



# Recent Advances in Synthetic Biology Applications of *Pichia* Species

Wan Sun, Yimeng Zuo, Zhanyi Yao, Jucan Gao, Zengyi Shao,  
and Jiazhang Lian

## Abstract

The rapid development of metabolic engineering and synthetic biology has recently attracted great attention to exploring many non-model microbial species. *Pichia pastoris* and *Pichia kudriavzevii* are the two representative *Pichia* species, which possess uniquely desired biochemical, metabolic, and physiological features favoring large-scale industrial production. This review begins with illustrating synthetic biology parts and tools that have been developed for manipulating these two species, followed by summarizing their applications in biomanufacturing as hosts to produce recombinant proteins, bulk

---

W. Sun, Y. Zuo, Z. Yao and J. Gao contributed equally to this work.

---

W. Sun · Z. Shao (✉)

Interdepartmental Microbiology Program, Iowa State University, Ames, IA, USA

e-mail: [zyshao@iastate.edu](mailto:zyshao@iastate.edu)

W. Sun · Z. Yao · Z. Shao

NSF Engineering Research Center for Biorenewable Chemicals, Iowa State University, Ames, IA, USA

W. Sun · Z. Shao

The Ames Laboratory, Ames, IA, USA

Y. Zuo · J. Gao · J. Lian (✉)

College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, Zhejiang, China

e-mail: [jzlian@zju.edu.cn](mailto:jzlian@zju.edu.cn)

Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou Zhejiang, China

Z. Yao · Z. Shao

Department of Chemical and Biological Engineering, Iowa State University, Ames, IA, USA

Z. Shao

4140 Biorenewables Research Laboratory, Iowa State University, Ames, IA, USA

chemicals, and natural products, and concludes on the challenges and potential strategies for advancing them for broader biotechnological applications.

---

## 1 Introduction

With an increasing demand for fuels and petrochemicals, in addition to the growing food shortage and global warming, researchers have been working on engineering microbes to generate desirable products (Choi et al. 2015; Curran and Alper 2012). Model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* have been studied extensively and engineered as cell factories for the production of value-added chemicals and proteins (Pontrelli et al. 2018; Lian et al. 2018). With the rapid development of next-generation sequencing and synthetic biology tools, non-conventional organisms with unique traits (e.g., thermotolerance (Varela et al. 2017), stress tolerance (Abdel-Mawgoud et al. 2018), special feedstock utilization (Agbogbo and Coward-Kelly 2008; Gong et al. 2015; Yaegashi et al. 2017; Yaguchi et al. 2017), and high protein secretion capacity (Li et al. 2007)) have been explored as well-suited hosts for industrial processes.

The most renowned *Pichia* species is *Pichia pastoris*, a methylotrophic yeast that can utilize methanol as the sole carbon source. It can grow to a very high cell density (>100 g/L dry cell weight) (Wang et al. 2012). The medium for growing *P. pastoris* is simple and inexpensive (containing only methanol or glycerol, biotin, salts, and trace elements), making it ideal for large-scale industrial production (Cereghino et al. 2002). In addition, *P. pastoris* can express heterologous proteins intracellularly or extracellularly at high levels whilst minimizing secretion of endogenous proteins. It possesses the machinery required for proper post-translational modifications (e.g., glycosylation, disulfide bond formation, and proteolytic processing) (Cereghino and Cregg 2000). As a “generally regarded as safe” (GRAS) yeast, it has been engineered to produce industrial enzymes and chemicals (Zhu 2019), therapeutic agents such as vaccines (Wang et al. 2016) and drugs (Yu et al. 2007), and protein-based polymers (Werten et al. 2019). In recent years, it has also been engineered to produce a leghemoglobin protein (LegH) from soy, to create a meat-like flavor in plant-based meat products (Fraser et al. 2018). The well-annotated genome sequence of *P. pastoris* (Love et al. 2016; Sturmberger et al. 2016) and a few genome-scale metabolic models are also available for engineering purposes (Ye et al. 2017; Saitua et al. 2017; Torres 2019).

This review change it to summarizes the current synthetic biology parts and tools available for *P. pastoris*, including promoters, terminators, plasmids, genome-editing tools, and signal peptides, followed by a discussion on the engineering efforts carried out to improve protein secretion and various applications of this yeast. In addition, we have also highlighted *Pichia kudriavzevii*, another *Pichia* species whose acid tolerance is attractive in the production of value-added organic acids such as succinic acid (Xiao et al. 2014), D-lactic acid (Park et al. 2018), and itaconic acid (Sun 2020). This review concludes with future perspectives in establishing *Pichia* species for biotechnological applications.

## 2 Synthetic Biology Parts and Tools

### 2.1 Promoters and Terminators

The strength and tunability of promoters are crucial for efficient production (Porro et al. 2005). Strong promoters are usually preferred because of the usually low expression levels of heterologous genes, whereas tunable promoters are desirable when multiple heterologous genes are involved in complex pathways. Commonly used promoters include inducible and constitutive promoters. The former allows genes of interest to be switched on or off at different stages through induction or repression via transcription factors. In practice, applying an inducible promoter can increase the cell density of cultures first and then initiate heterologous protein production. Such a separation of biomass accumulation and protein production provides a great advantage if the accumulated intermediates or products are toxic to the cells (Cereghino and Cregg 2000; Ahmad et al. 2014; Macauley-Patrick et al. 2005). However, using an inducible promoter requires an extra step during cultivation (e.g., carbon source swapping or compound supplementation), which incurs an additional cost for large-scale industrial production. In contrast, strong and steady expression of genes of interest mediated by constitutive promoters contributes to decreasing the operational cost while enhancing the yield, which is a commonly adopted strategy if the constitutive expression does not negatively affect cell growth (Vogl and Glieder 2013).

Inducible promoters are usually identified from unique biochemical pathways, and constitutive promoters generally originate from housekeeping genes. The two most commonly used promoters in *P. pastoris* are the inducible  $P_{AOXI}$  and the constitutive  $P_{GAP}$  (Rajamanickam et al. 2017).  $P_{AOXI}$  is the promoter of the alcohol oxidase gene *AOXI*, which is induced by the inexpensive carbon source, methanol. When methanol is used as a carbon source, the methanol-induced alcohol oxidase expression can reach 30% of the total soluble protein content (Cregg et al. 1993). It is one of the most effective promoters that have been found for protein expression in *P. pastoris* (Cregg et al. 1989; Jahic et al. 2006; Koutz et al. 1989). Heterologous protein synthesis at levels of approximately 20 g/L was achieved two decades ago (Hasslacher et al. 1997; Werten et al. 1999). Extensive efforts have been made to elucidate the regulatory mechanisms of  $P_{AOXI}$ . The determination of its *cis*-acting regulatory sequence elements has allowed researchers to engineer the promoter by means of deletion and duplication of putative transcription factor-binding sites, yielding a library of variants with strengths ranging from 6 to 160% of the wild-type  $P_{AOXI}$  (Hartner 2008). A dozen transcription factors have been identified to be involved in the induction of  $P_{AOXI}$ . This work provided various  $P_{AOXI}$  variants with tunable activities by combining *cis*-acting elements with the basal promoter identified based on deletion analysis.

$P_{GAP}$  is the promoter of the glycolytic glyceraldehyde 3-phosphate dehydrogenase gene *GAP*. It remains constitutively expressed, although its activity varies when different carbon sources are used (Waterham et al. 1997). Similar to the engineering strategy applied to  $P_{GAP}$ , a library of  $P_{GAP}$  variants was constructed

using mutagenesis, in which the activity varied from 0.6% to 1960% of the wild-type  $P_{GAP}$  (Qin et al. 2011). Several transcription factors have been examined and suggested to play roles in the regulation of  $P_{GAP}$ . Most of the available *P. pastoris* promoters are summarized in two review articles, with the strengths benchmarked to those of  $P_{AOXI}$  and  $P_{GAP}$  (Vogl and Glieder 2013; Turkanoglu Ozcelik et al. 2019). These promoters have been much less extensively studied but are highly desired for controlling multi-gene pathways. Most of their *cis*-acting regulatory sequences and mechanisms remain unclear. Beyond the ones in the two lists, there are also a strong native promoter ( $P_{CATI}$ ) from the catalase gene in *P. pastoris* and a strong heterogeneous one ( $P_{MOX}$ ) originating from the methanol oxidase gene in *Hansenula polymorpha*. The former is induced by methanol with the  $P_{CATI}$  variant P4 even stronger than  $P_{AOXI}$  (Nong et al. 2020); the latter was completely inactivated in the presence of xylose and sorbitol but showed strong activities in the glucose, glycerol, and methanol feeds (Mombeni 2020).

In conjunction to a promoter, a terminator (tt) also plays a critical role in regulating the expression level, mainly by influencing the mRNA stability and the subsequent translation process (Shalgi et al. 2005). However, compared with promoter engineering, much less attention has been focused on terminators. In *S. cerevisiae*, hundreds of terminators have been examined and characterized, which have been demonstrated to regulate protein expression levels over a broad range, whereas in *P. pastoris*, only a few recent studies have investigated the impact of terminators.

In *P. pastoris*, both endogenous and heterogeneous terminators have been used for heterologous expression. The commonly used endogenous terminators are mostly from either the methanol utilization pathway (i.e.,  $AOXI_{tt}$ ) or housekeeping genes (i.e.,  $GAP_{tt}$ ), whereas heterogeneous terminators are from other yeasts such as *S. cerevisiae* (e.g.,  $CYC1_{tt}$  from cytochrome C isoform 1,  $PRM9_{tt}$  from pheromone-regulated membrane protein 9, and  $VPS13_{tt}$  from vacuolar protein sorting-associated protein 13), *H. polymorpha* (e.g.,  $MOX_{tt}$  from methanol oxidase), and *Kluyveromyces lactis* (e.g.,  $LAC4_{tt}$  from beta-galactosidase). To date, the impact of different terminators on heterologous expression has been primarily examined by evaluating the expression level under the control of  $P_{AOXI}$  and  $P_{GAP}$ .

Vogl et al. examined 20 terminators, of which 15 originated from the endogenous methanol assimilation pathway and the remaining five from *S. cerevisiae*. These terminators provided comparable expression levels of green fluorescent protein (GFP) under the control of  $P_{AOXI}$ , with the lowest active terminator still reaching 57% of the highest active terminator (Vogl et al. 2016). Prielhofer et al. focused more on the terminators from highly expressed endogenous genes, including many ribosomal terminators. Using  $CYC1_{tt}$  from *S. cerevisiae* as a reference, all ten terminators yielded similar GFP levels under the control of  $P_{GAP}$  (Prielhofer et al. 2017). Interestingly, in both studies, the heterogeneous terminators of genes of interest offered comparable and sometimes even higher activities than the endogenous ones, indicating that the terminators isolated from other yeasts could be effectively recognized in *P. pastoris*. Recently, Ito et al. created a catalog of 72 terminators, including 28 endogenous terminators, 41 heterogeneous terminators

from *S. cerevisiae*, and three strong synthetic terminators developed originally for *S. cerevisiae*. Under the control of  $P_{GAP}$ , these terminators resulted in a 17-fold degree of tunability in *P. pastoris* (Ito et al. 2020). In these studies,  $AOXI_{It}$  seemed to yield the highest activity, regardless of the promoter being used (Karbalaei et al. 2020; Weninger et al. 2016).

A recent study using *Candida antarctica* lipase B (CALB) as a reporter protein showed that the activities of terminators are closely associated with those of promoters (Ramakrishnan et al. 2020). Ten terminators from the endogenous methanol utilization pathway, glycolysis, tricarboxylic acid (TCA) cycle, and other housekeeping genes and five terminators from *S. cerevisiae* were compared. Their activities were estimated by evaluating the corresponding CALB activity under the control of  $P_{AOXI}$  and  $P_{GAP}$ . Compared to  $AOXI_{It}$ , three terminators led to lower lipase activities when paired with  $P_{AOXI}$  but higher activities when paired with  $P_{GAP}$ , which suggested that the performance of terminators is not insulated from promoter influences and may also be subjected to regulatory mechanisms as seen in promoter studies. However, the mechanism by which individual terminators enhance expression along with different promoters remains unclear in *P. pastoris*. In addition, the terminator of dihydroxyacetone synthase ( $DHAS_{It}$ ) provided a slightly higher CALB expression level than  $AOXI_{It}$  under the control of  $P_{AOXI}$ , but nearly threefold higher activity under the control of  $P_{GAP}$ . Therefore,  $DHAS_{It}$  can potentially serve as a strong terminator when seeking high heterologous expression in *P. pastoris*. In general, terminators play a critical role in protein expression, but more studies are needed to elucidate terminator-mediated regulatory mechanisms.

## 2.2 Episomal Plasmids and Integration Plasmids

Most of the protein expression and metabolic engineering tasks in *P. pastoris* were achieved through genome integration, many of which targeted the  $AOXI$  gene via homologous recombination (HR) (Cereghino and Cregg 2000). However, unlike *S. cerevisiae*, non-homologous end joining (NHEJ) is the dominant mechanism for repairing chromosomal double-stranded breaks (DSBs) in *P. pastoris*. Transformants with targeted integration have to be identified using a laborious screening process because of the uncontrollable random integration events, including large-scale relocation of an integration locus, off-target integration potentially affecting cell growth, and even co-integration of the DNA elements originating from the F-plasmid and the genome of the *E. coli* host used to prepare the shuttle plasmid (Schwarzshans et al. 2016).

An alternative strategy is to use replicative plasmids, which yield higher transformation efficiency and are easier to screen, although stability is sometimes an issue in this case (Lee et al. 2005). An autonomously replicating sequence (ARS) is a key element in replicative plasmids. The first *P. pastoris*-specific ARS, designated  $PARSI$ , was identified over 35 years ago and enabled the use of replicative plasmids with high transformation efficiency (Cregg et al. 1985). Over the past decade, other elements have been discovered that can serve as alternative ARSs

in *P. pastoris*, such as a 452 bp *panARS* identified from *K. lactis* and a 1442 bp mitochondrial DNA fragment from *P. pastoris* itself (Liachko and Dunham 2014; Schwarzans et al. 2017). Unfortunately, plasmids using these ARSs demonstrate poor stability during mitotic segregation, which is fatal for industrial applications. Recent studies suggest that this inherent instability is caused by the lack of a centromere (CEN), another genomic element that guides stable segregation of chromosomes, and therefore, can be used to improve plasmid stability during cell division (Cao et al. 2017a).

CENs are DNA sequences recognized by kinetochore complexes, which subsequently interact with spindle microtubules and enable equal partitioning of chromosomes to the two dividing cells during mitosis and meiosis. In *S. cerevisiae*, a 125 bp CEN and an ARS have been widely applied as a combination in all low-copy episomal plasmids. Like ARS, a CEN is also species-specific, and recent studies have identified four putative CENs corresponding to the four *P. pastoris* chromosomes (Sturmberger et al. 2016; Coughlan et al. 2016). In a more recent study, a new autonomously replicating plasmid was constructed, harboring an entire putative centromeric region from chromosome 2 (*Cen2*). The plasmid can be replicated and stably distributed in *P. pastoris*. Within this *Cen2*, a ~111 bp sequence was found to enable autonomous replication, which can serve as a new ARS (Nakamura, et al. 2018). Another study confirmed that the entire CENs from chromosomes 1 and 4 (*Cen1* and *Cen4*) could confer replicative stability to plasmids (Piva 2020) although *Cen4* did not support a high number of transformants in a separate study (Nakamura, et al. 2018). Although these new plasmids exhibit relatively high stability and have the potential to expedite cloning and high-throughput screening, harboring the entire *Cen* sequence may lead to other undesired outcomes. First, this kind of plasmid can only be maintained at low copy numbers due to their chromosome-like segregation mechanism, in contrast to those plasmids using *PARS1*. Moreover, the sizes of the plasmids will be much larger since all *Cen* sequences are above 6 kb in length, which does not benefit transformation, especially when large pathways are cloned. Lastly, having the entire centromeric sequence may cause unexpected genomic integration or DNA exchange with chromosomes, which increases the difficulty of screening.

