The Pichia System

James Cregg, Ph.D., Keck Graduate Institute, Claremont, Calif.

Introduction

Pichia pastoris is a highly successful system for production of a wide variety of recombinant proteins. Several factors have contributed to its rapid acceptance, the most important of which include: (1) a promoter derived from the alcohol oxidase I (*AOX1*) gene of *P. pastoris* that is uniquely suited for the controlled expression of foreign genes; (2) the similarity of techniques needed for the molecular genetic manipulation of *P. pastoris* to those of *Saccharomyces cerevisiae*; (3) the strong preference of *P. pastoris* for respiratory growth, a key physiological trait that greatly facilitates its culturing at high-cell densities relative to fermentative yeasts; and (4) the decision in 1993 by Phillips Petroleum Company in Bartlesville, Okla. (and continued by Research Corporation Technologies [RCT]), to release the *P. pastoris* expression system to academic research laboratories, the consequence of which has been an explosion in the knowledge base on the system as described in numerous recent publications.

As a yeast, *P. pastoris* is a single-celled microorganism that is easy to manipulate and culture. However, it is also a eukaryote and capable of many of the post-translational modifications performed by higher eukaryotic cells such as proteolytic processing, folding, di-sulfide bond formation and glycosylation. Thus, many proteins that end up as inactive inclusion bodies in bacterial systems are produced as biologically active molecules in *P. pastoris*. The *P. pastoris* system is also generally regarded as being faster, easier, and less expensive to use than expression systems derived from higher eukaryotes such as insect and mammalian tissue culture cell systems and usually gives higher expression levels.

In this short introduction, we review basic aspects of the *P. pastoris* expression system. Further information on the *P. pastoris* system can be found in the numerous reviews describing the system, the *Pichia* Expression Kit Instruction Manual (Invitrogen Corporation, Carlsbad, Calif.). Web sites provide additional strains and vectors that can be obtained (http://www.invitrogen.com; and http://faculty.kgi.edu/cregg/index.htm), and the DNA sequence and structure of many Pichia vectors (http://www.invitrogen.com; and http://faculty.kgi.edu/cregg/index.htm) and, most recently, the sequence of the *P. pastoris* genome (contact: (vinayak@integratedgenomicscom).

1. A Brief History of the *P. pastoris* Expression System

The ability of certain yeast species to utilize methanol as a sole source of carbon and energy was discovered less than 40 years ago by Koichi Ogata. Because methanol could be inexpensively synthesized from natural gas (methane), there was immediate interest in exploiting these organisms for generating yeast biomass or single cell protein (SCP) to be marketed primarily as a high protein animal feed. During the 1970s, Phillips Petroleum Company developed media and methods for growing *P. pastoris* on methanol in continuous culture at high cell densities (<130 g/l dry cell weight). However, during this same period, the cost of methane increased dramatically due to the oil crisis and the cost of soy beans, the major alternative source of animal feed protein, decreased. As a result, the SCP process was never economically competitive.

In the early 1980s, Phillips Petroleum contracted with the Salk Institute Biotechnology/Industrial Associate Inc. (SIBIA), a biotechnology company located in La Jolla, Calif., to develop *P. pastoris* as a heterologous gene expression system. Researchers at SIBIA isolated the *AOX1* gene (and its promoter) and developed vectors, strains and methods for molecular genetic manipulation of *P. pastoris*. The combination of strong regulated expression under control of the *AOX1* promoter along with the fermentation media and methods developed for the SCP process resulted in strikingly high levels of foreign proteins in *P. pastoris*. In 1993, Phillips Petroleum sold its patent position with the *P. pastoris* expression system to RCT, the current patent holder. In addition, Phillips Petroleum licensed Invitrogen to sell components of the system to researchers worldwide, an arrangement that continues under RCT.

2. *P. pastoris* as a Methylotrophic Yeast

P. pastoris is one of approximately a dozen yeast species representing four different genera capable of metabolizing methanol. The other genera include *Candida*, *Hansenula* and *Torulopsis*. The methanol metabolic pathway appears to be the same in all yeasts and involves a unique set of pathway enzymes. The first step in the metabolism of methanol is the oxidation of methanol to formaldehyde, generating hydrogen peroxide in the process, by the enzyme alcohol oxidase (AOX). To avoid hydrogen peroxide toxicity, this first step in methanol metabolism takes place within a specialized organelle, called the peroxisome, which sequesters toxic

hydrogen peroxide away from the rest of the cell. AOX is a homo-octomer with each subunit containing one non-covalently bound FAD (flavin adenine di-nucleotide) co-factor. Alcohol oxidase has a poor affinity for O_2 and methylotrophic yeasts appear to compensate for this deficiency by synthesizing large amounts of the enzyme.

