Chapter 3

Preparation of expression plasmids for *Pichia pastoris*

Christel Logez¹, Fatima Alkhalfioui¹, Bernadette Byrne², Renaud Wagner^{1†}

¹Ecole Supérieure de Biotechnologie de Strasbourg – Centre National de la Recherche Scientifique, Département Récepteurs et Protéines Membranaires, 67412 Illkirch, France

²Division of Molecular Biosciences, Imperial College London, Exhibition Road, London, SW7 2AZ

[†]To whom correspondence should be addressed:

Dr. Renaud Wagner

Phone: +33 368 85 4731

Fax: +33 368 85 4829

E-mail: renaud.wagner@unistra.fr

Abstract

When planning any heterologous expression experiment, the very first critical step is related to the design of the overall strategy, hence to the selection of the most adapted expression vector. The very flexible *Pichia pastoris* system offers a broad range of possibilities for the expression of secreted, endogenous or membrane proteins thanks to a combination of various plasmid backbones, selection markers, promoters and fusion sequences introduced into dedicated host strains. The present chapter provides some guidelines on the choice of expression vectors and expression strategies. It also brings the reader a complete toolbox from which plasmids and fusion sequences can be picked and assembled to set up appropriate expression vectors. Finally, it provides standard starting protocols for the preparation of the selected plasmids and their use for host strain transformation.

Keywords: plasmid, expression, purification/detection tag, promoter, *Pichia pastoris* cell strains.

1. Introduction

Hundreds of proteins of various types, origins and functions have been produced in *Pichia pastoris* for many purposes and applications. Conveniently, a large set of representative examples have been listed in authoritative reviews and can advantageously serve the reader as points of reference (1-4). In these numerous studies, high yield expression is very often dependent on several parameters including the choice of the expression vector, the optimal gene sequence, the nature and site of insertion of any fusion tags, and the transformation and selection strategies. Thus, while no standards exist to predict which combination will enable the successful expression of a given protein, we instead propose the following steps that may help in determining the appropriate tools and methods to start with, and where to find them.

1.1. What plasmid to select?

Except for a limited number of autoreplicative plasmids that are not yet frequently employed (5-8), the usual expression vectors are designed to be maintained as integrative elements in the genome of *P. pastoris* (see section 1.6. below). They are built on a classical *E. coli*/yeast shuttle model with components required for *E. coli* amplification (classically one origin of replication and one antibiotic selection marker) and specific elements for heterologous gene expression in *P. pastoris*. These typically include selectable auxotrophy markers and/or antibiotic resistance bacterial genes, a range of promoter and terminator sequences, a cloning cassette and supplementary fusion sequences that can be added for improving the secretion, detection and purification of the expressed proteins. Each is schematically represented as a building block in **Figure 1** and these are further discussed in the following sections.

(Insert Figure 1 here)

1.2. What promoter to use?

P. pastoris harbours several strong or weaker promoters that can be exploited to drive heterologous expression of recombinant genes, both in an inducible or constitutive fashion (see **Table 1**). Inducible expression is usually the preferred strategy since it allows a convenient control of the experimental conditions applied before expression and is ideally adapted for the production of proteins that are toxic to the host. *P. pastoris* offers a panel of promoters that can be induced in the presence of various carbon or nitrogen sources (**9**), the promoter P_{AOXI} from the alcohol oxidase encoding gene (*AOXI*) being predominantly employed. This promoter is tightly repressed by glucose and strongly induced by methanol (**10**) allowing the cells to use methanol as the sole carbon source. A P_{AOXI} synthetic promoter library was recently developed revealing enhanced P_{AOXI} variants that resulted in high expression levels of a recombinant GFP (**11**). There are numerous cases, however, where constitutive expression performs as well as inducible expression, in particular when using the strong glyceraldehyde-3-phosphate dehydrogenase P_{GAP} promoter (**9, 12**). In addition, constitutive expression is more straightforward to manage since no switch of carbon source is required, which is particularly convenient when running fermentation procedures.

(Insert Table 1 here)

1.3. Do I need a secretion signal?

This is a non-trivial question since the choice of intracellular or extracellular localisation can have a direct impact both on yield and integrity of the recombinant protein, as well as on the procedures required for isolation. Secreting the recombinant proteins outside the cell has several advantages: soluble protein expression may be induced for longer periods of time since they are not accumulating in the limited volume of the cytoplasm where they might become toxic for the host. This can lead to an increased expression yield. Furthermore, no cell lysis step is required and secreted proteins can be recovered directly from the culture medium, which contains far fewer contaminating proteins than the cells therefore simplifying the purification process. One limitation is the frequent degradation of the secreted proteins by extracellular proteases and proteases released from lysed cells. In addition, proteins that are not naturally secreted may not be properly folded outside the cell. In this regard, intracellular expression is a valuable alternative (**13, 14**).