Therefore, genome integration appears to be an alternative for heterologous gene and pathway expression before the functioning mechanisms of ARS and CEN are clearly elucidated and the sequences are optimized in *P. pastoris*. Genome integration is generally achieved via a single crossover or double crossover. A single crossover requires that the circular vector contains a sequence identical to the target locus in the *P. pastoris* genome. After transformation, the linearized vector, including genes of interest, a selectable marker, and backbone, will be inserted into the target site, leaving two copies of the target site flanking the inserted vector. The *P<sub>AOX1</sub>* region and the auxotrophic gene *HIS4* have been widely selected as the target sites for single crossover-mediated integration, with an efficiency of 50–80% (Cereghino and Cregg 2000). However, this kind of integration will introduce elements beyond the expression cassette, such as the elements responsible for the replication of the shuttle vector in *E. coli*. In addition, a second

single crossover may occur again on the genome between two identical target sequences, especially when the selection pressure is removed, which results in a loss of integrated expression cassette. To achieve a double crossover, a selection marker-containing expression cassette flanked by two homologous arms to the target site is usually transformed, resulting in the direct replacement of the genomic sequence located between the two homologous sequences by the expression cassette. Flanking the desired genes of interest and a selectable marker with the 5' and 3' *AOX1* sequences leads to the disruption of *AOX1*, thereby changing the phenotypic substrate utilization of *P. pastoris*. Lastly, the selection markers commonly used for screening include auxotrophic genes such as *HIS4*, *URA3*, and *ADE1*, as well as the genes encoding resistance to zeocin and G418 sulfate (Daly and Hearn 2005; Papakonstantinou et al. 2009).

### 2.3 Genome Editing and Integration Loci

In addition to integration vectors or fragments, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology has become a revolutionary tool for achieving genomic integration in *P. pastoris* because Cas9, in principle can target any accessible locus containing a 2–6 bp protospacer adjacent motif with the help of a guide RNA (gRNA) molecule. However, it has been reported that a Cas9/gRNA complex may lead to toxic effects due to off-targeting and therefore, the expression level of Cas9 needs to be at an appropriate level to yield a desired efficiency (Weninger et al. 2016). For example, using a weaker promoter for Cas9 expression can lead to higher transformation efficiencies and growth rates (Prielhofer et al. 2017). Cas9 cleaves a genome and introduces a DSB that can be repaired by the NHEJ machinery when donor DNA is not provided, resulting in insertion and deletion (indel) mutations at the target locus. In *P. pastoris*, the inherently strong NHEJ activity enables highly efficient multi-locus disruption or deletion with co-transformation of different gRNAs. For example, Weninger et al. successfully mutated *GUT1* and *AOX1* simultaneously by expressing two gRNAs on a CRISPR/Cas9 plasmid. By testing different combinations of gRNAs, the highest double-editing efficiency of 69% was obtained (Weninger et al. 2016). However, the active NHEJ machinery is a hurdle for achieving precise locus-specific integrations via HR. When a donor DNA is provided, HR must compete with the predominant NHEJ to repair the DSB, and under the selection pressure, the marker can be randomly integrated via NHEJ. The HR activity in the wild-type *P. pastoris* is naturally much lower than that of NHEJ, yielding many false positives (Li et al. 2007). Therefore, it is much more difficult to achieve precise integration in *P. pastoris* than in other yeast species with HR dominance, such as *S. cerevisiae*. This limitation can be overcome by deleting the *KU70* and *KU80* genes encoding the two proteins comprising a heterodimer threading onto the broken DNA ends. In a recent study, it was reported that upon knock-out of *KU70*,

NHEJ was repressed to a great extent, thereby heightening the role of HR in repairing DSB repair and raising the target integration efficiency to approximately 100% (Weninger et al. 2016, 2018).

In addition to repressing NHEJ, future studies should focus on enhancing HR performance to increase editing efficiency when using CRISPR/Cas9 in *P. pastoris*. For example, the overexpression of RAD family recombinases can be considered. In *S. cerevisiae*, overexpression of Rad51 has been shown to elevate the integration correctness and overexpression of an engineered Rad51 variant, which has a higher affinity to recombinase Rad54, thus significantly increasing the targeting efficiency in *S. cerevisiae* (Liu et al. 2004). Furthermore, expressing Rad52 from *S. cerevisiae* in *Yarrowia lipolytica* has been shown to increase the targeting efficiency from 15 to 95% (Ji 2020). When *KU70* was disrupted, certain chemicals, such as hydroxyurea, were applied to synchronize *Y. lipolytica* cells to the S-phase of the cell cycle, which is when HR has the highest activity (Jang 2018). Similar strategies can be used to enhance the HR performance in *P. pastoris*.

Theoretically, any locus on a genome can serve as a potential target site for integration, except for those related to essential genes. However, several additional features directly affect the integration efficiency. First, a higher accessibility renders a higher chance for the Cas9/gRNA complex, and later, the donor DNA to form a complex with the genome during the cutting and repair process, which is especially important for the integration of large pathways. *AOX1*, *DAS1*, *DAS2*, and *GUT1* have been widely targeted owing to their easy accessibility (Pena et al. 2018). In addition, the nano-environment surrounding the integration locus is key to determining the expression level and dynamics of the integrated gene. Although a high expression level does not necessarily lead to high production, it is usually preferred by the rate-limiting step for the synthesis of a product through a multi-step pathway.

Increasing the copy number of a gene of interest has been found to be an efficient way to increase heterologous protein production. Accordingly, ribosomal RNA (rRNA)-encoding loci, namely ribosomal DNA (rDNA), have been populated because of their highly repeated sequences, where several loci can be targeted simultaneously using a single plasmid containing one gRNA design (Marx et al. 2009). In *P. pastoris*, each repeat consists of 25S, 5.8S, and 18S rRNA genes arranged identically in a head-to-tail tandem array and a non-transcribed spacer (NTS) is located between two rDNA repeats. The sequences and locations of these rDNA repeats on the chromosomes in many yeasts have been specified for decades. For example, Wang et al. successfully integrated ten copies of the resveratrol biosynthetic pathway consisting of genes from *Herpetosiphon aurantiacus*, *Arabidopsis thaliana*, and *Vitis vinifera*, onto the NTS regions of *O. polymorpha* simultaneously via CRISPR/Cas9, without using a selection marker (Wang et al. 2018). In *Y. lipolytica*, Luu et al. obtained an integrant with eight copies of the genes encoding capsid proteins originating from red-spotted grouper nervous necrosis virus at the 26S rRNA loci through HR (Luu et al. 2017). In *P. pastoris*, high copy-number integration of human serum albumin and human superoxide dismutase was achieved by targeting the NTS region of the rDNA locus



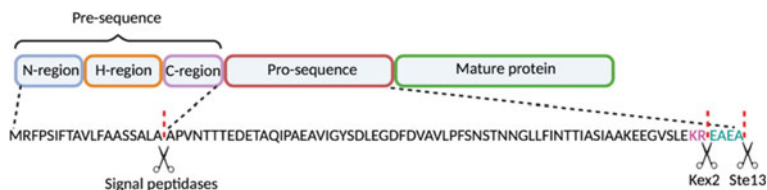
and repeated selection on increasing the concentration of zeocine™ (Marx et al. 2009). In addition to expressing proteins as products, this strategy has also been applied to enhance the production of D-lactic acid in *P. pastoris* by integrating the D-lactate dehydrogenase gene (*D-LDH*) into the rDNA locus, followed by copy number amplification enabled by gradually increased antibiotic concentration (Yamada et al. 2019).

## 2.4 Commonly Used Strains

Strains Y-11430 (CBS-7435), GS115, and X-33 are the most commonly used *P. pastoris* strains. Y-11430 is a wild-type strain that was originally isolated from California Black oak and deposited in the United States Department of Agriculture Culture Collection (USDA-NRRL). It is renowned because of its robust growth rate and high activity in the methanol utilization pathway. GS115 is a histidine auxotroph obtained by mutagenizing Y-11430 with nitrosoguanidine. It has become popular for its ease of integration and screening while using *HIS4* as a selectable marker (Cregg et al. 1985; Schutter et al. 2009). X-33 is the revertant of GS115 created by the complementation of *HIS4*, and zeocin or blasticidin can still be used to select X-33 transformants carrying the antibiotic resistance gene (Higgins et al. 1998). X-33 shares a few mutations with GS115 in genes encoding cell wall biosynthesis, which enhance the secretion of membrane-associated proteins and result in a higher transformation efficiency than other species with thick cell walls, making X-33 a popular commercial strain for heterologous protein expression (Brady et al. 2020).

## 2.5 Signal Peptides for Mediating Protein Secretion

Protein secretion is a complex process that involves multiple steps to generate a mature active protein. A signal peptide is a short sequence that is usually located at the N-terminus of a nascent polypeptide and directs a protein into the secretory pathway (Owji et al. 2018). The most commonly used signal peptide for recombinant protein expression in *P. pastoris* is  $\alpha$ -mating factor ( $\alpha$ -MF) prepropeptide originating from *S. cerevisiae* (Fig. 1). The  $\alpha$ -MF prepropeptide consists of a 19-amino acid presignal sequence and a 66-amino acid pro-sequence. The pre-peptide typically has three domains: a positively charged N-terminal region, a central hydrophobic region, and a polar C-terminal region. There are three main steps in processing an  $\alpha$ -MF secretion signal. First, the presignal is cleaved by signal peptidases in the endoplasmic reticulum (ER). Kex2 endopeptidase in the Golgi then cleaves the pro-leader sequence at the dibasic KR site, and finally, Ste13 protein removes the EA repeats (Julius et al. 1984; Brake et al. 1984).  $\alpha$ -MF has been used in *P. pastoris* to produce recombinant proteins such as endoglucanase III from *Trichoderma harzianum* (Generoso et al. 2012), human P53 protein



**Fig. 1** General organization of the  $\alpha$ -mating factor ( $\alpha$ -MF) secretion signal originating from *S. cerevisiae* and applied in mediating protein secretion in *P. pastoris*. The pre-sequence consists of three parts and is cleaved by signal peptides in ER. The pro-sequence is cleaved at KR site by endoprotease Kex2 and the two EA repeats are removed by dipeptidyl aminopeptidase Ste13 in the Golgi

(Abdelmoula-Souissi et al. 2013), and manganese superoxide dismutase (PoMn-SOD) from *Pleurotus ostreatus* (Yin et al. 2014), etc. Many recombinant proteins have been successfully expressed in *P. pastoris* with their native signal peptides. For example, the activity of alkaline protease from *Aspergillus oryzae* with its native signal peptide is 1.5-fold higher than that with the  $\alpha$ -MF secretion signal peptide (Guo and Ma 2008); the activity of the laccase from white-rot fungus *Polyporus gramocephalus* TR16 with its native signal peptide is threefold higher than that with the  $\alpha$ -MF secretion signal peptide (Huang et al. 2011).

To enhance the efficiency of the  $\alpha$ -MF signal sequence, various approaches such as codon optimization (Xiong et al. 2005; Ahn et al. 2016), error-prone PCR mutagenesis (Rakestraw et al. 2009), deletion mutagenesis (Aggarwal and Mishra 2020; Chahal et al. 2017; Lin-Cereghino et al. 2013), and synthetic signal peptides (Aza et al. 2021; Obst et al. 2017) have been applied. Upon codon optimization of  $\alpha$ -MF, the phytase yield increased approximately sevenfold and CALB production increased by 132–295% compared to the version prior to codon optimization (Xiong et al. 2005; Ahn et al. 2016). Through error-prone PCR and library screening, one  $\alpha$ -MF mutant could increase the secretion of a single-chain antibody 4m5.3 up to 16-fold, compared to the wild type. This improvement has also been found in the production of other single-chain antibody fragments and two structurally unrelated proteins, interleukin-2 (IL-2) and horseradish peroxidase (HRP) (Rakestraw et al. 2009). A recent study in *S. cerevisiae* resulted in an optimized  $\alpha$ -MF named  $\alpha_{OPT}$  with four mutations ( $A\alpha 9D$ ,  $A\alpha 20T$ ,  $L\alpha 42S$ , and  $D\alpha 83E$ ), through a bottom-up (i.e., iterations of directed evolution on the native  $\alpha$ -MF) and top-down strategy (i.e., examining the evolved signal peptide, namely  $\alpha_{9H2}$  leader, and removing potential deleterious or neutral mutations) (Aza et al. 2021). The obtained  $\alpha_{OPT}$  could increase the secretion of two laccases, PK2 and ApL, approximately 14- and 26-fold compared to  $\alpha$ -MF, respectively. Combinatorial saturation mutagenesis at positions 86 and 87 of the  $\alpha_{OPT}$  leader could further enhance laccase secretion. It is appealing to apply this  $\alpha_{OPT}$  to protein expression in *P. pastoris*. In addition, deletion of amino acids 57–70 in the propeptide of  $\alpha$ -MF enhanced the HRP activity by more than 50% and CALB activity

approximately onefold compared to the wild-type  $\alpha$ -MF signal sequence (Lin-Cereghino et al. 2013). In a separate study, this strategy led to an increased titer of granulocyte colony-stimulating factor (G-CSF) to  $39.4 \pm 1.4$  mg/L (Aggarwal and Mishra 2020). Structural studies suggested that a specific orientation between both the N- and C-termini of  $\alpha$ -MF pro-peptide is required to interact with secretion machinery and therefore facilitate protein secretion. Mutations generated near these termini usually impact secretion negatively, and changes within the interior of the pro-peptide could benefit secretion if these mutations can stabilize the N- and C-termini (Chahal et al. 2017). By combining established leader sequences and  $\alpha$ -MF with deletions, Obst et al. designed several synthetic secretion signal peptides and characterized them with a red fluorescent protein (RFP) and yeast-enhanced green fluorescent protein (yEGFP) as reporters under different promoters. However, although these synthetic hybrid peptides yielded a more than tenfold variation in secretion efficiency, all except  $\alpha$ MF\_no\_EAEA with certain promoters were less efficient than  $\alpha$ -MF (Obst et al. 2017). The fusion of the *S. cerevisiae* Ost1 signal sequence and  $\alpha$ -MF pro-region with two mutations could enhance the secretion of far-red fluorescent protein E2-Crimson by 20-fold and lipase BTL2 by tenfold (Barrero et al. 2018).

In addition to modifications in  $\alpha$ -MF, putative secretory signal peptides can be determined by in silico analysis and further confirmed by experiments. Using five computer programs, SignalP4.1, Phobius, WolfPsort0.2, ProP1.0, and NetNG-lyc1.0, Massahi and Calik were able to identify eight signal peptides from the sequences of 56 endogenous and exogenous proteins that had higher D-scores than that of *S. cerevisiae*  $\alpha$ -MF (Massahi and Çalık 2015). Among the eight signal peptides, five with D-scores higher than 0.8 (SP13, SP23, SP24, SP26, and SP34) were selected for investigation of their efficiency in secreting recombinant human growth hormone. SP23 had the highest secretion efficiency, reaching 70%–80% of the efficiency of  $\alpha$ -MF (Massahi and Çalık 2016). There are also eight commercially available signal peptides (The PichiaPink™ Secretion Signal Set) for protein expression in *P. pastoris*. Table 1 summarizes the major signal peptides reported in the literature.

