*, 190 – 199.
[85] V. Juturu, J. C. Wu, *World J. Microbiol. Biotechnol.* **2013**, *29*, 249 – 255.
[86] V. Juturu, C. Aust, J. C. Wu, *World J. Microbiol. Biotechnol.* **2013**, *29*, 597 – 605.
[87] A. Akbarzadeh, S. O. RanaeiSiadat, M. Motallebi, M. R. Zamani, M. BarshanTashnizi, S. Moshtaghi, *Appl. Biochem. Biotechnol.* **2014**, *172*, 2253 – 2265.
[88] A. Várnai, C. Tang, O. Bengtsson, A. Atterton, G. Mathiesen, V. G. H. Eijsink, *Microb. Cell Fact.* **2014**, *13*, 57.
[89] D. H. X. Quay, F. D. A. Bakar, A. Rabu, M. Said, R. M. Illias, N. M. Mahadi, O. Hassan, A. M. A. Murad, *African J. Biotechnol.* **2011**, *10*, 2101 – 2111.
[90] D. Liu, R. Zhang, X. Yang, Z. Zhang, S. Song, Y. Miao, Q. Shen, *Microb. Cell Fact.* **2012**, *11*, 25.
[91] A. Mellitzer, R. Weis, A. Glieder, K. Flicker, *Microb. Cell Fact.* **2012**, *11*, 61.
[92] a) V. Ciofalo, N. Barton, J. Kreps, I. Coats, D. Shanahan, *Regul. Toxicol. Pharmacol.* **2006**, *45*, 1 – 8; b) www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/NoticeInventory/ucm153734.htm.
[93] K. Okabayashi, T. Ohmura, K. Yokoyama, H. Kawabe (The Green Cross Corporation), US 5643792 A, **1997**.
[94] L. Ye, H. Zhao, Z. Li, J. C. Wu, *Bioresour. Technol.* **2013**, *135*, 459 – 463.
[95] X. Zhang, X. F. Zhang, H. P. Li, L. Y. Wang, C. Zhang, X. H. Xing, C. Y. Bao, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5387 – 5396.
[96] T. U. Gerngross (Glycofi, Inc.), US 7029872 B2, **2006**.
[97] T. U. Gerngross, S. Wildt, B.-K. Choi, J. H. Nett, P. Bobrowicz, S. R. Hamilton, R. C. Davidson (Glycofi, Inc.), US 7449308 B2, **2008**.
[98] R. Contreras, N. L. M. Callewaert, S. C. J. Geysens (Flanders Interuniversity Institute For Biotechnology), US 6803225 B2, **2004**.
[99] R. Contreras, N. L. M. Callewaert, S. C. J. Geysens, V. Kaigorodov, W. Vervecken (Flanders Interuniversity Institute For Biotechnology), US 7252933 B2, **2007**.
[100] T. U. Gerngross (Glycofi, Inc.), US 7326681 B2, **2008**.
[101] R. Contreras, N. L. M. Callewaert, W. Vervecken, V. Kaigorodov (Research Corporation Technologies, Inc.), US 7507573 B2, **2009**.
[102] S. R. Hamilton, (Glycofi, Inc.), US 2006286637 A1, **2006**.
[103] M. Miura, M. Hirose, T. Miwa, H. Irie, S. Kuwae, K. Miyano, W. Otani, H. Ohi (Mitsubishi Pharma Corporation), US 7198921 B2, **2007**.
[104] M. Miura, M. Hirose, T. Miwa, H. Irie, S. Kuwae, K. Miyano, W. Otani, H. Ohi (Mitsubishi Tanabe Pharma Corporation), US 7488591 B2, **2009**.
[105] P. Bobrowicz, W. J. Cook, W. Kett (Merck Sharp & Dohme Corp.), US 8795984 B2, **2014**.

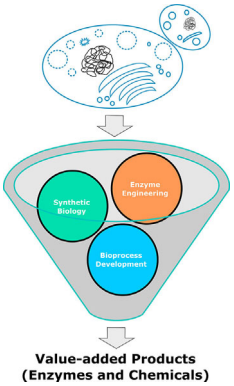
Manuscript received: July 28, 2017
Version of record online: ■■ ■■, 0000

REVIEWS

*V. Juturu, J. C. Wu**

■■ – ■■

Heterologous Protein Expression in *Pichia pastoris*: Latest Research Progress and Applications



A powerful tool kit for recombinant protein expression has been developed by using wild-type *P. pastoris*. A diverse spectrum of proteins and enzymes has been successfully produced by extracellular expression. Cytoplasmic expression and cell-surface display created whole-cell biocatalysts for biotransformation reactions, while metabolically engineered strains were used as microbial cell factories to produce value-added chemicals.