When opting for a secretion strategy, the target protein needs to be identified as secreted by the presence of a signal sequence (see **Table 2**). Successful secretion of many proteins from *P. pastoris* has been reported using a range of different signal sequences. These include a protein's native secretion signal, the *Saccharomyces cerevisiae* α -mating factor prepro leader sequence (α -MF), the *P. pastoris* acid phosphatase (*PHO1*) signal sequence and the invertase (*SUC2*) signal sequence (see (**1**) for an extensive list). Further information on the range and use of prepro peptides can be found in (**15**) and this may provide a useful resource for selecting a signal sequence.

(Insert Table 2 here)

In the case of integral membrane proteins, adding a secretion sequence may be highly beneficial for expression in *P. pastoris*. Such an approach has proved highly effective for the production of GPCRs (**16, 17**). However the presence of the signal sequence has variable effects on the expression of other MPs. In the case of aquaporins for instance, where the N and C termini are both located intracellularly, protein expression has been evaluated with or without a fused signal sequence. In both cases high yields of high quality protein suitable for structural studies were obtained (**18, 19**).

1.4. What kind of additional sequences do I need?

Whatever the objectives to be achieved when producing a protein with *P. pastoris* (biochemical and/or biophysical characterization, structural studies, pharmaceutical or food production), recovering the protein in its most native form is generally mandatory. There are many examples of expression and subsequent isolation of untagged proteins requiring development of specific and often tedious purification procedures. Alternatively, adding epitope tags allows detection, and isolation of the target protein using generic techniques and tools. An ideal tag should not only (i) exert a minimal effect on the tertiary structure and the biological activity of the protein it is fused to, but should also (ii) allow a one-step adsorption purification, (iii) be easily and specifically removed to produce the native protein and (iv) be applicable to a number of different proteins. While it is difficult to decide on the best fusion sequence and position to be used for a specific protein, we present in **Table 3** a list of tags and protease cleavage sites to release them that have proven helpful for the production of proteins in *P. pastoris* (**20**).

(Insert Table 3 here)

1.5. How can I optimize the sequence of my protein-encoding gene?

Even if recombinant genes are most often cloned and expressed in their native form, several adjustments in their sequence can be made to best fit the transcription and translation machineries of the yeast and very often result in dramatic improvement of the protein yields. The sequence parameters that were notably shown to positively influence the expression levels include (i) an optimal translation initiation sequence (the yeast consensus is A/YAA/UAAUGUCU), (ii) an adaptation to the codon usage of yeasts (**21, 22**), (iii) an increase of the GC-content (**22, 23**), (iv) a decreased occurrence of AT-rich regions (**24**) and (v) an adapted isoelectric point of the protein (**24**). With the very recent release of the whole genome sequence of *P. pastoris* (**25**), even more accurate sequence optimizations are now possible.

Answering this series of questions should help the researcher to assemble the most suitable vector(s) for expression of their target protein in *P. pastoris*. A significant set of plasmids is commercially available from Invitrogen (see the Protein and Expression section in <http://www.invitrogen.com> and **Table 4**) that could either be used as is, or further engineered to best suit the selected expression strategy.

(Insert Table 4 here)

1.6. My construct is now ready, how do I proceed with *P. pastoris* transformation?

As for many other yeasts, transformation of *P. pastoris* is straightforward. Several robust methods are available, either based on chemically competent (spheroplasts, PEG1000, LiCl) or electrocompetent cells. Moreover, these protocols are well described and can be easily found on numerous websites: convenient *Pichia* manuals can be downloaded from invitrogen.com.

The number of strains usually employed for heterologous expression is rather limited (see **Table 5**). They mainly differ in their auxotrophies, principally relying on a histidinol dehydrogenase deficiency (*his4*), allowing, upon transformation, for the positive selection of recombinant expression vectors. Some of them bear additional deficiencies in endogeneous proteases (SMD series), other were recently engineered for their capacity in performing “human-like” N-glycosylations (**26**).

(Insert Table 5 here)

As already mentioned in the first section, most of the transforming expression vectors are designed to be maintained as integrative elements in the genome of *P. pastoris*. This is generally achieved through recombination events between linearized sequences borne by the plasmids (typically *HIS4* or P_{AOXI}) and their homologous sequence counterparts present on the genome, leading to the targeted insertion of the expression vectors. Moreover, such plasmid insertions frequently occur in tandem in yeasts and thus lead to multiple integration of the gene of interest with an associated impact on subsequent expression levels.

Alternatively, integration can be obtained by a gene replacement strategy. In this case, a double recombination event must be performed between the *AOXI* promoter and terminator sequences present on the transforming DNA (containing the gene of interest and a selection marker) and the corresponding homologous sequences present within the *P. pastoris* genome. This double recombination event results in the replacement of the *AOXI* gene by the construct of interest.