## 2.6 Co-expression of Chaperones to Facilitate Protein Folding

Secretory proteins enter the ER by translocation in an unfolded state and then undergo chaperone-assisted folding for maturation into their native conformation. Only properly folded proteins are exported from the ER to the Golgi apparatus for further modifications before delivery to intra- or extracellular destinations (Idiris et al. 2010). The folding of secretory proteins is error-prone. When unfolded proteins accumulate in the ER, unfolded protein responses are triggered to decrease the amount of newly unfolded proteins from entering the ER and increasing ER folding capacity. If the ER is overburdened by misfolded proteins, cell apoptosis occurs (Yu et al. 2015; Hetz et al. 2020). Misfolded proteins can also be transported from the ER to the cytosol for ubiquitination and subsequently degraded by

**Table 1** Signal peptides used for extracellular protein secretion in *P. pastoris*

Signal peptide	Origin	Target protein	References
$\alpha$ -MF	<i>S. cerevisiae</i> $\alpha$ -mating factor	Endoglucanase III (300 mg L <sup>-1</sup> )	Generoso et al. (2012)
W1	Synthetic signal peptide	$\beta$ -galactosidase (440 mg L <sup>-1</sup> )	Cao et al. (2017b)
nsB	<i>Candida antarctica</i> lipase B	Lipase A (220 U mL <sup>-1</sup> ) and lipase B (383 U mL <sup>-1</sup> )	Vadhana et al. (2013)
PHA-E	<i>Phaseolus vulgaris</i> agglutinin E	Snowdrop lectin (GNA) and GFP	Raemaekers et al. (1999)
PHOI	<i>P. pastoris</i> acid phosphatase 1	Rabies virus glycoprotein (RABV-G)	Ben Azoun et al. (2016)
SUC2	<i>S. cerevisiae</i> sucrose invertase	Recombinant human antithrombin (rAT, 324 mg L <sup>-1</sup> )	Kuwaie et al. (2005)
HFBII	<i>Trichoderma reesei</i> class II hydrophobin HFBII	Endoglucanase (EG27I, 47.7 mg L <sup>-1</sup> )	Su et al. (2017)
nsB-AP2	<i>Candida antarctica</i> lipase B + amphipathic peptide	Leech hyaluronidase (LHAase, 1.24 $\times$ 10 <sup>6</sup> U mL <sup>-1</sup> )	Kang et al. (2016)
SP23	<i>P. pastoris</i> protein disulfide isomerase	Human growth hormone (rhGH, 56 mg L <sup>-1</sup> )	Massahi and Çalik (2016)
SPIgG1	Murine IgG1	Anti-HIV antibody (VRC01, 3.05 $\mu$ g mL <sup>-1</sup> )	Aw et al. (2017)
HSA	<i>P. pastoris</i> human serum albumin	Human lysozyme (hLM)	Xiong and Chen (2008)
cSIG	Chicken lysozyme signal peptide	Classical swine fever virus glycoprotein E2	Li (2020)
SP killer protein	<i>S. cerevisiae</i> killer protein	Phytase	Helian et al. (2020)
SP invertase	<i>S. cerevisiae</i> invertase	Phytase	Helian, et al. (2020)
SP $\alpha$ -amylase	<i>A. niger</i> $\alpha$ -amylase	Phytase	Helian, et al. (2020)
SP Inulinase	<i>Kluyveromyces maxianus</i> Inulinase	Phytase	Helian, et al. (2020)
SP Lysozyme	<i>Gallus gallus</i> lysozyme	Phytase	Helian, et al. (2020)
Scw	<i>P. pastoris</i> Scw11p	EGFP, CALB	Liang et al. (2013)
Dse	<i>P. pastoris</i> Dse4p	EGFP, CALB	Liang et al. (2013)

(continued)

**Table 1** (continued)

Signal peptide	Origin	Target protein	References
Exg	<i>P. pastoris</i> Exg1p	EGFP, CALB	Liang et al. (2013)
MF4I	Codon-modified <i>Sc</i> $\alpha$ -mating factor	Phytase (6.1 g L <sup>-1</sup> )	Xiong et al. (2005)
HFBI	<i>Trichoderma reesei</i> class II hydrophobin HFBI	$\beta$ -galactosidase	Cao et al. (2017b)
AprE	<i>S. cerevisiae</i> , preregion of $\alpha$ -factor	$\beta$ -galactosidase	Cao et al. (2017b)
SP FAEC	<i>Talaromyces stipitatus</i> feruloyl esterase (FAEC)	FAEC (297 mg L <sup>-1</sup> )	Crepin et al. (2003)
SP Fae-1	<i>Neurospora crassa</i> Fae-1 feruloyl esterase	FAEC (260 mg L <sup>-1</sup> )	Crepin et al. (2003)
SP Bovine $\beta$ -casein	Bovine $\beta$ -casein	Xylanase (2,755.126 IU mL <sup>-1</sup> )	He et al. (2012)
Pptox	Viral K28 preprotoxin	GFP	Eiden-Plach et al. (2004)
SP pGKL	pGKL killer toxin	Mouse $\alpha$ -amylase (~240 mg L <sup>-1</sup> )	Kato et al. (2001)
SP DDDK	<i>PMT1</i> -gene-inactivated <i>P. pastoris</i> DDDK protein	Porcine carboxypeptidase B (CPB) and Erythrina trypsin inhibitor (ETI)	Govindappa et al. (2014)
Modified signal peptide (MSP)	Mouse salivary $\alpha$ -amylase (S8L) + pro-region of <i>S. cerevisiae</i> $\alpha$ -MF	Glucoamylase (GA, 12.619 IU mL <sup>-1</sup> )	Liu et al. (2005)

proteasome. This process is called ER-associated degradation (ERAD) (Römisch 2005). To enhance recombinant protein production in *P. pastoris*, endogenous or exogenous chaperones can be overexpressed to facilitate proper protein folding and secretion (Shen et al. 2012; Navone et al. 2021; Damasceno et al. 2007; Sallada et al. 2019; Jariyachawalid et al. 2012; Summpunn et al. 2018). There are two families of chaperones: molecular chaperones and chaperonins. Molecular chaperones bind to a short segment of substrate proteins and chaperonins form barrel-shaped folding chambers to sequester all or part of the unfolded proteins for proper folding (Evstigneeva et al. 2001).

Protein disulfide isomerase (Pdi) is a commonly used chaperone that is present in the ER lumen. It catalyzes both the formation and isomerization of disulfide bonds (i.e., changing an incorrectly bonded protein to a correct disulfide-bonded protein) and helps with the correct protein folding (Wilkinson and Gilbert 2004). Upon co-expression of Pdi with an IL-1 receptor antagonist and human serum albumin fusion protein (IH) that contains 18 disulfide bonds, there was a significant increase in the yield of IH, as compared to that from the strain expressing only the IH protein at a high copy (Shen et al. 2012). Another study with *E. coli* AppA

phytase that contains an extra non-consecutive disulfide bond showed that co-expression of Pdi increased the phytase ApV1 thermostability, and consequently, the production by ~12-fold compared to the expression of ApV1 alone (Navone et al. 2021). Immunoglobulin binding protein (BiP) is another abundant chaperone protein that resides in the ER. Belonging to the heat shock protein Hsp70 family, it facilitates protein folding and plays an important role in the ERAD pathway. Co-expression of BiP with an A33 single-chain antibody fragment (A33scFv) in *P. pastoris* increased the ER folding capacity and resulted in an approximately threefold increase in A33scFv secretion (Damasceno et al. 2007). Co-expression of the chaperon gene *KAR2* with different copies of the gene encoding hydrophobin (HFBI) also resulted in increased HFBI secretion. The highest HFBI secretion with 3-copy *HFBI* was  $22 \pm 1.6$ -fold higher than that of the strain overexpressing only single-copy *HFBI* (Sallada et al. 2019).

Apart from molecular chaperones, chaperonins have also been engineered to facilitate protein production in *P. pastoris*. D-phenylglycine aminotransferase (D-PhgAT) from *Pseudomonas stutzeri* ST-201 is an intracellular protein that is difficult to express in the soluble active form. Jariyachawalid et al. overexpressed this enzyme in *P. pastoris* and found that most of the D-PhgAT protein was insoluble. By co-expressing *E. coli* chaperonins GroEL-GroES intracellularly with D-PhgAT, a considerable amount of soluble D-PhgAT was produced, and the activity also increased significantly. Compared to the D-PhgAT gene expressed alone, a 14,400-fold higher volumetric activity was achieved when ten copies of chaperonins were co-expressed (Jariyachawalid et al. 2012). In another study, GroEL-GroES residing in the ER was co-expressed with extracellular bacterial phytase or intracellular D-PhgAT in *P. pastoris*. The volumetric activity of extracellular phytase was 1.5–2.3-fold higher than that of phytase expression alone. However, the majority of the D-PhgAT protein was inactive and found in the insoluble protein fraction (Sumppunn et al. 2018). These results suggested that the GroEL-GroES chaperone could potentially enhance the production of functional proteins in *P. pastoris* when they are present within the same compartment. Some of the major chaperones overexpressed in *P. pastoris* are summarized in Table 2.

## 2.7 Cell Surface Display

Cell surface display is a promising method for engineering functional proteins to be expressed on the cell surface through fusing with an anchor protein. Applications include, but are not limited to: whole-cell biocatalysts, bioadsorption and bioremediation, biosensor design, vaccine and antibody development, epitope mapping, library screening, protein engineering (Gai and Wittrup 2007; Kuroda and Ueda 2011; Tanaka et al. 2012; Ueda 2016; Andreu and Olmo 2018). The anchor protein can be fused with a target protein either at either the N-terminus or at the C-terminus (Tanaka et al. 2012). Both the fusion order and the linker between a target protein and an anchor protein can affect the display efficiency and functional properties (Ueda 2016). Commonly used anchor proteins in *P. pastoris* are Aga1 (Wang

**Table 2** Commonly used chaperones to increase recombinant protein production in *P. pastoris*

Chaperone	Target protein	References
Glutathione peroxidase 1 (Gpx1)	RABV-G	Ben Azoun et al. (2016)
Protein disulfide Isomerase (Pdi1p)	HFBI, lipase, rhG-CSF, and recombinant <i>Bombyx mori</i> acetylcholinesterase 2 (rBmAChE2)	Sallada et al. (2019), Sha et al. (2013), Zhang et al. (2006), Li (2021)
ER-oxidoreductin 1 protein (Ero1p)	HFBI	Sallada et al. (2019)
CNE1p, FAD1p	$\Delta^9$ -tetrahydrocannabinolic acid synthase (THCAS)	Zirpel et al. (2018)
Bmh2, Sso2, Ssa4, Vhb	<i>Rhizopus oryzae</i> lipase (ROL)	Jiao et al. (2018)
Sse1, Bfr2, Cup5, Kin2	Fab fragment of a monoclonal antibody fragment (2F5mAb)	Gasser et al. (2007)
Immunoglobulin binding protein (BiP/Kar2p)	A33scFv, HFBI	Damasceno et al. (2007), Sallada et al. (2019)
Ydj1p, Ssa1p, Sec63p	Lipase	Zhang et al. (2006)
HAC1	Chitinase, lipase	Jiao et al. (2018), Gasser et al. (2007), Song et al. (2020)
Sis1	Porcine growth hormone (pGH)	Deng et al. (2020)
GroEL-GroES	D-phenylglycine aminotransferase (D-PhgAT)	Jariyachawalid et al. (2012), Sumppunn et al. (2018)

et al. 2007; Su et al. 2010a; Dong 2013), Sed1 (Su et al. 2010b; Li et al. 2015a), Tip1 (Jo et al. 2011), Aga2 (Jacobs et al. 2008), and Flo1 (Jiang et al. 2007), all of which are from *S. cerevisiae*, as well as, Pir1 (Khasa et al. 2011; Yang et al. 2017) and Pir2 (Khasa et al. 2011) from *P. pastoris*. In another study, 13 endogenous glycosylphosphatidylinositol-modified cell wall proteins were identified upon screening the genome of *P. pastoris* GS115 (Zhang et al. 2013), three of which were chosen as anchor proteins for displaying CALB (Wang et al. 2017). These three anchors (i.e., GCW21, GCW51, and GCW61) have also been applied to display bacterial PETase on the surface of *P. pastoris*, to degrade highly crystallized polyethylene terephthalate (PET). The turnover rate of the whole-cell biocatalyst displaying PETase was approximately 36-fold higher than that of the purified PETase (Chen 2020). Another anchor protein identified in *P. pastoris* is Flo9. The displayed lipase B with Flo9 showed higher thermostability at 45 °C and stability in organic solvents (Moura 2015).

*P. pastoris* X-33 has also been engineered to assemble protein complexes such as minicellulosomes on the cell surface (Ou and Cao 2014). The truncated CipA, which contains a cellulose-binding module and two cohesin modules from *Clostridium acetobutylicum*, was fused to the C-terminus of the anchor flocculation

protein Flo1 from *S. cerevisiae*, whereas a *Nasutitermes takasagoensis* endoglucanase (NtEG) was fused with the dockerin. Fusion proteins were expressed separately in two *P. pastoris* X-33 strains, which were co-cultured for minicellulosome assembly. The surface displayed CipA and assembly of cohesin and dockerin were confirmed using immunofluorescence and western blotting. The hydrolysis efficiencies of NtEG for carboxymethyl cellulose (CMC), microcrystal cellulose (Avicel), and filter paper were enhanced by 1.4-fold, 2.0-fold, and 3.2-fold, respectively, when compared to free NtEG. Another study conducted by Dong et al. utilized an ultra-high-affinity IM7/CL7 protein pair for minicellulosome assembly (Dong et al. 2020). IM7 (including one, two, or three units) was fused to the N-terminus of the anchor protein SED1 from *S. cerevisiae* and expressed in *P. pastoris*. An endoglucanase (EG), an exoglucanase (CBH), a  $\beta$ -glucosidase (BGL), and a carbohydrate-binding module (CBM) from *Thermobifida fusca*, each fused with an N-terminal CL7 tag, were expressed individually in *E. coli*. The secreted proteins from *E. coli* cultures were assembled and displayed on the *P. pastoris* cell surface in vitro. The display system with two or three IM7 showed comparable or even higher efficiency for the hydrolysis of Avicel, phosphoric acid-swollen cellulose (PASC) and CMC, compared to free cellulases. The ethanol titer reached 5.1 g/L when three IM7 units were engaged in CMC fermentation.

In a recent study, Silva et al. displayed specific immunogenic epitopes of ZIKV envelope, NS1 protein, and both on the surface of *P. pastoris* GS115 by fusing these epitopes at the N-terminus of a partial Ag $\alpha$ 1 (C-terminal portion, nucleotides 970–1950) (Silva et al. 2021). The ability of the recombinant yeast to stimulate immune cells was evaluated in vitro using mouse immunological cells isolated from the spleen. *P. pastoris* displaying EnvNS1 epitopes showed better efficacy in producing IL-6, IL-10, and tumor necrosis factor-alpha (TNF- $\alpha$ ) cytokines and an increase in lymphocytes CD4<sup>+</sup>, CD8<sup>+</sup>, and CD16<sup>+</sup>, similar to ZIKV. These epitopes will be beneficial for the development of vaccines against ZIKV infection.

---

### 3 Applications of *P. pastoris* in Biomanufacturing

After more than 20 years of development, *P. pastoris* has become one of the most popular protein expression systems that is widely used in protein preparation, structural analysis, and functional characterization. As a GRAS microorganism approved by the United States Food and Drug Administration, thousands of proteins, including medicinal proteins (i.e., insulin, human serum albumin, hepatitis B surface antigen, and epidermal growth factor (Weinacker et al. 2013)) and industrial enzymes (i.e., mannanase, phytase, xylanase, and lipase (Rabert et al. 2013)) have been successfully expressed in *P. pastoris*. In addition, due to the development of pathway assembly and genome editing tools, a growing interest has been seen in establishing *P. pastoris* as a microbial cell factory to produce chemicals and natural products.



## 3.1 Recombinant Proteins

Recombinant proteins can be produced using bacterial, yeast, mold, insect, plant, and mammalian expression systems. *P. pastoris* is particularly attractive for the large-scale production of recombinant proteins. It naturally contains several highly expressed genes that encode methanol assimilation and dissimilation pathway enzymes, enabling growth with methanol as the sole carbon and energy source (Wegner and Harder 1987). Thus, high-level expression of target proteins can be readily achieved using these methanol-inducible promoters. When compared with the plant and mammalian expression systems, the *P. pastoris* expression system offers the advantages of low cost, fast growth, high cell density fermentation (HCDF), and consequently high expression levels. In contrast to prokaryotes, the biggest advantage of *P. pastoris* is the capability of post-translational modifications (e.g., O- and N-glycosylation and disulfide bond formation. When compared to *S. cerevisiae*, *P. pastoris* poses little concerns regarding over-glycosylation, and the secretory expression level of recombinant proteins is much higher (Karbalaei et al. 2020). Therefore, *P. pastoris* has been widely used to produce therapeutic glycoproteins. Moreover, *P. pastoris* can efficiently secrete the target proteins into the fermentation broth, making the downstream separation and purification process simpler, which is a paramount variable in designing viable industrial-scale processes.

### 3.1.1 Medicinal Proteins

Recombinant proteins represent a growing market in medical biotechnology. Many approved biopharmaceuticals are protein-based, such as monoclonal antibodies, growth factors, blood factors, hormones, interleukins, anticoagulants, interferons, and vaccines. Some of the representative medicinal proteins produced by *P. pastoris* are listed in Table 3.

Vaccines represent the largest class of recombinant medicinal proteins produced by *P. pastoris*. Vaccines can be divided into three types: inactivated vaccines, live attenuated vaccines, and recombinant subunit vaccines (Gasser et al. 2006). Recent studies have found that *P. pastoris* is preferred for producing recombinant subunit vaccines compared to other expression systems (Gasser et al. 2007). Compared to the non-glycosylated antigen, the mannose-glycosylated antigen produced by *P. pastoris* has enhanced antigen presentation and T cell activation. Enterovirus 71 (EV71) is the main pathogen that causes hand-foot-mouth disease in children. The establishment of a microbial system for large-scale and safe production of the EV71 vaccine has value in medicinal applications. Yang et al. cloned *P1* and *3C* genes of EV71 and established a microbial system for efficient production of recombinant EV71-VLP (virus-like particle) in *P. pastoris*. Expression levels of as high as 270 mg/L EV71-VLPs antigen have been achieved (Yang et al. 2020). Another example is related to cervical cancer, the fourth most common cancer that threatens the health of women worldwide. Although there is already a market-oriented vaccine, its high price limits its wide application. Recently, Sanchooli et al. inserted the cross-neutralizing epitope of L2-HPV-16 into L1-HPV-16 to form

**Table 3** Representative medicinal proteins expressed in *P. pastoris*

Product	Strain	Vector	Titer (mg/L)	Application	References
Bovine IFN- $\alpha$	GS115	pPIC9K	200	Prevention and therapy of viral diseases	Tu et al. (2016)
hPAB- $\beta$	GS115	pPIC9K	241	Antibacterial peptide	Chen et al. (2011)
Bovine lactoferrin	KM71H	pJ902	3500	Transferrin and antibacterial protein	Iglesias-Figueroa, et al. (2016)
Human topoisomerase I	GS115 SMD1168H	pPICZ $\alpha$	11	Antitumor	Chan et al. (2018)
Insulin	SuperMan5	pPICZ $\alpha$	5	Diabetes	Baeshen et al. (2016)
Human gastric lipase	X-33	pGAPZ $\alpha$	7	Diseases of the digestive system	Sams et al. (2017)
Rabies virus glycoprotein	GS115	pHIL-S1	1.23	Rabies vaccine	Ben Azoun et al. (2016)
L1-L2 proteins of HPV virus type 16	GS115	pBLHIS-IX	23.61	Cervical cancer	Bredell et al. (2018)
Human sialyltransferase	KM71H	pPICZ $\alpha$ B	130	Pharmacological uses	Luley-Goedl et al. (2016)
Human antiplatelet scFv antibody	X-33	pPICZ $\alpha$ A	600	Treatment of atherosclerosis	Vallet-Courbin (2017)
IL-1 $\beta$	GS115 SMD1168 X-33	pPICZ $\alpha$ A	250	Proinflammatory cytokine	Li et al. (2016a)
IL-3	X-33	pPICZ $\alpha$ A	2230	Multipotent hematopoietic cytokine	Dagar and Khasa (2018)
IL-11	GS115	pPINK $\alpha$ HC	300	Thrombopoietic growth factor	Yu et al. (2018)
IL-15	X-33	pPICZ $\alpha$ A	75	Differentiation and proliferation of T, B, and NK cells	Sun et al. (2016)

an L1/L2-HPV-16 chimeric fragment, which was cloned into the pPICZA plasmid for heterologous expression. The chimeric protein could be positively detected by both L1-HPV-16 and L2-HPV-16 antibodies (Sanchooli et al. 2018). Meanwhile, when Bredell et al. expressed the HPV-16L1/L2 chimeric protein (VLP) in *P. pastoris* KM71 (Mut<sup>S</sup>) or GS115 (Mut<sup>+</sup>) under a constant dissolved oxygen level

(DO stat) fed-batch culture supplemented with methanol, they achieved a titer of 23.61 mg/L for the chimeric protein (Bredell et al. 2018).

Although a panel of medicinal proteins has been expressed in *P. pastoris*, correct folding of target proteins is a major concern. Due to the lack of sufficient chaperone factors, a considerable number of recombinant proteins cannot fold into their correct configurations. To overcome this limitation and improve the expression level, rational design and reverse engineering strategies can be adopted to improve the protein folding microenvironment. As mentioned above, genes related to protein folding in the ER, such as BiP (an Hsp70 chaperone), can help secretory proteins fold correctly. The secretion of A33scFv fragment was increased approximately threefold upon overexpression of the chaperone protein BiP in *P. pastoris* (Damasceno et al. 2007). Overexpression of Pdi (responsible for the formation of disulfide bonds) in *P. pastoris* can increase the expression of the antibody protein 2F5Fab by 2-fold (Gasser et al. 2006). Zhang et al. tested three factors related to protein transport from *S. cerevisiae* (Sec63p, Ydj1p, and Ssa1p), whose overexpression increased the expression of GCSF by 2.8-, 3.6-, and 6.8-fold, respectively, in *P. pastoris*. Therefore, finding suitable protein mates remains a major challenge in establishing an efficient recombinant protein expression system (Zhang et al. 2006). Gasser et al. identified new protein chaperone genes, including *CUP5*, *SSA4*, *BMH2*, *KIN2*, *SSE1*, and *BFR2* at the transcriptional level. The overexpression of these genes significantly enhanced the secretion of 2F5Fab antibody in *P. pastoris*, with the final titer of 2F5Fab reaching up to 47.27 mg/L (Gasser et al. 2007). Stadlmayr et al. established a cDNA overexpression library in *P. pastoris* and identified three new protein chaperones as the secretion-enhancing factors, which increased the expression level of the model protein by up to 45% (Stadlmayr et al. 2010). Huang et al. identified six significantly upregulated genes related to recombinant protein production using comparative proteomic analysis. In particular, the co-expression of *TPX*, *FBA*, and *PGAM* increased the expression level of the reporter gene by 2.46-, 1.58-, and 1.33-fold, respectively (Huangfu et al. 2015). Noteworthy, owing to the different properties of foreign proteins, there is a dearth of generally applicable engineering approaches. In other words, the optimal protein chaperone or secretory factors can be different case by case and should be evaluated individually. For example, overexpression of Pdi in *P. pastoris* failed to increase the production of A33scFv antibody protein and overexpression of BiP even reduced the yield of glucose oxidase by 10-fold (Heide et al. 2002).

### 3.1.2 Industrial Enzymes

In recent years, industrial enzymes have been increasingly leveraged in the chemical, food and beverage, pharmaceutical, cosmetic, and textile industries. Owing to the increasing demand for industrial enzymes, the development of production strategies has accelerated. In this regard, the effectiveness of *P. pastoris* as a host for high-level expression of recombinant proteins has attracted increasing attention, because of the presence of strong methanol-inducible promoters (more than 30% of the total proteins) and HCDF (higher than 200 g/L biomass). In addition, *P. pastoris* has a strong ability to secrete target proteins into the fermentation

medium, facilitating downstream purification at a much lower cost. Therefore, *P. pastoris* has been regarded as a favorable host for large-scale production of industrial enzymes. Representative industrial enzymes produced in *P. pastoris* using HCDF are listed in Table 4.

As a type of hydrolase, lipase demonstrates high regioselectivity and stereoselectivity and can catalyze ester hydrolysis and transesterification. Therefore, lipases are widely used in food, cosmetics, and pharmaceuticals. Zheng et al. cloned

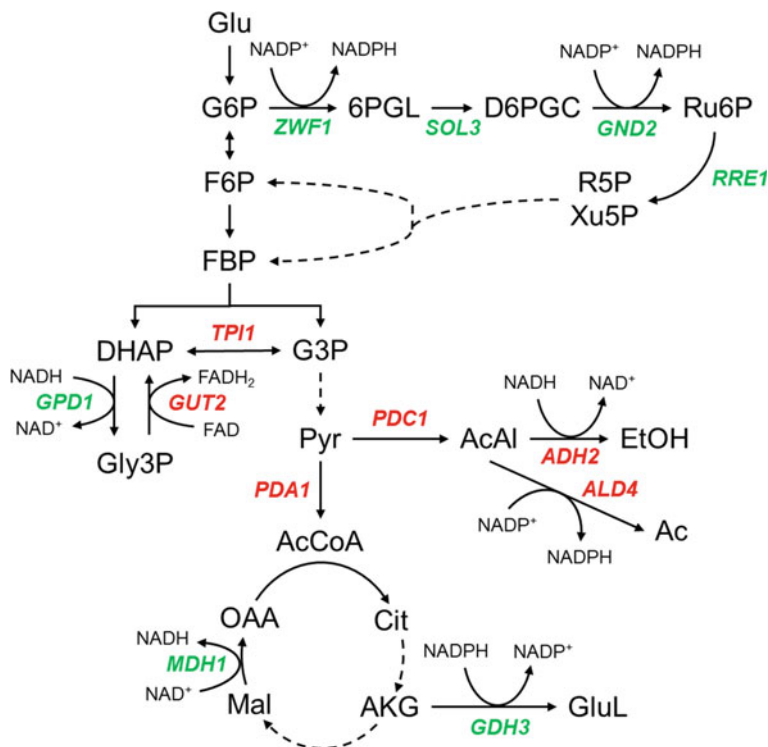
**Table 4** Representative industrial enzymes expressed in *P. pastoris*

Product	Strain	Vector	Scale (L)	Titer (g/L)	Application	References
$\beta$ -Mannanase	GS115	pPIC9K	5	0.17	Production of manno-oligosaccharides	Li et al. (2018)
Invertase	KM71	pPIC9	7	2.5	Producing inverted sugar	Veana et al. (2014)
Hyaluronidase	GS115	pPIC9K	3	0.42	Degradation of hyaluronic acid	Jin et al. (2014)
Xylanase	X-33	pPICZ $\alpha$ A	50	9.88	Bakery, beverages, starch processing	Li et al. (2015b)
Pectinase	GS115	pAO815	–	0.35	Brewing, papermaking	Peng et al. (2019)
Glucose oxidase	GS115	pPIC9	40	0.83	Bakery, eggs, starch processing	Meng et al. (2014)
Formate dehydrogenase	X-33	pPICZ $\alpha$ A	3	0.05	Cofactor regeneration, CO <sub>2</sub> reduction	Duman (2020)
Peroxidase	DSMZ 70,382	pUC57	14	0.94	Diagnostics, bioremediation, lignin degradation	Majeke (2020)
L-Glutamate oxidase	GS115	pHBM905A	20	11	Transformation of L-glutamic acid, industrial fermentation, biosensors	YaPing et al. (2017)
Laccase	X-33	pGAPZA	5	0.495	Dye decolorization, beverages	Kittl et al. (2012)
Hydrolases	X-33	pPICZ $\alpha$ A	5	0.61	Triacylglycerol lipase	Zheng et al. (2019)
Lipase B	SMD1168	–	10	2	Resolution of chiral compounds	Jahic et al. (2003)
Tannin acyl hydrolase	KM71	pPIC9K	7	0.074	Detannification of food	Zhong et al. (2004)
Phytase	GS115	pPICZ $\alpha$ A	10	9.58	Animal feeding	Li et al. (2015c)

the *Aspergillus oryzae* lipase gene (*AOL*) to yield the plasmid pPICZ $\alpha$ A-AOL, which was subsequently integrated into the genome of *P. pastoris* X-33. Using the methanol feeding strategy, AOL with a specific activity of 432 U/mg was obtained in a 5 L bioreactor (Zheng et al. 2019). To increase the yield of lipase, Zhang et al. employed a fusion expression strategy by fusing small ubiquitin modifying protein (SUMO) with *Aspergillus niger* lipase (ANL) to obtain SANL. The resultant chimeric gene was cloned into pPIC9K for heterologous expression in *P. pastoris* GS115. The highest activity of SANL was ~960 U/mL in a 3 L fermenter, which was 1.85-fold higher than that of its parent ANL (Zhang et al. 2019a).

Although the expression of recombinant proteins is mainly induced by methanol, a co-substrate culture strategy has been found to increase the yield of industrial enzymes. Berrios et al. cloned the *Rhizopus oryzae* lipase gene (*ROL*) and constructed the plasmid pPICZ $\alpha$ A-ROL for heterologous expression in *P. pastoris* X-33. The engineered strain was continuously cultured with methanol and glycerol as co-substrates in a 1.5 L BioStatAplus bioreactor. The results showed that using glycerol as a co-substrate at 22 and 30 °C could increase the volumetric productivity of recombinant lipase and reduce the consumption of methanol (Berrios et al. 2017). In addition, the co-substrate culture could also be applied for the industrial production of phytase. As an animal feed additive, phytase can decompose phytic acid and greatly reduce the input of animal feed. Li et al. engineered phytase production in *P. pastoris* by modifying *P<sub>AOX1</sub>* and the  $\alpha$  factor signal peptide and increasing gene copy numbers. Phytase activity as high as 2,119 U/mL with a corresponding titer of 0.75 g/L was achieved, which was 4.12-fold higher than that of the parent strain. In a 10 L fermenter, using glycerol and methanol as co-substrates for fed-batch fermentation, the titer and enzyme activity of phytase could be further improved to as high as 9.58 g/L and 35,032 U/mL, respectively (Li et al. 2015c).

Besides the traditional strategies in engineering secretion signals and modifying *P<sub>AOX1</sub>* promoter, genome-scale metabolic models can be employed to regulate the metabolic fluxes from a systems perspective, to improve the expression level of recombinant proteins. Saitua et al. employed the dynamic flux balance analysis (dFBA) framework to establish a dynamic genome-scale metabolic model, to simulate recombinant protein expression process in *P. pastoris* (Saitua et al. 2017). Starting with seven state variables including glucose, biomass, and fermentation quantity, they analyzed the kinetics of substrate assumption and distribution of metabolic flux. On this basis, Nocon et al. optimized the dFBA algorithm and predicted gene targets (including both gain- and loss-of-function targets), to enhance the production of recombinant proteins. Overexpression targets were identified to reside in the pentose phosphate pathway and the TCA cycle, whereas knockout targets were found to belong to several branch points of glycolysis (Fig. 2). Five out of the nine predicted targets were found to increase the expression level of cytosolic human superoxide dismutase (hSOD). More importantly, most of the same genetic modifications led to enhanced expression of bacterial  $\beta$ -glucuronidase, indicating the general applicability of the identified metabolic engineering targets (Nocon et al. 2014).



**Fig. 2** Implementation of a genome-scale metabolic model to predict gene overexpression and knockout targets of the central metabolism for increased production of recombinant proteins in *P. pastoris* (Nocon et al. 2014). Overexpressed genes are shown in green and deleted genes are shown in red. *ZWF1*: glucose-6-phosphate dehydrogenase; *SOL3*: 6-phosphogluconolactonase; *GND2*: phosphogluconate dehydrogenase; *MDH1*: malate dehydrogenase; *TPI1*: triose-phosphate isomerase; *ADH2*: alcohol dehydrogenase; *ALD4*: aldehyde dehydrogenase; *PDA1*: pyruvate dehydrogenase; *PDC1*: pyruvate decarboxylase; *RPE1*: ribulose 5-phosphate 3-epimerase; *GPD1*: glycerol-3-phosphate dehydrogenase; *GUT2*: glycerol-3-phosphate dehydrogenase; *GDH3*: glutamate dehydrogenase

### 3.2 Bulk Chemicals

With the rise of synthetic biology, yeast has become an important cell factory to produce fine chemicals. Currently, *S. cerevisiae* is the preferred host to produce a wide range of chemicals, including, but not limited to: glycerol, L-propanediol, lactic acid, succinic acid, and isoprene. Comparative metabolomics indicated that the intermediate metabolites in *P. pastoris* could cover more than 90% of those in *S. cerevisiae*, indicating great potential for chemical production in *P. pastoris* (Carnicer et al. 2012). Currently, the production of S-adenosyl-L-methionine (Chu et al. 2013), xylitol (Louie et al. 2021), hyaluronic acid (Oliveira et al. 2016), gluconic acid (Liu et al. 2016), and lactic acid (Lima et al. 2016) has been

reported, confirming the possibility of producing simple and complex chemicals in *P. pastoris*.

To achieve efficient and cost-effective production of chemicals, a combination of metabolic engineering modification modifications and bioprocess optimization is generally employed. Cheng et al. constructed the glucose-D-arabitol-D-xylulose-xylitol pathway to produce xylitol from glucose for the first time. The D-arabitol dehydrogenase gene from *Klebsiella pneumoniae* and the xylitol dehydrogenase gene from *Gluconobacter oxydans* were cloned into pPIC9K for subsequent integration into the genome of the GS225 strain, a derivative strain of *P. pastoris* GS115 strain after adaptive evolution. The yield of xylitol was 0.078 g/g glucose when it was fermented in a 3 L fermenter (Cheng et al. 2014).

Mixed carbon source fermentation may be more beneficial for promoting product production, and this strategy has been used in the production of a variety of chemicals, such as glucaric acid. As an organic acid, glucaric acid is considered to be one of “the most valuable chemicals from biomass” and plays an important role in the synthesis of many biodegradable substances. Much attention has been paid to the production of glucaric acid using a microbial cell factory. Liu et al. overexpressed the inositol oxygenase gene (*MIOX*) and urinate dehydrogenase gene (*UDH*) from *Pseudomonas aeruginosa* KT2440 for glucaric acid production in *P. pastoris*. As *MIOX* was determined to be rate-limiting, fusion expression of *MIOX* with *UDH* with a flexible linker was employed to improve the conversion efficiency. With glucose and myo-inositol as the co-substrates in fed-batch fermentation, the engineered *P. pastoris* strain was able to produce glucaric acid at a titer as high as  $6.61 \pm 0.30$  g/L (Liu et al. 2016).