The phenotype of the resulting transformants is not solely dependent on the selection marker present in the chosen vector (auxotrophy and/or antibiotic resistance). The integration strategy itself dictates the methanol utilization phenotype of the transformed cells since plasmid insertion results in a Mut⁺ (methanol utilization plus) phenotype, while the gene replacement of *AOXI* leads to a Mut^s (methanol utilization slow) phenotype. In several cases, these differences in methanol utilization have been reported as an important parameter to consider for enhancing the performance of recombinant protein expression (27).

1.7. Where can I find a practical illustration of the construction and preparation of an expression vector?

The next sections present the material and protocols needed for the cloning of the gene encoding the adenosine A_{2A} receptor (AA2AR), a G protein-coupled receptor (GPCR), into an engineered pPIC9K plasmid (see **Table 1**). This vector was modified by standard molecular biology procedures to incorporate a Flag-tag, a TEV protease cleavage sequence and a deca histidine-tag (10His) upstream of the the *Bam*HI and *Spe*I cloning sites for insertion of the target gene, as well as a second TEV site and a Biotinylation-tag downstream (28). This combination was selected on the basis of

previous studies showing enhanced expression levels of other GPCRs when fused to the α -MF signal sequence (present on pPIC9K) and the biotinylation-tag (**16, 17**). The Flag and 10His tag were inserted for detection and purification purposes, the TEV cleavage sites were added to allow cleavage of the N- and C-terminally fused sequences following purification. The brief protocols presented here illustrate a very standard way of generating the desired *P. pastoris* expression plasmid as well as preparation prior to yeast transformation.

2. Materials

2.1. Cloning the gene of interest into the expression vector

1. A cDNA template containing the full-length AA2AR_HUMAN encoding gene.
2. A 30 base-long AA2A specific forward primer bearing an additional 5' adapter specifically designed to introduce a **BamHI** restriction site (fwd. sequence: 5'-GAAGACAGGATCCATGCCCATCATGGGCTCCTCGGTGTACATC-3'), and a similar reverse primer, bearing a 5' adapter introducing **SpeI** (rev. sequence: 5'-GAAGACAAGTGGACACTCCTGCTCCATCCTGGGCCAGGGG-3') (see **Note 1**).
3. A high-fidelity polymerase, typically the PrimeSTAR (Takara) or the Phusion (Finnzyme) DNA polymerase, and its specific buffer and dNTP mix.
4. Standard restriction enzymes and their related buffers, here *BamHI* and *SpeI* (Fermentas, Germany).
5. A T4 DNA ligase, here the Rapid DNA ligation kit (Fermentas).
6. *E. coli* competent cells, here the TOP10 chemically competent cells (Invitrogen).
7. Liquid and agar plates of LB media (1 % tryptone, 0.5 % yeast extract, 1 % NaCl) supplemented with 50 μ g/ml kanamycin

8. A robust nucleic acid extraction and purification kit, here the NucleoSpin kit (Macherey-Nagel).
9. Standard equipment, consumables and chemicals for routine molecular biology techniques including PCR amplification of DNA fragments, DNA separation and visualization, UV spectrophotometry and *E. coli* culturing.

2.2. Preparation of the expression vector

1. Liquid LB medium.
2. NucleoSpin Plasmid kit from Macherey-Nagel.
3. Restriction enzyme *PmeI* and its specific buffer (Fermentas).
4. Phenol.
5. 24:1 (v/v) chloroform-isoamyl alcohol.
6. Ice-cold 100% ethanol.
7. Ice-cold 70% ethanol.
8. 3M sodium acetate pH 4.8.
9. Sterile H₂O.
10. Agarose gels (1%) supplemented with ethidium bromide.

2.3. Transformation of *Pichia pastoris*

All materials and solutions must be sterile.

1. YPD rich medium: 1 % yeast extract, 2 % peptone, 2 % dextrose.
2. Agar plates made with YPD rich medium supplemented with 2% agar.
3. A fresh SMD1163 colony streaked on a YPD plate.
4. 1 M Hepes pH 8
5. 1 M dithiothreitol (DTT)

6. 1 M cold sorbitol
7. Sterile cold H₂O.
8. Electroporation instrument and sterile 0.2 cm electroporation cuvettes.
9. MD plates: 1.34 % Yeast Nitrogen Base w/o amino acids, 2 % dextrose, 4×10^{-5} % biotin.
10. YPD plates supplemented with 0.1 and 0.25 mg/ml geneticin.

3. Methods

3.1. Cloning the AA2AR gene into the modified pPIC9K expression vector

3.1.1. PCR amplification and preparation of the AA2AR gene

1. Prepare the PCR reaction mix on ice: typically 1 to 10 ng of the template cDNA, 5 µl each of the 2 µM stock solution of the forward and reverse primers, 10 µl of 5 X PCR buffer, 4 µl of a dNTP mixture (2.5 mM each), 1 U of high fidelity PrimeSTAR polymerase, and sterile water to a final volume of 50 µl.
2. Run the PCR reaction in a thermocycler with a standard 30 cycle protocol alternating 15 sec at 98 °C, 15 sec at 55 °C and 1 min at 72 °C.
3. Pipet 25 µl of the PCR reaction, add 5 µl of 6 X loading dye and load the mixture on a 1 % agarose gel to analyze the amplified product after migration.
4. Extract the desired DNA fragment using the protocol detailed in the NucleoSpin kit (see **Note 2**).