2-Phenylethanol (2-PE) is widely used in cosmetics and high-end perfumes because of its rose flavor. In a recent study, Kong et al. overexpressed the 2-ketoacid decarboxylase gene (*ARO10*), aldehyde reductase gene (*ADH6*), and aromatic aminotransferase gene (*ARO8*) from *S. cerevisiae*, together with feedback inhibition-insensitive mutant genes, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (*aroG<sup>fbr</sup>*) and chorismate mutase/prephenate dehydratase (*pheA<sup>fbr</sup>*) from *E. coli*, under the control of the strong constitutive promoter, *P<sub>GAP</sub>*, and achieved de novo biosynthesis of 2-PE from glucose for the first time. Using shake flask fermentation for 36 h, 1,169 mg/L of 2-PE was found to accumulate in the engineered *P. pastoris* strain (Kong 2020). Notably, the titer of 2-PE synthesized by *P. pastoris* was higher than those achieved in *E. coli* and *S. cerevisiae*, indicating that *P. pastoris* has excellent potential as a host strain to produce chemicals.

As mentioned earlier, although *P. pastoris* has been used to synthesize various compounds, most of the engineered strains still use glucose and other fermentable sugars as substrates (Pena et al. 2018). While it is often added as an inducer, methanol has rarely been used as a substrate. Owing to the increasing concerns about sustainability and the abundant availability of methanol, the use of methanol as a substrate to produce chemicals has become a research hotspot. Cai et al. used methanol as the sole carbon and energy source to produce Monachine J and lovastatin in metabolically engineered *P. pastoris*, with titers reaching 60.0 mg/L

and 14.4 mg/L, respectively. After bioprocess optimization, including the employment of a co-culture strategy and fed-batch fermentation with glycerol as the co-substrate, the yield of Monachine J and lovastatin reached 594 mg/L and 251 mg/L, respectively. In terms of methanol conversion, Yamada et al. integrated *D-LDH* into the *rDNA* loci of *P. pastoris*. Multicopy integration of the *D-LDH* expression cassette was achieved following post-transformational gene amplification and selection on gradually increasing zeocine concentrations. The optimally engineered *P. pastoris* strain produced D-lactic acid with a titer of 3.48 g/L by means of test-tube fermentation for 96 h, with methanol as the sole carbon source. This is the first report on the establishment of *P. pastoris* as a microbial cell factory for the conversion of methanol to lactic acid (Yamada et al. 2019). This study provides a basis for the application of gene integration strategy at the *rDNA* loci in *P. pastoris* for the construction of microbial cell factories for the production of value-added chemicals. Although research on the synthesis of chemicals from methanol is still in its infancy and the titer is still not high enough for industrial production, the challenges in methanol conversion should be able to be addressed using metabolic engineering and synthetic biology approaches in the future. *P. pastoris* is believed to play an increasingly important role in biorefinery and biomanufacturing in the near future.

### 3.3 Natural Products

Key enzymes of secondary metabolite (natural product) biosynthetic pathways are often found to have low expression levels and/or limited enzymatic activities, which becomes the bottleneck for efficient biosynthesis of high-value natural products. Considering the advantages of high-level expression and post-translational modifications of complex eukaryotic proteins (i.e., cytochrome P450s, CYPs), *P. pastoris* is a promising microbial cell factory for the synthesis of complex biologically active molecules. Currently, natural products synthesized in *P. pastoris* mainly include terpenoids, polyketides, and flavonoids (Table 5).

Terpenoids are hydrocarbons that are widely found in plants and microorganisms. Many terpenoids have important physiological activities and therefore are important research targets for the development of new drugs. The biosynthetic pathways of a series of terpenoids, such as lycopene, carotene, astaxanthin, (+)-nootkatone, and dammarenediol-II, have been successfully constructed in *P. pastoris*. (+)-Nootkatone is a sesquiterpene compound of great commercial value, with a grapefruit aroma and various biological activities. Although the existing chemical synthesis technology of (+)-nootkatone can meet industrial and commercial needs, heavy metals and flammable compounds are involved in the synthesis process. Therefore, the synthesis of (+)-nootkatone using biological methods is a new trend in the near future (Fig. 3). Through co-expression of the prenaspirodiene oxygenase (HPO) from *Hyoscyamus muticus* and the cytochrome P450 reductase from *A. thaliana*, *trans*-nootkatol was produced by hydroxylation of (+)-valencene, and *trans*-nootkatol was further oxidized to (+)-nootkatone by



**Table 5** Production of natural products in *P. pastoris* cell factories

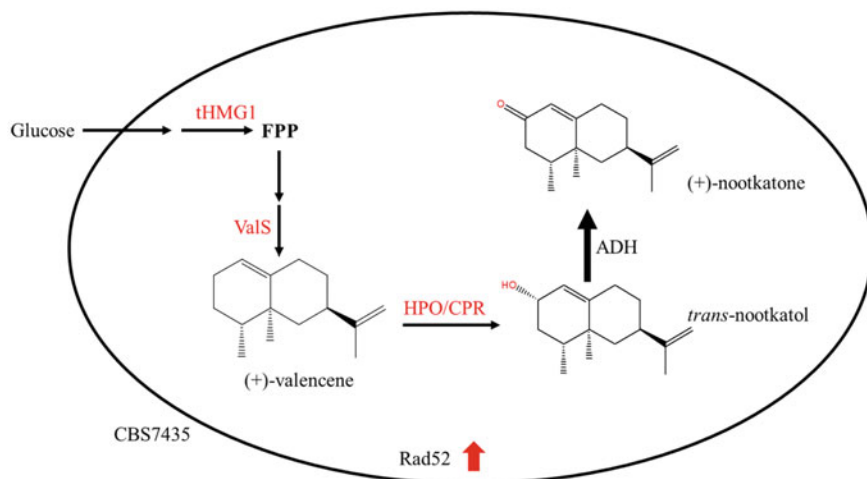
Nature product	Host	Genes	Promoter	Yield	References
Lycopene	X-33	<i>crtE, crtB, crtI</i>	$P_{GAP}$	1.141 $\mu\text{g/g}$ DCW	Araya-Garay et al. (2012a)
$\beta$ -Carotene	X-33	<i>crtE, crtB, crtI, crtL</i>	$P_{GAP}$	339 $\mu\text{g/g}$ DCW	Araya-Garay et al. (2012a)
Astaxanthin	X-33	<i>crtE, crtB, crtI, crtL, crtW, crtZ</i>	$P_{GAP}$	3.7 $\mu\text{g/g}$ DCW	Araya-Garay et al. (2012b)
Nootkatone	CBS7435 <i>his4 ku70</i>	<i>CPR, HPO, ADH, truncated HMG1</i>	$P_{AOX1}$	17 mg/L 208 mg/L (fed-batch)	Wriessnegger et al. (2014)
Dammarenediol-II	GS115	<i>ERG1, ERG7</i>	$P_{AOX1}$	1.073 mg/L DCW	Liu et al. (2015)
		<i>Repression</i>	$P_{THL11}$		
		<i>PgDDS</i>	$P_{AOX1}$		
Menaquinone-4	GS115	<i>SaGGPPS</i>	$P_{GAP}$	0.24 mg/g DCW	Sun et al. (2019)
		<i>HsUBIADI</i>	$P_{AOX1}$		
		<i>IDI</i>	$P_{GAP}$		
		<i>NpgA, ATX</i>	$P_{AOX1}$	2.2 g/L (fed-batch)	Gao et al. (2013)
Citrimin	GS115	<i>pksCT, MPL1, MPL2</i>	$P_{AOX1}$	0.6 $\pm$ 0.1 mg/L	Xue et al. (2017)
		<i>MPL4, MPL6</i>			
		<i>MPR7</i>			
Lovastatin	GS115	<i>LovA, LovB</i>	$P_{AOX1}$	24.6 mg/L	Liu et al. (2018)
		<i>LovC, LovD</i>		250.8 mg/L (fed-batch)	
		<i>LovF, LovG</i>			

(continued)

**Table 5** (continued)

Nature product	Host	Genes	Promoter	Yield	References
17-Hydroxyprogesterone	GS115	<i>NpgA</i> Human CYP17 cDNA	$P_{Aox1}$	Not reported	Kolar et al. (2007)

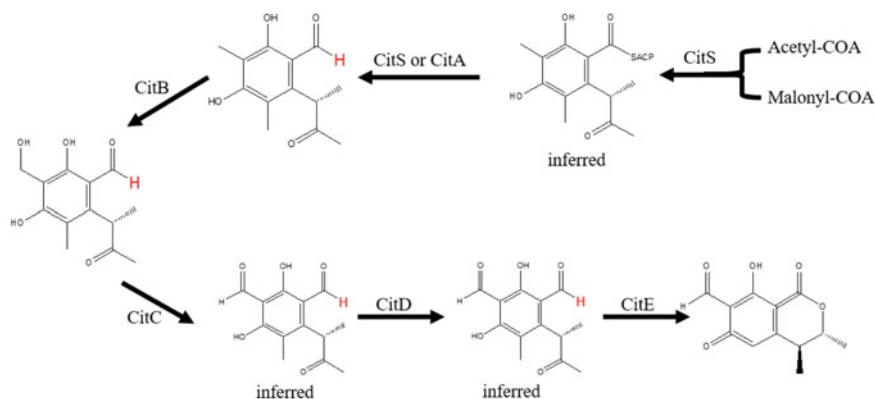
*Note* *CrtE*: geranylgeranyl diphosphate synthase; *CrtB*: phytoene synthase; *CrtI*: phytoene desaturase; *CrtL*: lycopene  $\beta$ -cyclase; *CrtW*: xanthophylls; *CrtZ*:  $\beta$ -carotene hydroxylase; *CPR*: *A. thaliana* cytochrome P450 reductase; *HPO*: premmaspirodiene oxygenase of *H. muticus*; *ADH*: alcohol dehydrogenase; truncated *HMG1*: truncated hydroxy-methylglutaryl-CoA reductase; *ERG1*: 2,3-oxidosqualene synthase; *ERG7*: responsible for conversion of 2,3-oxidosqualene to lanosterol; *PgDDS*: dammarenediol-II synthase from *Panax ginseng*; *SaGGPPS*: geranylgeranyl diphosphate synthase from the archaeobacterium *Sulfolobus acidocaldarius*; *HsUBIAD1*: a human homologue of *E. coli* prenyltransferase menA and mammalian mitochondrial prenyltransferase COQ2; *IDI*: IPP isomerase; *NpgA*: *A. nidulans* phosphopantetheinyl transferase; *ATX*: *Aspergillus terreus* 6-methylsalicylic acid synthase; *pk5CT*: *M. purpureus* citrinin polyketide synthase; *MPL1*: serine hydrolase; *MPL2*: oxygenase; *MPL4*: dehydrogenase; *MPL6*: short-chain dehydrogenase; *MPL7*: oxidoreductase; *MPRI*: a transporter; *LovA*: a cytochrome P450 monooxygenase; *LovB*: lovastatin monaketide synthase; *LovC*: enoyl-reductase; *LovD*: acyl-transferase; *LovF*: lovastatin diketide synthase; *LovG*: thioesterase; Human CYP17 cDNA: catalyzing the conversion of progesterone to 17-hydroxyprogesterone as well as 16-hydroxyprogesterone



**Fig. 3** Synthetic pathway of (+)-nootkatone in *P. pastoris*. (+)-Valencene is synthesized from farnesyl pyrophosphate by valencene synthase (ValS). HPO/CPR converts (+)-valencene into *trans*-nootkatol, which is further converted by the endogenous alcohol dehydrogenase (ADH) to form (+)-nootkatone. Enzymes that can convert *trans*-nootkatol to (+)-nootkatone naturally exist in *P. pastoris*. Overexpression of the endogenous ADH and the truncated hydroxymethylglutaryl coenzyme A reductase (tHMG1) from *S. cerevisiae*, as well as the endogenous Rad52 significantly increased the production of (+)-nootkatone. Exogenous genes are shown in red and bold arrows represent overexpressed endogenous genes. HPO: premnaspriodiene oxygenase from *H. muticus*; CPR: cytochrome P450 reductase from *A. thaliana*

the endogenous dehydrogenase of *P. pastoris*. Further introduction of valencene synthase (ValS) and the truncated *S. cerevisiae* hydroxy-methylglutaryl CoA reductase resulted in the construction of a strain capable of de novo production of (+)-nootkatone from glucose. The production of (+)-nootkatone reached a titer of 17 mg/L in shake flasks and 208 mg/L in the bioreactors (Wriessnegger et al. 2014). Notably, overexpression of *RAD52* and the optimization of the medium have increased the production of *trans*-nootkatone by 5-fold (Wriessnegger, et al. 2016).

Polyketides are a class of secondary metabolites with various structures and biological activities. With 6-methylsalicylic acid (6-MSA) being the first polyketide produced in *P. pastoris*, the biosynthetic pathways of citrinin, monacolin, and other polyketides have also been successfully reconstituted in this host. Synthesized by a relatively small gene cluster (Fig. 4), citrinin serves as a representative for polyketide biosynthesis in *P. pastoris* (Shimizu et al. 2007; Sakai et al. 2008). A total of seven foreign genes was heterologously expressed to construct the citrinin biosynthetic pathway, including the citrinin polyketide synthase gene *PksCT* (*CitS*) from *Monascus purpureus*, the phosphoubiquitin transferase gene *NpgA* from *Aspergillus nidulans*, as well as the cluster genes, including *MPL1* (*CitA*), *MPL2* (*CitB*), *MPL4* (*CitD*), *MPL6* (*CitE*), and *MPL7* (*CitC*) from *M. purpureus*.



**Fig. 4** Reconstitution of the citrinin biosynthetic pathway in *P. pastoris*. CitS (*pksCT*): polyketide synthase; CitA (*MPL1*): serine hydrolase; CitB (*MPL2*): iron II oxidase; CitD (*MPL4*): aldehyde dehydrogenase; CitE (*MPR1*): short-chain dehydrogenase

After 24 h of cultivation with methanol as the sole carbon source, citrinin was produced up to a concentration of 0.6 mg/L (Xue et al. 2017).

#### 4 Recent Advances in Engineering *P. kudriavzeii*

*P. kudriavzeii* is a non-conventional yeast that can be found in various fermented foods (Choi et al. 2017; Vuyst et al. 2016; Qin et al. 2016), cocoa beans (Delgado-Ospina, et al. 2020), fruits (Park, et al. 2018), wastewater (Pajot et al. 2011), etc. Other names include *Issatchenkia orientalis*, *Candida glycerinogenes*, and *Candida krusei* (Douglass 2018). It is a multistress-tolerant yeast that can grow at low pH (Xiao et al. 2014; Park et al. 2018; Sun et al. 2020; Toivari et al. 2013; Hisamatsu et al. 2006), high temperature (as high as 50 °C) (Park et al. 2018; Yuangsaard et al. 2013; Chamnipa et al. 2018), and high concentrations of salt conditions (Isono et al. 2012); thus, it has been engineered to produce organic acids such as D-xylonic acid (Toivari et al. 2013), succinic acid (Xiao et al. 2014), D-lactic acid (Park et al. 2018), and itaconic acid (Sun et al. 2020). For example, Toivari et al. engineered *P. kudriavzeii* VTT C-79090T to express a D-xylose dehydrogenase gene from *Caulobacter crescentus* at the *PDC1* locus, resulting in 146 g/L D-xylonate production at pH 3.0 (Toivari et al. 2013). Xiao et al. engineered *I. orientalis* SD108 by overexpressing three native genes (i.e., encoding pyruvate carboxylase, malate dehydrogenase, and fumarase) and fumarate reductase that was previously codon-optimized for expression in *S. cerevisiae* via genome integration, enabling production of succinic acid at a titer of 11.63 g/L in shake flask (Xiao et al. 2014). In addition, by replacing the gene encoding pyruvate decarboxylase 1 with the gene encoding D-lactate dehydrogenase from *Lactobacillus plantarum* followed by adaptive evolution, the engineered *P. kudriavzeii* NG7 strain was

able to produce D-lactic acid at a titer of 135 g/L (pH 3.6) and 154 g/L (pH 4.7) (Park et al. 2018). Our group also engineered *P. kudriavzevii* YB4010 to produce 1.23 g/L itaconic acid at pH 3.9 in fed batch fermentation by overexpressing a *cis*-aconitic acid decarboxylase gene from *Aspergillus terreus* and a native mitochondrial tricarboxylate transporter in the strain with the isocitrate dehydrogenase gene deleted (Sun et al. 2020). *P. kudriavzevii* can also produce ethanol at a high salt concentration (50 g/L Na<sub>2</sub>SO<sub>4</sub>) at pH 2.0 or a temperature as high as 43 °C (Isono et al. 2012). Other applications have been demonstrated in wine fermentation (Mónaco et al. 2014, 2016), production of potential probiotics (Greppi et al. 2017), biological control (Bajaj et al. 2013), and bioremediation for heavy metal removal (Li et al. 2016b; Zhang et al. 2019b). To date, the registered *P. kudriavzevii* strains are mainly diploid (Xiao et al. 2014; Xi et al. 2021), although triploid and aneuploid strains have also been reported (Douglass 2018).