3.1.2. Preparation of the plasmid and ligation with the insert DNA

1. In individual eppendorf tubes, cut the amplified insert fragment coding for the gene of interest and the pPIC9K vector with the *Bam*HI and *Spe*I enzymes.

2. Load the digestion products on a 1 % agarose gel and following separation extract the cut insert and vector fragments separately (see **Note 3**).
3. Prepare the ligation reaction with a 5: 1 ratio of insert : plasmid. Typically 50-100 mg of linearized plasmid is used. Add 1 U of T4 DNA ligase together with the ligase buffer and make the final volume to 20 µl with sterile water.

3.1.3. Transformation, selection and control of *E. coli* recombinant clones

1. Use about 5 µl of the ligation mixture to transform 50 µl of TOP10 chemically competent cells. Incubate on ice for 5 to 30 min.
2. Heat-shock the cells for 30 s at 42 °C and immediately transfer the tubes on ice.
3. Add 250 µl of regeneration medium (typically SOB or SOC medium) and let the cells regenerate for 1 h at 37 °C.
4. Spread 100 to 200 µl of the transformation mixture on prewarmed LB agar plates supplemented with 50 µg/ml kanamycin and incubate overnight at 37 °C.
5. The following day, pick 6 to 12 colonies and use to inoculate 2 ml LB supplemented with 50 µg/ml kanamycin. Grow the cultures overnight at 37 °C in an incubator shaker. The presence of the insert in a particular clone can also be checked using colony PCR.
6. Purify the plasmid DNA of each clone using the plasmid purification kit following the manufacturer's instructions.
7. Perform restriction digest analysis of the plasmids using the *Bam*HI and *Spe*I enzymes to confirm the presence of the insert.
8. Final check the integrity of the insert by DNA sequencing (see **Note 4**).

9. Store the plasmid containing the correct sequence at - 20 °C. In addition prepare a glycerol stock by adding 700 µl of culture containing correct clone to 300 µl 20 % glycerol LB medium in cryotubes and storing at - 80 °C.

3.2. Preparation of the expression vector

3.2.1. Amplification and linearization of the expression vector

1. Inoculate 5-10 ml LB with the *E. coli* clone containing the expression pPIC9K plasmid and incubate at 37°C overnight.
2. Extract and purify the plasmid DNA using the plasmid preparation kit (Macherey-Nagel) according to the manufacturer's instructions.
3. Prepare a restriction digest solution by adding 5 to 7 µg of purified plasmid to 25 U of *PmeI*, 20 µl of 10 X corresponding buffer and sterile water to a final volume of 200 µl. Incubate the reaction for 2 h at 37 °C (see **Note 5**).

3.2.2. Phenol-chloroform extraction of the linearized plasmid

1. Add 400 µl of phenol:chloroform (1:1) to the 200 µl digestion mixture.
2. Centrifuge 5 min at 18,000 g and transfer the superior phase to a new tube.
3. Add 400 µl of chloroform and vortex thoroughly.
4. Centrifuge 5 min at 18,000 g and transfer the superior phase to a new tube.
5. Add 1 ml of 100 % ethanol, 50 µl of 3 M sodium acetate and incubate for at least 1 h at -20 °C to precipitate the DNA.
6. Centrifuge for 30 min at 4°C at 18,000 g.
7. Wash the pellet with 100 µl of 70 % ethanol.
8. Centrifuge for 5 min at 4°C at 18,000 g.
9. Dry the pellet then resuspend in 15 µl sterile H₂O (see **Note 6**).

10. Check the DNA linearization by loading 1 µl of the solution on a 1% agarose gel.

3.3. Transformation of *P. pastoris*

3.3.1. Preparation of competent cells (see Note 7)

1. Inoculate 100 ml of YPG medium ~~supplemented with??[rm2]~~ with a fresh SMD1163 colony and incubate the preculture overnight at 30°C in an incubator shaker.
2. Measure the OD_{600nm} of the preculture, ~~and~~ dilute it in 400 ml YPG medium in order to ~~reach-obtain~~ an OD_{600nm} of ~~1-0.25 and incubate after 2 generations (about 4 h)~~ at 30 °C. ~~Might want to explain this a little more in particular what the doubling time of the culture is - some kind of estimate of the OD the culture should be diluted to would also be good.[rm3]~~
3. When the culture reaches an OD_{600nm} of 1 ~~(approximately after 4 h)~~, harvest the cells by centrifugation in sterile tubes at 2,000 g and 4°C for 5 min.
4. Resuspend the cells in 100 ml YPD, 20 ml of 1 M Hepes pH 8 and 2.5 ml of 1 M DTT; mix gently until the pellet is resuspended.
5. Incubate for 15 min at 30°C.
6. Transfer onto ice and add sterile and cold H₂O to a final volume of 500 ml.
7. Harvest the cells by centrifugation at 2,000 g and 4°C for 5 min.
8. Wash the cell pellet with 250 ml of sterile and cold H₂O.
9. Harvest the cells by centrifugation at 2,000 g and 4°C for 5 min.
10. Resuspend the pellet in 20 ml of cold 1 M sorbitol by gently mixing.
11. Harvest the cells by centrifugation at 2,000 g and 4°C for 5 min.
12. Resuspend the pellet in 500 µl of cold 1 M sorbitol by gently mixing.

3.3.2. Electrotransformation

Preparation of expression plasmids for *Pichia pastoris*

1. Place an electroporation cuvette on ice at least 10-15 min before performing the transformation.
2. Mix 40 μ l of competent cells, with 7.5 μ l of the linearized DNA in the cuvettes, mix gently with the pipette and incubate for 5 min on ice.
3. Adjust the electroporation settings as follows: 1 500 V, 25 μ F, 600 Ω .
4. Place the cuvette in the electroporator chamber and apply the electric pulse.
5. Immediately resuspend the electroporated mixture in 1 ml of cold 1 M sorbitol and transfer cells to a sterile tube.
6. Allow the cells to recover for about 1 h at 30 °C then harvest by centrifugation at 2,000 g for 10 min.
5. Discard the supernatant and resuspend the pellet in 500 μ l 1 M sorbitol.

3.3.2. Selection of recombinant clones

1. Spread 2 x 250 μ l of electrotransformed cells on 2 MD plates and incubate for 2-3 days at 30°C.
3. Harvest the His⁺ transformants by adding 1 ml of YPD onto the plates and scrape off all the clones using a sterile scaper.
4. Perform 10 x and 100 x dilutions and measure the OD_{600nm} for each dilution.
5. Spread an equivalent of 10⁵ cells/plate (OD_{600nm} of 1 is equivalent to approximately 5x10⁷ cells/ml) on YPD plates supplemented with 0.1 or 0.25 mg/ml geneticin.
6. Incubate for 2-3 days at 30°C. Colonies should only grow if the cells contain the insert together with the selection marker integrated into the genome. Those colonies which grow on plates with the higher geneticin concentration should contain more copies of the selection marker and more copies of the target gene.

4. Notes

1. In the present example, the primers are designed to amplify a gene that will be fused in frame with tag sequences at both 5' and 3' ends. Therefore no stop codon is introduced in the sequence of the reverse primer.
2. If the PCR products appear as pure and single bands corresponding to the desired DNA fragment, you can directly use the PCR mixture in the next step. However, if the PCR reaction results in a number of non-specific products, optimise the reaction conditions by increasing the annealing temperature or the length of the annealing step.
3. If you use restriction enzymes that can be heat inactivated, this step is optional.
4. The sequencing step is mandatory here as the DNA insert has been obtained by PCR. This step is of course not necessary if the insert fragment is obtained from digestion of a previously-validated plasmid.
5. Check for the absence of the *PmeI* site in the gene to be expressed, otherwise you will generate several fragments instead of a linearized vector. In case *PmeI* is present, select another enzyme that will cut only once in the vector (*SacI* or *SalI* for instance).
6. Make sure the DNA pellet is completely dried otherwise it will be poorly resuspended and not at the concentration required for transformation.
7. This protocol describes the preparation of electrocompetent cells. Alternative protocols for chemically competent cells are also available in the Invitrogen *Pichia* manual, which is freely available at www.invitrogen.com.

References

1. Cereghino JL and Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24: 45–66.
2. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13: 329-332.
3. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22: 249–270.
4. Alkhalfioui F, Logez C, Bornert O, Wagner R. (2010) Expression systems: *Pichia pastoris*. In Robinson (Ed.): Production of Membrane Proteins – Strategies for Expression and Isolation. Wiley-VCH (in press).
5. Cregg JM, Barringer KJ, Hessler AY, Madden KR. (1985) *Pichia pastoris* as a host system for transformations. Mol Cell Biol. 5: 3376–3385.
6. Lee CC, Williams TG, Wong DWS, Robertson GH. (2005) An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid 54: 80–85.
7. Hong IP, Lee SJ, Kim YS, Choi SG (2007) Recombinant expression of human cathelicidin (hCAP18/LL-37) in *Pichia pastoris*. Biotechnol Lett 29: 73–78.
8. Sandstrom AG, Engstrom K, Nyhlen J, Kasrayan A, Backvall JE. (2009) Directed evolution of *Candida antarctica* lipase A using an episomal replicating yeast plasmid. Protein Eng Des Sel. 22: 413–420.
9. Zhang AL, Luo JX, Zhang TY, Pan YW, Tan YH, Fu CY, Tu FZ. (2009) Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. Mol Biol Rep 36: 1611–9.