Prior to the creation of episomal plasmids, engineering works in *P. kudriavzevii* were usually performed by directly transforming a linear fragment carrying the target gene(s) and a selection marker flanked by homologous arms of the target integration site (Park et al. 2018). *URA3* is often used as a selection marker because of the relatively easy protocol for marker recycling. Considering that the strain is a diploid, a single-round integration will likely create a heterozygous strain, and a second-round integration to the wild-type allele is recommended to improve genetic stability. Recently, Tran and Cao et al. created an episomal plasmid consisting of an ARS and *LEU2* originating from *S. cerevisiae*, *P. kudriavzevii* *URA3*, and a GFP reporter gene. The percentage of GFP<sup>+</sup> cells in the culture grown from a single colony was approximately 60% after 24 h (Cao et al. 2020). They isolated a centromere-like (CEN-L) sequence from the *P. kudriavzevii* genome with the assistance of an in silico GC<sub>3</sub> analysis to identify the “GC<sub>3</sub> valley” on each chromosome, followed by sequence alignment to identify the conserved regions (Cao et al. 2020). As a CEN is responsible for faithful chromosome segregation and plays a critical role in stabilizing plasmids, the newly constructed plasmid including CEN-L led to an increased percentage of GFP<sup>+</sup> cells, to 81% after cultivation for 24 h and to 67% after cultivation for 120 h.

RNA sequencing is usually implemented to identify strong constitutive promoters and terminators. Cao et al. grew *P. kudriavzevii* under four growth conditions (YNB medium with or without lignocellulosic biomass inhibitors under aerobic or anaerobic conditions) and analyzed the transcriptome. Thirty-five promoters of the most highly expressed genes identified based on RNA-sequencing data were selected and cloned with the *GFP* reporter gene and *TEF1* terminator on an episomal plasmid containing ARS. Strong, medium, and weak constitutive promoters were categorized based on flow cytometry. For terminator identification, 14 terminators of the strong promoters identified above were selected for further comparison. Double reporter genes (i.e., *GFP* and *mCherry*) were placed between the *TDH3* promoter and the *PGK1* terminator. Each of the candidate terminator sequences was cloned between the *GFP* and the *mCherry* open reading frames (ORFs) with a random sequence or no sequence inserted between the two ORFs as controls. Quantitative PCR was used to calculate the transcriptional ratios of

*mCherry* and *GFP*. Thirteen of the 14 candidate terminators had ratios below 0.03 and were categorized as strong terminators. Moreover, similar to *S. cerevisiae*, *P. kudriavzeii* has a relatively high HR efficiency. An HR-mediated DNA assembly method was developed to facilitate rapid plasmid construction in a single step. Co-transforming five linear DNA fragments, with 70–80 bp overlaps designed between the adjacent fragments, directly into *I. orientalis* SD108 led to the successful construction of a 14.5 kb plasmid containing the xylose utilization pathway with an assembly efficiency of 100% (Cao et al. 2020).

Genome editing tools have been developed for *P. kudriavzeii*. The promoter used to transcribe the gene encoding sgRNA needs to be an RNA polymerase (RNAP) III promoter because a typical RNAP II promoter used to transcribe proteins will make the gene undergo post-transcriptional modifications such as 5'-end capping and 3'-end polyadenylation, which may inactivate the Cas9/gRNA complex (Gao and Zhao 2014). Tran and Cao et al. chose a series of RNAP III promoters including tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup>, 5S rRNA, RPR1 (the RNA component of RNase P, the 250 bp upstream sequence of RPR1), and fusions of 5S rRNA or RPR1' (the 250 bp upstream sequence of RPR1 and the first 120 bp of RPR1) with tRNA<sup>Leu</sup>. An iCas9 (containing D147Y and P411T) that possesses a higher activity than Cas9 from *Streptococcus pyogenes* was tagged with a nuclear localization sequence and expressed by an episomal plasmid, together with each of the sgRNA cassettes. Targeting *ADE2*, *LEU2*, *HIS3*, and *TRP1* in *I. orientalis* SD108 showed that RPR1'-tRNA<sup>Leu</sup> led to the highest single-gene disruption efficiency of 97–100% and was therefore used to create double and triple knockouts. Double-gene knockouts of *ADE2/TRP1* and *ADE2/HIS3* were attained with efficiencies of 72.8% and 89.9%, respectively, whereas triple-gene knockout efficiency for *ADE2/HIS3/SDH2* was approximately 47% (Tran et al. 2019). In parallel, our group also carried out a similar study by evaluating *ADE2* disruption efficiencies with five versions of Cas9 (including *S. pyogenes* Cas9, iCas9, a codon-optimized Cas9 version for *Homo sapiens*, *Candida albicans*, and *Scheffersomyces stipitis*) and RPR1 (the 311 bp upstream sequence of RPR1) as the sgRNA promoter in *P. kudriavzeii* YB4010. The highest efficiency (42%) was achieved using iCas9 (Sun et al. 2020).

Another interesting area for exploring *P. kudriavzeii* as a production host for organic acids is the identification of their transporters. Previously, our group identified a mitochondrial tricarboxylate transporter, Pk\_MttA, which can potentially transport citrate and *cis*-aconitate from the mitochondria to the cytosol, thereby increasing itaconic acid production in *P. kudriavzeii* YB4010 (Sun et al. 2020). A recent genome sequencing and transcriptome analysis of *P. kudriavzeii* CY902 led to the identification of two JEN family carboxylate transporters (PkJEN2-1 and PkJEN2-2), which can import succinate into cell (Xi et al. 2021). Substrate specificity analysis showed that both PkJEN2-1 and PkJEN2-2 are dicarboxylate importers for succinate, L-malate, and fumarate. In addition, PkJEN2-1 can import  $\alpha$ -ketoglutarate, whereas PkJEN2-2 can also uptake citrate but not  $\alpha$ -ketoglutarate. The structural basis of PkJEN2-2 specificity toward tricarboxylate substrates was studied using model-based structure analysis and rational design. By inactivating

both transporters, enhanced extracellular succinate accumulation can be achieved in the late stages of fermentation. This study highlights an important direction for future studies to improve organic acid production in *P. kudriavzevii*.

---

## 5 Conclusions and Perspectives

A panel of synthetic biology tools has been developed to establish *P. pastoris* as a microbial cell factory for the efficient production of recombinant proteins, chemicals, and natural products. Although *P. pastoris* has been widely employed for high-level expression of heterologous proteins, its application in the biosynthesis of chemicals and natural products is still limited to a few examples. On the other hand, although *P. pastoris* can utilize methanol as the sole carbon and energy source, the methanol-to-chemical conversion efficiency is still rather low, as most of the methanol is converted to CO<sub>2</sub> via the dissimilatory pathway for energy generation. Thus, the redirection of methanol flux toward the assimilatory pathway rather than the dissimilatory pathway is a prerequisite for establishing *P. pastoris* as a cell factory for chemical production from methanol. In other words, the methanol conversion pathway for the assimilation of methanol and biosynthesis of the desired product should be carefully engineered. Particularly, Lu et al. constructed a three-step synthetic acetyl-CoA (SACA) pathway for the synthesis of acetyl-CoA from formaldehyde by combining an engineered glycolaldehyde synthetase variant (GALS), acetyl phosphate synthetase (ACPS), and phosphate acetyltransferase (PTA). SACA represents the shortest pathway for acetyl-CoA biosynthesis ever reported and is promising for the efficient production of acetyl-CoA-derived compounds from C1 carbon sources (Lu et al. 2019). Considering the capability of formaldehyde formation and the role of acetyl-CoA as an important biosynthesis precursor, the SACA pathway is expected to be employed in *P. pastoris* to produce value-added compounds from methanol in the near future.

The complexity and our insufficient understanding of the cellular metabolic and regulatory networks have largely limited our capability of designing efficient *Pichia* cell factories. For example, the expression of recombinant proteins is affected by central metabolisms in *P. pastoris* (Nocon et al. 2014). One strategy is to establish genome-scale metabolic models, such as PpaMBEL1254 (Sohn et al. 2010), iPP668 (Chung et al. 2010), and iLC915 (Caspeta et al. 2012), to describe the cellular metabolic network from a systems biology perspective. Genome-scale metabolic models have been employed to guide the design of *P. pastoris* cell factories with increased expression of hSOD, human lysozyme, and antibody fragment Fab-3H6 (Cankorur-Cetinkaya et al. 2018). Unfortunately, the complexity of cellular metabolic and regulatory networks is still beyond the reach of the genome-scale metabolic models. For example, the overexpression of a non-intuitive gene, *RAD52*, encoding a protein responsible for DNA recombination was found to be beneficial for the expression of CYPs and correspondingly the production of *trans*-nootkatone (Wriessnegger, et al. 2016). Similar to that of *P. pastoris*,

a genome-scale metabolic model, iIsor850, has been developed for *P. kudriavzevii* (i.e., *I. orientalis* SD108). This model contains 850 genes, 1826 reactions, and 1702 metabolites (Suthers 2020). Biomass composition data and the estimated ATP maintenance requirements were collected by cultivating *I. orientalis* SD108 in a chemostat under carbon limitation, to improve the predictive power of the model. The consistency of the model predictions was validated using assessment of substrate utilization and gene knockouts. This model was used to propose engineering strategies for enhanced succinic acid production using the OptKnock framework. This model will be beneficial for the metabolic engineering of *P. kudriavzevii* for other organic acid production.

However, so far, only a small portion of genes is included in the genome-scale metabolic models, and the predictions are not sufficiently accurate and far from perfection. Therefore, an alternative strategy is to employ synthetic biology-based genome-scale engineering, which can perturb many genes in a combinatorial manner, extending our limited knowledge on cellular metabolism and regulation. Such a creation-driven-understanding technology aims to identify non-intuitive engineering targets, to increase the expression of recombinant proteins, as well as the production of desirable compounds. Currently, CRISPR-based genome-scale engineering tools have been developed for *E. coli* (Garst et al. 2017), *S. cerevisiae* (Lian et al. 2017, 2019), and mammalian cells (Gilbert et al. 2014), and are expected to be established in *P. pastoris* and *P. kudriavzevii* in the near future, which can be employed for the construction of efficient cell factories in a high-throughput manner.

**Acknowledgements** YZ, JG, and JL were supported by the National Key Research and Development Program of China (2018YFA0901800). WS and ZS were supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research through Ames Laboratory. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358.

## References

- Abdel-Mawgoud AM et al (2018) Metabolic engineering in the host *Yarrowia lipolytica*. *Metab Eng* 50:192–208
- Abdelmoula-Souissi S et al (2013) Secreted recombinant P53 protein from *Pichia pastoris* is a useful antigen for detection of serum p53: autoantibody in patients with advanced colorectal adenocarcinoma. *Mol Biol Rep* 40(5):3865–3872
- Agbogbo FK, Coward-Kelly G (2008) Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast *Pichia stipitis*. *Biotechnol Lett* 30(9):1515–1524
- Aggarwal S, Mishra S (2020) Differential role of segments of  $\alpha$ -mating factor secretion signal in *Pichia pastoris* towards granulocyte colony-stimulating factor emerging from a wild type or codon optimized copy of the gene. *Microb Cell Fact* 19(1):199
- Ahmad M et al (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* 98(12):5301–5317
- Ahn J et al (2016) Codon optimization of *Saccharomyces cerevisiae* mating factor alpha prepro-leader to improve recombinant protein production in *Pichia pastoris*. *Biotechnol Lett* 38(12):2137–2143
- Andreu C, Del Olmo ML (2018) Yeast arming systems: pros and cons of different protein anchors and other elements required for display. *Appl Microbiol Biotechnol* 102(6):2543–2561



- Araya-Garay JM et al (2012a) Construction of new *Pichia pastoris* X-33 strains for production of lycopene and beta-carotene. *Appl Microbiol Biotechnol* 93(6):2483–2492
- Araya-Garay JM et al (2012b) Construction of a novel *Pichia pastoris* strain for production of xanthophylls. *AMB Express* 2(1):24
- Aw R et al (2017) Expressing anti-HIV VRC01 antibody using the murine IgG1 secretion signal in *Pichia pastoris*. *AMB Express* 7(1):70
- Aza P et al (2021) Design of an improved universal signal peptide based on the  $\alpha$ -factor mating secretion signal for enzyme production in yeast. *Cell Mol Life Sci* 78(7):3691–3707
- Baeshen MN et al (2016) Expression and purification of C-peptide containing insulin using *Pichia pastoris* expression system. *Biomed Res Int* 2016:3423685
- Bajaj BK, Raina S, Singh S (2013) Killer toxin from a novel killer yeast *Pichia kudriavzevii* RY55 with idiosyncratic antibacterial activity. *J Basic Microbiol* 53(8):645–656
- Barrero JJ et al (2018) An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*. *Microb Cell Fact* 17(1):161
- Ben Azoun S et al (2016) Molecular optimization of rabies virus glycoprotein expression in *Pichia pastoris*. *Microb Biotechnol* 9(3):355–368
- Berrios J et al (2017) A comparative study of glycerol and sorbitol as co-substrates in methanol-induced cultures of *Pichia pastoris*: temperature effect and scale-up simulation. *J Ind Microbiol Biotechnol* 44(3):407–411
- Brady JR et al (2020) Comparative genome-scale analysis of *Pichia pastoris* variants informs selection of an optimal base strain. *Biotechnol Bioeng* 117(2):543–555
- Brake AJ et al (1984) Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 81(15):4642–4646
- Bredell H et al (2018) Expression of unique chimeric human papilloma virus type 16 (HPV-16) L1–L2 proteins in *Pichia pastoris* and *Hansenula polymorpha*. *Yeast* 35(9):519–529
- Cankorur-Cetinkaya A et al (2018) Process development for the continuous production of heterologous proteins by the industrial yeast *Komagataella Phaffii*. *Biotechnol Bioeng* 115(12):2962–2973
- Cao M et al (2017a) Centromeric DNA facilitates nonconventional yeast genetic engineering. *ACS Synth Biol* 6(8):1545–1553
- Cao L et al (2017b) Improving the secretion yield of the  $\beta$ -galactosidase Bgal1-3 in *Pichia pastoris* for use as a potential catalyst in the production of prebiotic-enriched milk. *J Agric Food Chem* 65(49):10757–10766
- Cao M et al (2020) A genetic toolbox for metabolic engineering of *Issatchenkia orientalis*. *Metab Eng* 59:87–97
- Carnicer M et al (2012) Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* 8(2):284–298
- Caspeta L et al (2012) Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. *BMC Syst Biol* 6:24
- Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24(1):45–66
- Cereghino GP et al (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr Opin Biotechnol* 13(4):329–332
- Chahal S et al (2017) Structural characterization of the  $\alpha$ -mating factor prepro-peptide for secretion of recombinant proteins in *Pichia pastoris*. *Gene* 598:50–62
- Chamnipa N et al (2018) The potential of the newly isolated thermotolerant yeast *Pichia kudriavzevii* RZ8-1 for high-temperature ethanol production. *Braz J Microbiol* 49(2):378–391
- Chan MK et al (2018) Expression of stable and active human DNA topoisomerase I in *Pichia pastoris*. *Protein Expr Purif* 141:52–62
- Chen Z et al (2011) Recombinant antimicrobial peptide hPAB-beta expressed in *Pichia pastoris*, a potential agent active against methicillin-resistant *Staphylococcus aureus*. *Appl Microbiol Biotechnol* 89(2):281–291
- Chen Z et al (2020) Efficient biodegradation of highly crystallized polyethylene terephthalate through cell surface display of bacterial PETase. *Sci Total Environ* 709:136138