10. Hartner FS, Glieder A. (2006) Regulation of methanol utilisation pathway genes in yeasts. *Microb Cell Fact* 5: 39.
11. Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg JM, Glieder A. (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 36: e76.
12. Cos O, Ramón R, Montesinos JL, Valero F. (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: A review. *Microb Cell Fact* 5: 17.
13. Fantoni A, Bill RM, Gustafsson L, Hedfalk K. (2007) Improved yields of full-length functional human FGF1 can be achieved using the methylotrophic yeast *Pichia pastoris*. *Protein Expr Purif* 52: 31–9.
14. Delroisse JM, Dannau M, Gilsoul JJ, El Mejdoub T, Destain J, Portetelle D, Thonart P, Haubruge E, Vandenbol M. (2005) Expression of a synthetic gene encoding a *Tribolium castaneum* carboxylesterase in *Pichia pastoris*. *Protein Expr Purif*. 42: 286–94.
15. Daly R, Hearn MTW. (2005) Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production. *J Mol Recognit* 18: 119–138.
16. Weiss HM, Haase W, Michel H, Reilander H. (1995) Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: Pharmacological characterization and localization. *FEBS Lett* 377, 451–456.
17. Grunewald S, Haase W, Molsberger E, Michel H. Reilander H. (2004) Production of the human D2S receptor in the methylotrophic yeast *P. pastoris*. *Receptors Channels* 10: 37–50.

18. Fischer G, Kosinska-Eriksson U, Aponte-Santamaria C, Palmgren M, Geijer C, Hedfalk K, Hohmann S, de Groot BL, Neutze R, Lindkvist-Petersson K. (2009) Crystal structure of a yeast aquaporin at 1.15 Å reveals a novel gating mechanism. PLOS Biol 7: e1000130.
19. Tornroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P. (2006) Structural mechanism of plant aquaporin gating. Nature 439: 688–694.
20. Terpe K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60: 523–33.
21. Woo JH, Liu YY, Mathias A, Stavrou S, Wang Z, Thompson J, Neville DM Jr. (2002) Gene optimization is necessary to express a bivalent anti-human anti-T cell immunotoxin in *Pichia pastoris*. Protein Expr Purif 25: 270–82.
22. Sinclair G and Choy FY (2002) Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. Protein Expr. Purif. 26: 96–105.
23. Tull D, Gottschalk TE, Svendsen I, Kramhøft B, Phillipson BA, Bisgård-Frantzen H, Olsen O, Svensson B. (2001) Extensive N-glycosylation reduces the thermal stability of a recombinant alkalophilic bacillus alpha-amylase produced in *Pichia pastoris*. Protein Expr Purif 21: 13–23.
24. Boettner M, Steffens C, von Mering C, Bork P, Stahl U, Lang C. (2007) Sequence-based factors influencing the expression of heterologous genes in the yeast *Pichia pastoris*--A comparative view on 79 human genes. J Biotechnol 130: 1–10.

25. De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert, N. (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat Biotechnol* 27: 561–569.
26. Hamilton SR, Gerngross TU. (2007) Glycosylation engineering in yeast: The advent of fully humanized yeast. *Curr Opin Biotechnol* 18; 387–392.
27. Pla IA, Damasceno LM, Vannelli T, Ritter G, Batt CA, Shuler ML. (2006) Evaluation of Mut⁺ and MutS *Pichia pastoris* phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration. *Biotechnol Prog* 22: 881–8.
28. André N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, Pattus F, Michel H, Wagner R, Reinhart C. (2006) Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci* 15: 1115–1126.
29. Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras, TR. (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol Cell Biol* 5: 1111–1121.
30. Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Tomomitsu K. (2000) Addition of oleic acid increases expression of recombinant human serum albumin by the AOX2 promoter in *Pichia pastoris*. *J Biosci Bioeng* 89: 479–484.
31. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186: 37–44.

32. Shen SG, Sulter G, Jeffries TW and Cregg JM. (1998) A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. *Gene* 216: 93–102.
33. Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. (1987) Expression of the lacZ gene from two methanol regulated promoters in *Pichia pastoris*. *Nucl Acids Res* 15, 3859–3876.
34. Menendez J, Valdes I, Cabrera N. (2003) The ICL1 gene of *Pichia pastoris*, transcriptional regulation and use of its promoter. *Yeast* 20: 1097-1108.
35. Sears IB, O'Connor J, Rossanese OW, Glick BS. (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. *Yeast* 14: 783-790.
36. Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, Choi E, Jung J, Lee H. (2007) Translation elongation factor 1-alpha gene from *Pichia pastoris*: molecular cloning, sequence, and use of its promoter. *Appl Microbiol Biotechnol* 74: 601–608.
37. Ahn J, Hong J, Park M, Lee H, Lee E, Kim C, Lee J, Choi E, Jung J, Lee H. (2009) Phosphate-responsive promoter of a *Pichia pastoris* sodium phosphate symporter. *Appl and Env Microb* 75: 3528–3534.
38. Liu H, Tan XQ, Russell KA, Veenhuis M, Cregg JM. (1995) PER3, a gene required for peroxisome biogenesis in *Pichia pastoris*, encodes a peroxisomal membrane-protein involved in protein import. *J Biol Chem* 270: 10940-10951.
39. Hashimoto Y, Koyabu N, Imoto T. (1998) Effects of signal sequences on the secretion of hen lysozyme by yeast: construction of four secretion cassette vectors. *Prot Eng* 11: 75-77.

40. Payne WE, Gannon PM, Kaiser CA. (1995) An inducible acid phosphatase from the yeast *Pichia pastoris*—characterisation of the gene and its product. *Gene* 163, 19–26.
41. Paifer E, Margolles E, Cremata J, Montesino R, Herrera L, Delgado J.M. (1994) Efficient expression and secretion of recombinant alpha amylase in *Pichia pastoris* using two different signal sequences. *Yeast* 10: 1415–1419.
42. Raemaekers RJM, de Muro L, Gatehouse JA, Fordham-Skelton AP. (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: 394–403.
43. Kato S, Ishibashi M, Tatsuda D, Tokunaga H, Tokunaga M. (2001) Efficient expression, purification and characterization of mouse salivary alpha-amylase secreted from methylotrophic yeast, *Pichia pastoris*. *Yeast* 18 643–655.
44. Tokunaga M, Kawamura A, Omor A, Hishinuma F. (1992) Structure of yeast pGKL 128-kDa killer-toxin secretion signal sequence. Processing of the 128-kDa killer-toxin-secretion-signal - a-amylase fusion protein. *Eur J Biochem* 203, 415–423.
45. Oka C, Tanaka M, Muraki M, Harata K, Suzuki K, Jigami Y. (1999) Human lysozyme secretion increased by alpha-factor pro-sequence in *Pichia pastoris*. *Biosci Biotechnol Biochem* 63: 1977–1983.
46. Cregg JM, Madden KR. (1987) Development of yeast transformation systems and construction of methanol-utilization-defective mutants of *Pichia pastoris* gene disruption. *Biological Research on Yeasts*, II, 1–18.

47. Cregg JM, Madden KR. (1989) Use of site-specific recombination to regenerate selectable markers. *Mol Gen Genet* 219: 320-323.

48. White CE, Hunter MJ, Meininger DP, White LR, Komives EA. (1995) Large-scale expression, purification and characterization of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388. *Protein Eng* 8: 1177-1187.

Figure legend

Figure 1: Your Favorite *P. pastoris* Expression Vector (YFPpEV). The five different boxes represent the main elements that must be considered when building an expression vector. GOI: gene of interest.

TABLE 1: <i>P. pastoris</i> promoters for recombinant expression				
Name	Original gene	Nature (<i>expression condition</i>)	Strength	Ref.
P _{AOX1}	Alcohol oxidase I	Inducible (<i>methanol</i>)	Strong	29
P _{AOX2}	Alcohol oxidase II	Inducible (<i>methanol</i>)	Strong	30
P _{GAP}	Glyceraldehyde-3-phosphate dehydrogenase	Constitutive (<i>glucose, glycerol</i>)	Strong	31
P _{FLD1}	Formaldehyde dehydrogenase	Inducible (<i>methanol, methylamine</i>)	Strong	32
P _{DHAS}	Dihydroxyacetone synthase	Inducible (<i>methanol</i>)	Weak	33
P _{ICL}	Isocitrate lyase	Inducible (<i>ethanol</i>)	Not determined	34
P _{YPT1}	GTP binding protein	Constitutive (<i>glucose, methanol, mannitol</i>)	Weak	35
P _{TEF1}	Translation elongation factor 1 α	Constitutive (<i>glucose</i>)	Strong	36

P _{PHO89}	Sodium phosphate symporter	Inducible (<i>phosphate starvation</i>)	Strong	37
P _{PEX8}	Peroxisomal matrix protein	Inducible (<i>methanol</i>)	Weak	38

Table 2: Signal sequences used for extracellular expression

Name	Protein signal sequence	Sequence structure	Ref.
α factor	<i>S. cerevisiae</i> α mating factor	85 residues 3 N-glycosylation sites Kex2/Ste13 processing sites	39
PHO1	<i>P. pastoris</i> acid phosphatase	15 residues 6 N-glycosylation sites	40
SUC2	<i>P. pastoris</i> invertase	19 residues	39, 41
PHA-E	<i>Phaseolus vulgaris</i> phytohemagglutinin	21 residues	42
KILM1	Killer toxin type I	44 residues	42
pGKL	128 kDa killer protein	29 residues Kex2 processing site	43, 44
CLY	Chicken lysozyme signal peptide	18 residues	45
CLY-L8	Engineered leucine-rich signal peptide	16 residues	45

Table 3A: Additional tag sequences

Name	Peptidic sequence	Size
His-tag	HHHHHH(HHHH)	6 to 10 aa.
Flag-tag	DYKDDDDK	8 aa.
Strep II tag	WSHPQFEK	8 aa.
1D4 antibody recognition sequence	TETSQVAPA	9 aa.
HA epitope tag	YPYDVPDYAS	10 aa.
c-Myc epitope	EQKLISEEDL	11 aa.