- Cheng H et al (2014) Genetically engineered *Pichia pastoris* yeast for conversion of glucose to xylitol by a single-fermentation process. *Appl Microbiol Biotechnol* 98(8):3539–3552
- Choi S et al (2015) Biorefineries for the production of top building block chemicals and their derivatives. *Metab Eng* 28:223–239
- Choi DH, Park EH, Kim MD (2017) Isolation of thermotolerant yeast *Pichia kudriavzevii* from nuruk. *Food Sci Biotechnol* 26(5):1357–1362
- Chu J et al (2013) Progress in the research of S-adenosyl-L-methionine production. *Appl Microbiol Biotechnol* 97(1):41–49
- Chung BK et al (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb Cell Fact* 9:50
- Coughlan AY et al (2016) Centromeres of the yeast *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. *Genome Biol Evol* 8(8):2482–2492
- Cregg JM et al (1985) *Pichia pastoris* as a host system for transformations. *Mol Cell Biol* 5(12):3376–3385
- Cregg JM et al (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Mol Cell Biol* 9(3):1316–1323
- Cregg JM, Vedvick TS, Raschke WC (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology (n y)* 11(8):905–910
- Crepin VF, Faulds CB, Connerton IF (2003) Production and characterization of the *Talaromyces stipitatus* feruloyl esterase FAEC in *Pichia pastoris*: identification of the nucleophilic serine. *Protein Expr Purif* 29(2):176–184
- Curran KA, Alper HS (2012) Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metab Eng* 14(4):289–297
- Dagar VK, Khasa YP (2018) Combined effect of gene dosage and process optimization strategies on high-level production of recombinant human interleukin-3 (hIL-3) in *Pichia pastoris* fed-batch culture. *Int J Biol Macromol* 108:999–1009
- Daly R, Hearn MT (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18(2):119–138
- Damasceno LM et al (2007) Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. *Appl Microbiol Biotechnol* 74(2):381–389
- de Lima PB et al (2016) Novel homologous lactate transporter improves L-lactic acid production from glycerol in recombinant strains of *Pichia pastoris*. *Microb Cell Fact* 15(1):158
- de Oliveira JD et al (2016) Genetic basis for hyper production of hyaluronic acid in natural and engineered microorganisms. *Microb Cell Fact* 15(1):119
- De Schutter K et al (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* 27(6):561–566
- De Vuyst L et al (2016) Yeast diversity of sourdoughs and associated metabolic properties and functionalities. *Int J Food Microbiol* 239:26–34
- Del Mónaco SM et al (2014) Selection and characterization of a Patagonian *Pichia kudriavzevii* for wine deacidification. *J Appl Microbiol* 117(2):451–464
- Del Mónaco SM, Rodríguez ME, Lopes CA (2016) *Pichia kudriavzevii* as a representative yeast of North Patagonian winemaking terroir. *Int J Food Microbiol* 230:31–39
- Delgado-Ospina J et al (2020) Functional biodiversity of yeasts isolated from Colombian fermented and dry cocoa beans. *Microorganisms* 8(7)
- Deng J et al (2020) Co-expressing GroEL-GroES, Ssa1-Sis1 and Bip-PDI chaperones for enhanced intracellular production and partial-wall breaking improved stability of porcine growth hormone. *Microb Cell Fact* 19(1):35
- Dong JX et al (2013) Surface display and bioactivity of *Bombyx mori* acetylcholinesterase on *Pichia pastoris*. *PLoS One* 8(8):e70451
- Dong C et al (2020) Engineering *Pichia pastoris* with surface-display minicellulosomes for carboxymethyl cellulose hydrolysis and ethanol production. *Biotechnol Biofuels* 13:108
- Douglass AP et al (2018) Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. *PLoS Pathog* 14(7):e1007138

- Duman ZE et al (2020) High-level heterologous expression of active *Chaetomium thermophilum* FDH in *Pichia pastoris*. *Enzyme Microb Technol* 137:109552
- Eiden-Plach A et al (2004) Viral preprotoxin signal sequence allows efficient secretion of green fluorescent protein by *Candida glabrata*, *Pichia pastoris*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*. *Appl Environ Microbiol* 70(2):961–966
- Evstigneeva ZG, Solov'eva NA, Sidel'nikova LI (2001) Structure and functions of chaperones and chaperonins (Review). *Prikl Biokhim Mikrobiol* 37(1):5–18
- Fraser RZ et al (2018) Safety evaluation of soy leghemoglobin protein preparation derived from *Pichia pastoris*, intended for use as a flavor catalyst in plant-based meat. *Int J Toxicol* 37(3):241–262
- Gai SA, Witttrup KD (2007) Yeast surface display for protein engineering and characterization. *Curr Opin Struct Biol* 17(4):467–473
- Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* 56(4):343–349
- Gao L et al (2013) Engineered fungal polyketide biosynthesis in *Pichia pastoris*: a potential excellent host for polyketide production. *Microb Cell Fact* 12:77
- Garst AD et al (2017) Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol* 35(1):48–55
- Gasser B et al (2006) Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol Bioeng* 94(2):353–361
- Gasser B et al (2007) Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. *Appl Environ Microbiol* 73(20):6499–6507
- Generoso WC et al (2012) Recombinant expression and characterization of an endoglucanase III (cel12a) from *Trichoderma harzianum* (Hypocreaceae) in the yeast *Pichia pastoris*. *Genet Mol Res* 11(2):1544–1557
- Gilbert LA et al (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159(3):647–661
- Gong Z et al (2015) Efficient conversion of acetate into lipids by the oleaginous yeast *Cryptococcus curvatus*. *Biotechnol Biofuels* 8:189
- Govindappa N et al (2014) A new signal sequence for recombinant protein secretion in *Pichia pastoris*. *J Microbiol Biotechnol* 24(3):337–345
- Greppi A et al (2017) Potential probiotic *Pichia kudriavzevii* strains and their ability to enhance folate content of traditional cereal-based African fermented food. *Food Microbiol* 62:169–177
- Guo JP, Ma Y (2008) High-level expression, purification and characterization of recombinant *Aspergillus oryzae* alkaline protease in *Pichia pastoris*. *Protein Expr Purif* 58(2):301–308
- Hartner FS et al (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 36(12):e76
- Hasslacher M et al (1997) High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree *Hevea brasiliensis* in microbial hosts. *Protein Expr Purif* 11(1):61–71
- He Z et al (2012) Comparison of alpha-factor preprosequence and a classical mammalian signal peptide for secretion of recombinant xylanase xynB from yeast *Pichia pastoris*. *J Microbiol Biotechnol* 22(4):479–483
- Helian Y et al (2020) A multistrategy approach for improving the expression of *E. coli* phytase in *Pichia pastoris*. *J Ind Microbiol Biotechnol* 47(12):1161–1172
- Hetz C, Zhang K, Kaufman RJ (2020) Mechanisms, regulation and functions of the unfolded protein response. *Nat Rev Mol Cell Biol* 21(8):421–438
- Higgins DR et al (1998) Small vectors for expression based on dominant drug resistance with direct multicopy selection. *Methods Mol Biol* 103:41–53
- Hisamatsu M et al (2006) Isolation and identification of a novel yeast fermenting ethanol under acidic conditions. *J Appl Glycosci* 53(2):111–113
- Huang SJ et al (2011) Molecular cloning and characterization of a novel laccase gene from a white-rot fungus *Polyporus gramocephalus* TR16 and expression in *Pichia pastoris*. *Lett Appl Microbiol* 52(3):290–297

- Huangfu J et al (2015) Novel helper factors influencing recombinant protein production in *Pichia pastoris* based on proteomic analysis under simulated microgravity. *Appl Microbiol Biotechnol* 99(2):653–665
- Idiris A et al (2010) Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl Microbiol Biotechnol* 86(2):403–417
- Iglesias-Figueroa B et al (2016) High-Level expression of recombinant bovine lactoferrin in *Pichia pastoris* with antimicrobial activity. *Int J Mol Sci* 17(6)
- Isono N et al (2012) A comparative study of ethanol production by *Issatchenkia orientalis* strains under stress conditions. *J Biosci Bioeng* 113(1):76–78
- Ito Y et al (2020) Exchange of endogenous and heterogeneous yeast terminators in *Pichia pastoris* to tune mRNA stability and gene expression. *Nucleic Acids Res* 48(22):13000–13012
- Jacobs PP et al (2008) *Pichia* surface display: display of proteins on the surface of glycoengineered *Pichia pastoris* strains. *Biotechnol Lett* 30(12):2173–2181
- Jahic M et al (2003) Temperature limited fed-batch technique for control of proteolysis in *Pichia pastoris* bioreactor cultures. *Microb Cell Fact* 2(1):6
- Jahic M et al (2006) Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. *Biotechnol Prog* 22(6):1465–1473
- Jang IS et al (2018) Improving the efficiency of homologous recombination by chemical and biological approaches in *Yarrowia lipolytica*. *PLoS One* 13(3):e0194954
- Jariyachawalid K et al (2012) Effective enhancement of *Pseudomonas stutzeri* D-phenylglycine aminotransferase functional expression in *Pichia pastoris* by co-expressing *Escherichia coli* GroEL-GroES. *Microb Cell Fact* 11:47
- Ji Q et al (2020) Improving the homologous recombination efficiency of *Yarrowia lipolytica* by grafting heterologous component from *Saccharomyces cerevisiae*. *Metab Eng Commun* 11:e00152
- Jiang ZB et al (2007) Cell surface display of functionally active lipases from *Yarrowia lipolytica* in *Pichia pastoris*. *Protein Expr Purif* 56(1):35–39
- Jiao L et al (2018) Efficient heterologous production of *Rhizopus oryzae* lipase via optimization of multiple expression-related helper proteins. *Int J Mol Sci* 19(11)
- Jin P et al (2014) High-yield novel leech hyaluronidase to expedite the preparation of specific hyaluronan oligomers. *Sci Rep* 4:4471
- Jo JH et al (2011) Surface display of human lactoferrin using a glycosylphosphatidylinositol-anchored protein of *Saccharomyces cerevisiae* in *Pichia pastoris*. *Biotechnol Lett* 33(6):1113–1120
- Julius D et al (1984) Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. *Cell* 37(3):1075–1089
- Kang Z, Zhang N, Zhang Y (2016) Enhanced production of leech hyaluronidase by optimizing secretion and cultivation in *Pichia pastoris*. *Appl Microbiol Biotechnol* 100(2):707–717
- Karbalaeei M, Rezaee SA, Farsiani H (2020) *Pichia pastoris*: a highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol* 235(9):5867–5881
- Kato S et al (2001) Efficient expression, purification and characterization of mouse salivary alpha-amylase secreted from methylotrophic yeast *Pichia Pastoris*. *Yeast* 18(7):643–655
- Khasa YP et al (2011) Isolation of *Pichia pastoris* PIR genes and their utilization for cell surface display and recombinant protein secretion. *Yeast* 28(3):213–226
- Kittl R et al (2012) Constitutive expression of *Botrytis aclada* laccase in *Pichia pastoris*. *Bioengineered* 3(4):232–235
- Kolar NW et al (2007) Functional expression and characterisation of human cytochrome P4501alpha in *Pichia pastoris*. *J Biotechnol* 129(4):635–644
- Kong S et al (2020) De novo biosynthesis of 2-phenylethanol in engineered *Pichia pastoris*. *Enzyme Microb Technol* 133:109459
- Koutz P et al (1989) Structural comparison of the *Pichia pastoris* alcohol oxidase genes. *Yeast* 5(3):167–177
- Kuroda K, Ueda M (2011) Cell surface engineering of yeast for applications in white biotechnology. *Biotechnol Lett* 33(1):1–9

- Kuwaie S et al (2005) Production of recombinant human antithrombin by *Pichia pastoris*. *J Biosci Bioeng* 99(3):264–271
- Lee CC et al (2005) An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. *Plasmid* 54(1):80–85
- Li P et al (2007) Expression of recombinant proteins in *Pichia pastoris*. *Appl Biochem Biotechnol* 142(2):105–124
- Li W et al (2015a) Cell surface display and characterization of *Rhizopus oryzae* lipase in *Pichia pastoris* using Sed1p as an anchor protein. *Curr Microbiol* 71(1):150–155
- Li YY et al (2015b) High-level expression and characterization of a thermostable xylanase mutant from *Trichoderma reesei* in *Pichia pastoris*. *Protein Expr Purif* 108:90–96
- Li C et al (2015c) Combined strategies for improving expression of *Citrobacter amalonaticus* phytase in *Pichia pastoris*. *BMC Biotechnol* 15:88
- Li P et al (2016a) High-level secretory expression and purification of recombinant human interleukin 1 beta in *Pichia pastoris*. *Protein Pept Lett* 23(8):763–769
- Li C et al (2016b) Efficient removal of zinc by multi-stress-tolerant yeast *Pichia kudriavzevii* A16. *Bioresour Technol* 206:43–49
- Li YX et al (2018) High-level expression of an engineered beta-mannanase (mRmMan5A) in *Pichia pastoris* for manno-oligosaccharide production using steam explosion pretreated palm kernel cake. *Bioresour Technol* 256:30–37
- Li D et al (2020) Optimized expression of classical swine fever virus E2 protein via combined strategy in *Pichia pastoris*. *Protein Expr Purif* 167:105527
- Li J et al (2021) Preparation of a Bombyx mori acetylcholinesterase enzyme reagent through chaperone protein disulfide isomerase co-expression strategy in *Pichia pastoris* for detection of pesticides. *Enzyme Microb Technol* 144:109741
- Liachko I, Dunham MJ (2014) An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res* 14(2):364–367
- Lian J et al (2017) Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat Commun* 8(1):1688
- Lian J, Mishra S, Zhao H (2018) Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. *Metab Eng* 50:85–108
- Lian J et al (2019) Multi-functional genome-wide CRISPR system for high throughput genotype-phenotype mapping. *Nat Commun* 10(1):5794
- Liang S et al (2013) Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in *Pichia pastoris*. *Biotechnol Lett* 35(1):97–105
- Lin-Cereghino GP et al (2013) The effect of  $\alpha$ -mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene* 519(2):311–317
- Liu L, Maguire KK, Kmiec EB (2004) Genetic re-engineering of *Saccharomyces cerevisiae* *RAD51* leads to a significant increase in the frequency of gene repair *in vivo*. *Nucleic Acids Res* 32(7):2093–2101
- Liu SH et al (2005) Improved secretory production of glucoamylase in *Pichia pastoris* by combination of genetic manipulations. *Biochem Biophys Res Commun* 326(4):817–824
- Liu XB et al (2015) Metabolic engineering of *Pichia pastoris* for the production of dammarenediol-II. *J Biotechnol* 216:47–55
- Liu Y et al (2016) Production of glucaric acid from myo-inositol in engineered *Pichia pastoris*. *Enzyme Microb Technol* 91:8–16
- Liu Y et al (2018) Engineered monoculture and co-culture of methylotrophic yeast for *de novo* production of monacolin J and lovastatin from methanol. *Metab Eng* 45:189–199
- Louie TM et al (2021) Production of bio-xylitol from D-xylose by an engineered *Pichia pastoris* expressing a recombinant xylose reductase did not require any auxiliary substrate as electron donor. *Microb Cell Fact* 20(1):50
- Love KR et al (2016) Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genomics* 17:550
- Lu X et al (2019) Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. *Nat Commun* 10(1):1378

- Luley-Goedl C et al (2016) Combining expression and process engineering for high-quality production of human sialyltransferase in *Pichia pastoris*. *J Biotechnol* 235:54–60
- Luu VT et al (2017) Development of recombinant *Yarrowia lipolytica* producing virus-like particles of a fish nervous necrosis virus. *J Microbiol* 55(8):655–664
- Macauley-Patrick S et al (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22(4):249–270
- Majeke BM et al (2020) Synergistic codon optimization and bioreactor cultivation toward enhanced secretion of fungal lignin peroxidase in *Pichia pastoris*: enzymatic valorization of technical (industrial) lignins. *Enzyme Microb Technol* 139:109593
- Marx H et al (2009) Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Res* 9(8):1260–1270
- Massahi A, Çalık P (2015) *In-silico* determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production. *J Theor Biol* 364:179–188
- Massahi A, Çalık P (2016) Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: from *in-silico* analysis to fermentation. *J Theor Biol* 408:22–33
- Meng Y et al (2014) Production and characterization of recombinant glucose oxidase from *Aspergillus niger* expressed in *Pichia pastoris*. *Lett Appl Microbiol* 58(4):393–400
- Mombeni M et al (2020) pMOX: a new powerful promoter for recombinant protein production in yeast *Pichia pastoris*. *Enzyme Microb Technol* 139:109582
- Moura MV et al (2015) Displaying lipase B from *Candida Antarctica* in *Pichia pastoris* using the yeast surface display approach: prospection of a new anchor and characterization of the whole cell Biocatalyst. *PLoS One* 10(10):e0141454
- Nakamura Y et al (2018) A stable, autonomously replicating plasmid vector containing *Pichia pastoris* centromeric DNA. *Appl Environ Microbiol* 84(15)
- Navone L et al (2021) Disulfide bond engineering of AppA phytase for increased thermostability requires co-expression of protein disulfide isomerase in *Pichia pastoris*. *Biotechnol Biofuels* 14(1):80
- Nocon J et al (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab Eng* 24:129–138
- Nong L et al (2020) Engineering the regulatory site of the catalase promoter for improved heterologous protein production in *Pichia pastoris*. *Biotechnol Lett* 42(12):2703–2709
- Obst U, Lu TK, Sieber V (2017) A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth Biol* 6(6):1016–1025
- Ou J, Cao Y (2014) Incorporation of *Nasutitermes takasagoensis* endoglucanase into cell surface-displayed minicellulosomes in *Pichia pastoris* X33. *J Microbiol Biotechnol* 24(9):1178–1188
- Owji H et al (2018) A comprehensive review of signal peptides: structure, roles, and applications. *Eur J Cell Biol* 97(6):422–441
- Pajot HF et al (2011) Unraveling the decolourizing ability of yeast isolates from dye-polluted and virgin environments: an ecological and taxonomical overview. *Antonie Van Leeuwenhoek* 99(3):443–456
- Papakonstantinou T, Harris S, Hearn MT (2009) Expression of GFP using *Pichia pastoris* vectors with zeocin or G-418 sulphate as the primary selectable marker. *Yeast* 26(6):311–321
- Park HJ et al (2018) Low-pH production of d-lactic acid using newly isolated acid tolerant yeast *Pichia kudriavzevii* NG7. *Biotechnol Bioeng* 115(9):2232–2242
- Park HJ et al (2018) Draft genome sequence of a multistress-tolerant yeast, *Pichia kudriavzevii* NG7. *Genome Annou* 6(3)
- Pena DA et al (2018) Metabolic engineering of *Pichia pastoris*. *Metab Eng* 50:2–15
- Peng XB et al (2019) High-level secretive expression of a novel achieved *Talaromyces cellulolyticus* endo-polygalacturonase in *Pichia pastoris* by improving gene dosage for hydrolysis of natural pectin. *World J Microbiol Biotechnol* 35(6):84
- Piva LC et al (2020) Construction and characterization of centromeric plasmids for *Komagataella phaffii* using a color-based plasmid stability assay. *PLoS One* 15(7):e0235532
- Pontrelli S et al (2018) *Escherichia coli* as a host for metabolic engineering. *Metab Eng* 50:16–46
- Porro D et al (2005) Recombinant protein production in yeasts. *Mol Biotechnol* 31(3):245–259