V5 epitope	GKPIPNPLLGLDST	14 aa.
Calmodulin binding peptide	KRRWKKNFIAVSAANRFKKISSSGAL	26 aa.
Bio-tag	Peptide	85 aa.
GFP	Protein	35 kDa

Table 3B: Protease cleavage sites

Name	Recognized sequence	Protease size (kDa)
Enterokinase	DDDDK	26
Tobacco etch virus (TEV) protease	ENLYFQS	28
Thrombin	LVPRGS/F	35
Factor Xa.	IE/DGR	43

Table 4: Commercially available expression vectors

Name	Selection marker	Phenotype of transformants	Promoter	Secretion sequence	Added tags
pAO815	<i>HIS4</i>	His ⁺	P _{AOX1}	None	None
pPIC3.5K	<i>HIS4</i> , Kan	His ⁺ , G418 ^R	P _{AOX1}	None	None
pPIC9K	<i>HIS4</i> , Kan	His ⁺ , G418 ^R	P _{AOX1}	α factor	None
pPICZ A, B, C	Ble	Zeo ^R	P _{AOX1}	None	c-Myc/his6
pPICZ α A, B, C	Ble	Zeo ^R	P _{AOX1}	α factor	c-Myc/his6
pPIC6 A, B, C	Bsd	Bla ^R	P _{AOX1}	None	c-Myc/his6
pHIL-D2	<i>HIS4</i>	His ⁺	P _{AOX1}	None	None
pHIL-S2	<i>HIS4</i>	His ⁺	P _{AOX1}	PHO1	None
pFLD	Ble	Zeo ^R	P _{FLD1}	None	V5 epitope/his6
pFLD α	Ble	Zeo ^R	P _{FLD1}	α factor	V5 epitope/his6
pGAPZ A, B, C	Ble	Zeo ^R	P _{GAP}	None	c-Myc/his6
pGAPZ α A, B, C	Ble	Zeo ^R	P _{GAP}	α factor	c-Myc/his6

pPink-HC	<i>ADE2</i>	Ade ⁺	P _{AOX1}	None	None
pPink-LC	<i>ADE2</i>	Ade ⁺	P _{AOX1}	None	None
pPink α -HC	<i>ADE2</i>	Ade ⁺	P _{AOX1}	α factor	None

HIS4 and *ADE2*: auxotrophy markers for histidine and adenine, respectively; Kan, Ble and Bsd: bacterial genes conferring *P. pastoris* resistance to G418 (or geneticine), zeocine and blasticidine antibiotics, respectively.

Table 5: *P. pastoris* most commonly used strains

Strain	Genotype	Phenotype	Ref.
Y11430	wild-type	Mut ⁺	NRRL
X-33	wild-type	Mut ⁺	NRRL
GS115	<i>his4</i>	Mut ⁺ , His ⁻	46
KM71	<i>his4, arg4, aox1::ARG4</i>	Mut ^S , His ⁻ , Arg ⁺	46
MC100-3	<i>his4, arg4, aox1::ARG4, aox2::HIS4</i>	Mut ⁺ , His ⁺ , Arg ⁺	47
SMD1163	<i>his4, pep4, prb1</i>	Mut ⁺ , His ⁻ , Prot ⁻ (A ⁻ , B ⁻ , CarbY ⁻)	48
SMD1165	<i>his4, prb1</i>	Mut ⁺ , His ⁻ , Prot ⁻ (B ⁻)	48
SMD1168	<i>his4, ura3, pep4::URA3</i>	Mut ⁺ , His ⁻ , Prot ⁻ (A ⁻ , B ^S , CarbY ⁻)	48
PichiaPink [®] Strain 1	<i>ade2</i>	Mut ⁺ , Ade ⁻ ,	Invitrogen
PichiaPink [®] Strain 2	<i>ade2, pep4</i>	Mut ⁺ , Ade ⁻ , Prot ⁻ (A ⁻ , B ^S , CarbY ⁻)	Invitrogen
PichiaPink [®] Strain 3	<i>ade2, prb1</i>	Mut ⁺ , Ade ⁻ , Prot ⁻ (B ⁻)	Invitrogen
PichiaPink [®] Strain 4	<i>ade2, pep4, prb1</i>	Mut ⁺ , Ade ⁻ , Prot ⁻ (A ⁻ , B ⁻ , CarbY ⁻)	Invitrogen

NRRL: Northern Regional Research Laboratories, Peoria, IL.