- Prielhofer R et al (2017) GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast *Pichia pastoris*. *BMC Syst Biol* 11(1):123
- Qin X et al (2011) GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. *Appl Environ Microbiol* 77(11):3600–3608
- Qin H et al (2016) Microbial diversity and biochemical analysis of suanzhou: a traditional Chinese fermented cereal gruel. *Front Microbiol* 7:1311
- Robert C et al (2013) Recombinant proteins for industrial uses: utilization of *Pichia pastoris* expression system. *Braz J Microbiol* 44(2):351–356
- Raemaekers RJ et al (1999) Functional phytohemagglutinin (PHA) and *Galanthus nivalis* agglutinin (GNA) expressed in *Pichia pastoris* correct N-terminal processing and secretion of heterologous proteins expressed using the PHA-E signal peptide. *Eur J Biochem* 265(1):394–403
- Rajamanickam V et al (2017) A novel bi-directional promoter system allows tunable recombinant protein production in *Pichia pastoris*. *Microb Cell Fact* 16(1):152
- Rakestraw JA et al (2009) Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 103(6):1192–1201
- Ramakrishnan K et al (2020) Transcriptional control of gene expression in *Pichia pastoris* by manipulation of terminators. *Appl Microbiol Biotechnol* 104(18):7841–7851
- Römisch K (2005) Endoplasmic reticulum-associated degradation. *Annu Rev Cell Dev Biol* 21:435–456
- Saitua F et al (2017) Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. *BMC Syst Biol* 11(1):27
- Sakai K et al (2008) Construction of a citrinin gene cluster expression system in heterologous *Aspergillus oryzae*. *J Biosci Bioeng* 106(5):466–472
- Sallada ND, Harkins LE, Berger BW (2019) Effect of gene copy number and chaperone coexpression on recombinant hydrophobin HFBI biosurfactant production in *Pichia pastoris*. *Biotechnol Bioeng* 116(8):2029–2040
- Sams L et al (2017) Constitutive expression of human gastric lipase in *Pichia pastoris* and site-directed mutagenesis of key lid-stabilizing residues. *Biochim Biophys Acta Mol Cell Biol Lipids* 1862(10 Pt A):1025–1034
- Sanchooli A et al (2018) VLP production from recombinant L1/L2 HPV-16 protein expressed in *Pichia Pastoris*. *Protein Pept Lett* 25(8):783–790
- Schwarzans JP et al (2016) Non-canonical integration events in *Pichia pastoris* encountered during standard transformation analysed with genome sequencing. *Sci Rep* 6:38952
- Schwarzans JP et al (2017) A mitochondrial autonomously replicating sequence from *Pichia pastoris* for uniform high level recombinant protein production. *Front Microbiol* 8:780
- Sha C et al (2013) Enhancement of lipase r27RCL production in *Pichia pastoris* by regulating gene dosage and co-expression with chaperone protein disulfide isomerase. *Enzyme Microb Technol* 53(6–7):438–443
- Shalgi R et al (2005) A catalog of stability-associated sequence elements in 3' UTRs of yeast mRNAs. *Genome Biol* 6(10):R86
- Shen Q et al (2012) The effect of gene copy number and co-expression of chaperone on production of albumin fusion proteins in *Pichia pastoris*. *Appl Microbiol Biotechnol* 96(3):763–772
- Shimizu T, Kinoshita H, Nihira T (2007) Identification and in vivo functional analysis by gene disruption of *ctnA*, an activator gene involved in citrinin biosynthesis in *Monascus purpureus*. *Appl Environ Microbiol* 73(16):5097–5103
- Silva AJD et al (2021) *Pichia pastoris* displaying ZIKV protein epitopes from the envelope and NS1 induce in vitro immune activation. *Vaccine* 39(18):2545–2554
- Sohn SB et al (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. *Biotechnol J* 5(7):705–715
- Song W et al (2020) Multiple strategies to improve the yield of chitinase a from *Bacillus licheniformis* in *Pichia pastoris* to obtain plant growth enhancer and GlcNAc. *Microb Cell Fact* 19(1):181

- Stadlmayr G et al (2010) Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*. *Biotechnol Bioeng* 105(3):543–555
- Sturmberger L et al (2016) Refined *Pichia pastoris* reference genome sequence. *J Biotechnol* 235:121–131
- Su GD et al (2010a) Display of *Candida antarctica* lipase B on *Pichia pastoris* and its application to flavor ester synthesis. *Appl Microbiol Biotechnol* 86(5):1493–1501
- Su GD, Zhang X, Lin Y (2010b) Surface display of active lipase in *Pichia pastoris* using Sed1 as an anchor protein. *Biotechnol Lett* 32(8):1131–1136
- Su X et al (2017) High-level expression and purification of a molluscan endoglucanase from *Ampullaria crosseana* in *Pichia pastoris*. *Protein Expr Purif* 139:8–13
- Sumppunn P, Jomrit J, Panbangred W (2018) Improvement of extracellular bacterial protein production in *Pichia pastoris* by co-expression of endoplasmic reticulum residing GroEL-GroES. *J Biosci Bioeng* 125(3):268–274
- Sun W et al (2016) High level expression and purification of active recombinant human interleukin-15 in *Pichia pastoris*. *J Immunol Methods* 428:50–57
- Sun X et al (2019) Construction of a novel MK-4 biosynthetic pathway in *Pichia pastoris* through heterologous expression of HsUBIAD1. *Microb Cell Fact* 18(1):169
- Sun W et al (2020) Metabolic engineering of an acid-tolerant yeast strain *Pichia kudriavzevii* for itaconic acid production. *Metab Eng Commun* 10:e00124
- Suthers PF et al (2020) Genome-scale metabolic reconstruction of the non-model yeast *Issatchenkia orientalis* SD108 and its application to organic acids production. *Metab Eng Commun* 11:e00148
- Tanaka T et al (2012) Recent developments in yeast cell surface display toward extended applications in biotechnology. *Appl Microbiol Biotechnol* 95(3):577–591
- Toivari M et al (2013) Low pH D-xylonate production with *Pichia kudriavzevii*. *Bioresour Technol* 133:555–562
- Torres P et al (2019) Contextualized genome-scale model unveils high-order metabolic effects of the specific growth rate and oxygenation level in recombinant *Pichia pastoris*. *Metab Eng Commun* 9:e00103
- Tran VG et al (2019) Development of a CRISPR/Cas9-based tool for gene deletion in *Issatchenkia orientalis*. *mSphere* 4(3)
- Tu Y et al (2016) Extracellular expression and antiviral activity of a bovine interferon-alpha through codon optimization in *Pichia pastoris*. *Microbiol Res* 191:12–18
- Turkanoglu Ozcelik A, Yilmaz S, Inan M (2019) *Pichia pastoris* promoters. *Methods Mol Biol* 1923:97–112
- Ueda M (2016) Establishment of cell surface engineering and its development. *Biosci Biotechnol Biochem* 80(7):1243–1253
- Vadhana AK et al (2013) Improved secretion of *Candida antarctica* lipase B with its native signal peptide in *Pichia pastoris*. *Enzyme Microb Technol* 52(3):177–183
- Vallet-Courbin A et al (2017) A Recombinant human anti-platelet scFv antibody produced in *Pichia pastoris* for atheroma targeting. *PLoS One* 12(1):e0170305
- van der Heide M et al (2002) Overproduction of BiP negatively affects the secretion of *Aspergillus niger* glucose oxidase by the yeast *Hansenula polymorpha*. *Appl Microbiol Biotechnol* 58(4):487–494
- Varela JA et al (2017) Applications of *Kluyveromyces marxianus* in biotechnology. In: Satyanarayana T, Kunze G (eds) *Yeast diversity in human welfare*. Springer Singapore, Singapore, pp 439–453
- Veana F et al (2014) Gene encoding a novel invertase from a xerophilic *Aspergillus niger* strain and production of the enzyme in *Pichia pastoris*. *Enzyme Microb Technol* 63:28–33
- Vogl T, Glieder A (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. *N Biotechnol* 30(4):385–404



- Vogl T et al (2016) A toolbox of diverse promoters related to methanol utilization: functionally verified parts for heterologous pathway expression in *Pichia pastoris*. *ACS Synth Biol* 5(2):172–186
- Wang Q et al (2007) Construction of a novel system for cell surface display of heterologous proteins on *Pichia pastoris*. *Biotechnol Lett* 29(10):1561–1566
- Wang X et al (2012) Constitutive expression of *Yarrowia lipolytica* lipase LIP2 in *Pichia pastoris* using GAP as promoter. *Appl Biochem Biotechnol* 166(5):1355–1367
- Wang M, Jiang S, Wang Y (2016) Recent advances in the production of recombinant subunit vaccines in *Pichia pastoris*. *Bioengineered* 7(3):155–165
- Wang P et al (2017) Accurate analysis of fusion expression of *Pichia pastoris* glycosylphosphatidylinositol-modified cell wall proteins. *J Ind Microbiol Biotechnol* 44(9):1355–1365
- Wang L et al (2018) Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. *Biotechnol Biofuels* 11:277
- Waterham HR et al (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186(1):37–44
- Wegner GH, Harder W (1987) Methylophilic yeasts-1986. *Antonie Van Leeuwenhoek* 53(1):29–36
- Weinacker D et al (2013) Applications of recombinant *Pichia pastoris* in the healthcare industry. *Braz J Microbiol* 44(4):1043–1048
- Weninger A et al (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylophilic yeast *Pichia pastoris*. *J Biotechnol* 235:139–149
- Weninger A et al (2018) Expanding the CRISPR/Cas9 toolkit for *Pichia pastoris* with efficient donor integration and alternative resistance markers. *J Cell Biochem* 119(4):3183–3198
- Werten MW et al (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast* 15(11):1087–1096
- Werten MWT et al (2019) Production of protein-based polymers in *Pichia pastoris*. *Biotechnol Adv* 37(5):642–666
- Wilkinson B, Gilbert HF (2004) Protein disulfide isomerase. *Biochim Biophys Acta* 1699(1–2):35–44
- Wriessnegger T et al (2014) Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris*. *Metab Eng* 24:18–29
- Wriessnegger T et al (2016) Enhancing cytochrome P450-mediated conversions in *P. pastoris* through RAD52 over-expression and optimizing the cultivation conditions. *Fungal Genet Biol* 89:114–125
- Xi Y et al (2021) Characterization of JEN family carboxylate transporters from the acid-tolerant yeast *Pichia kudriavzevii* and their applications in succinic acid production. *Microb Biotechnol* 14(3):1130–1147
- Xiao H et al (2014) Exploiting *Issatchenkia orientalis* SD108 for succinic acid production. *Microb Cell Fact* 13:121
- Xiong R, Chen J (2008) Secreted expression of human lysozyme in the yeast *Pichia pastoris* under the direction of the signal peptide from human serum albumin. *Biotechnol Appl Biochem* 51(Pt 3):129–134
- Xiong AS et al (2005) High level expression of a recombinant acid phytase gene in *Pichia pastoris*. *J Appl Microbiol* 98(2):418–428
- Xue Y et al (2017) Methylophilic yeast *Pichia pastoris* as a chassis organism for polyketide synthesis via the full citrinin biosynthetic pathway. *J Biotechnol* 242:64–72
- Yaegashi J et al (2017) *Rhodospiridium toruloides*: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnol Biofuels* 10:241
- Yaguchi A et al (2017) Metabolism of aromatics by *Trichosporon oleaginosus* while remaining oleaginous. *Microb Cell Fact* 16(1):206
- Yamada R et al (2019) Toward the construction of a technology platform for chemicals production from methanol: D-lactic acid production from methanol by an engineered yeast *Pichia pastoris*. *World J Microbiol Biotechnol* 35(2):37

- Yang S et al (2017) Cell-surface displayed expression of trehalose synthase from *Pseudomonas putida* ATCC 47054 in *Pichia pastoris* using Pir1p as an anchor protein. *Front Microbiol* 8:2583
- Yang Z et al (2020) Development and characterization of an enterovirus 71 (EV71) virus-like particles (VLPs) vaccine produced in *Pichia pastoris*. *Hum Vaccin Immunother* 16(7):1602–1610
- YaPing W et al (2017) High-level expression of l-glutamate oxidase in *Pichia pastoris* using multi-copy expression strains and high cell density cultivation. *Protein Expr Purif* 129:108–114
- Ye R et al (2017) Comprehensive reconstruction and evaluation of *Pichia pastoris* genome-scale metabolic model that accounts for 1243 ORFs. *Bioresour Bioprocess* 4(1):22
- Yin C et al (2014) High-level expression of a manganese superoxide dismutase (PoMn-SOD) from *Pleurotus ostreatus* in *Pichia pastoris*. *Appl Biochem Biotechnol* 174(1):259–269
- Yu X et al (2007) Expression and purification of ancrod, an anticoagulant drug *Pichia Pastoris*. *Protein Expr Purif* 55(2):257–261
- Yu P et al (2015) Improving the secretory production of the heterologous protein in *Pichia pastoris* by focusing on protein folding. *Appl Biochem Biotechnol* 175(1):535–548
- Yu KM et al (2018) Efficient expression and isolation of recombinant human interleukin-11 (rhIL-11) in *Pichia pastoris*. *Protein Expr Purif* 146:69–77
- Yuangsaard N et al (2013) Selection and characterization of a newly isolated thermotolerant *Pichia kudriavzevii* strain for ethanol production at high temperature from cassava starch hydrolysate. *Antonie Van Leeuwenhoek* 103(3):577–588
- Zhang W et al (2006) Enhanced secretion of heterologous proteins in *Pichia pastoris* following overexpression of *Saccharomyces cerevisiae* chaperone proteins. *Biotechnol Prog* 22(4):1090–1095
- Zhang L et al (2013) Screening for glycosylphosphatidylinositol-modified cell wall proteins in *Pichia pastoris* and their recombinant expression on the cell surface. *Appl Environ Microbiol* 79(18):5519–5526
- Zhang XF et al (2019a) High-level expression of *Aspergillus niger* lipase in *Pichia pastoris*: characterization and gastric digestion in vitro. *Food Chem* 274:305–313
- Zhang D et al (2019b) Improved cadmium resistance and removal capacity in *Pichia kudriavzevii* A16 by sucrose preincubation. *J Basic Microbiol* 59(9):867–878
- Zheng JY et al (2019) High-level expression and characterization of a stereoselective lipase from *Aspergillus oryzae* in *Pichia pastoris*. *Protein Expr Purif* 155:1–7
- Zhong X et al (2004) Secretion, purification, and characterization of a recombinant *Aspergillus oryzae* tannase in *Pichia pastoris*. *Protein Expr Purif* 36(2):165–169
- Zhu T et al (2019) *Pichia pastoris* as a versatile cell factory for the production of industrial enzymes and chemicals: current status and future perspectives. *Biotechnol J* 14(6):e1800694
- Zirpel B et al (2018) Optimization of  $\Delta(9)$ -tetrahydrocannabinolic acid synthase production in *Komagataella phaffii* via post-translational bottleneck identification. *J Biotechnol* 272–273:40–47