There are two genes in *P. pastoris* that code for AOX—*AOX1* and *AOX2*—but the *AOX1* gene is responsible for the vast majority of alcohol oxidase activity in the cell. Expression of the *AOX1* gene is tightly regulated and induced by methanol to high levels. In methanol-grown shake-flask cultures, this level is typically about 5 percent of total soluble protein but can be ≥ 30 percent in cells fed methanol at growth limiting rates in fermentor cultures. Expression of the *AOX1* gene is controlled at the level of transcription. In methanol-grown cells, approximately 5 percent of polyA^{+} RNA is from the *AOX1* gene whereas in cells grown on other carbon sources, AOX1 message is undetectable. The regulation of the *AOX1* gene is similar to the regulation of the *GAL1* gene of *S. cerevisiae* in that control appears to involve a two mechanisms: a repression/de-repression mechanism plus an induction mechanism. However, unlike *GAL1* regulation, de-repressing conditions (e.g. the absence of a repressing carbon source such as glucose in the medium) do not result in substantial transcription of the *AOX1* gene. The presence of methanol appears to be essential to induce high levels of transcription.

3. Secretion of Heterologous Proteins

With *P. pastoris*, heterologous proteins can either be expressed intracellularly or secreted into the medium. Because *P. pastoris* secretes only low levels of endogenous proteins, and because its culture medium contains no added proteins, a secreted heterologous protein comprises the vast majority of the total protein in the medium. Thus, secretion serves as a major first step in purification, separating the foreign protein from the bulk of cellular proteins. However, the option of secretion is usually limited to foreign proteins that are normally secreted by their native hosts. Secretion requires the presence of a signal sequence on the foreign protein to target it to the secretory pathway. While several different secretion signal sequences have been used successfully, including the native secretion signal present on some heterologous proteins, success has been variable. The secretion signal sequence from the *S. cerevisiae* α factor prepro peptide has been used with the most success.

4. Common Expression Strains

All *P. pastoris* expression strains are derivatives of NRRL-Y 11430 (Northern Regional Research Laboratories, Peoria, Ill.). Some have a mutation in one or more auxotrophic genes [e.g., GS115 (*his4*)] to allow for selection of expression vectors containing the complementing biosynthetic gene (e.g. *HIS4*) upon transformation. All of these strains grow on complex media but require supplementation with histidine (or other appropriate nutrient) for growth on minimal media.

Three types of host strains are available that vary with regard to their ability to utilize methanol due to deletions in one or both *AOX* genes. Strains with deleted AOX genes sometimes are better producers of a foreign protein than wild-type strains. These strains also require much less methanol to induce expression, which can be useful in large fermenter cultures where large amounts of methanol is sometimes considered a significant fire hazard. However, the most commonly used expression host is wild-type or GS115 (*his4*), which are wild type with regard to the *AOX1* and *AOX2* genes and grow on methanol at the wild-type rate (methanol utilization plus or Mut^+ phenotype). KM71 (*his4 arg4 aox1Δ::ARG4*) is a strain in which the chromosomal *AOX1* gene is largely deleted and replaced with the *S. cerevisiae ARG4* gene. As a result, this strain must rely on the much weaker *AOX2* gene for AOX and grows on methanol at a slow rate (methanol utilization slow or Mut^S phenotype). With many *P. pastoris* expression vectors, it is possible to insert an expression cassette and simultaneously delete the *AOX1* gene of a Mut^+ strain. The third host, MC100-3 (*his4 arg4 aox1Δ::SARG4 aox2Δ::Phis4*), is deleted for both AOX genes and is totally unable to grow on methanol (methanol utilization minus or Mut^- phenotype).

Some secreted foreign proteins are unstable in the *P. pastoris* culture medium and are rapidly degraded by proteases there. Major vacuolar proteases appear to be a significant factor in degradation, particularly in fermenter cultures, due to the high cell density environment combination with the lysis of a small percentage of cells. The use of host strains that are defective in these proteases has proven to help reduce degradation in several instances. SMD1163 (*his4 pep4 prb1*), SMD1165 (*his4 prb1*), or SMD1168 (*his4 pep4*) are a series of protease-deficient strains that may provide a more suitable environment for expression of certain heterologous proteins. The *PEP4* gene encodes proteinase A, a vacuolar aspartyl protease required for the activation of other vacuolar proteases such as carboxypeptidase Y and proteinase

B. Proteinase B, prior to processing and activation by proteinase A, has about half the activity of the processed enzyme. The *PRB1* gene codes for proteinase B. Therefore, *pep4* mutants display a substantial decrease or elimination in proteinase A and carboxypeptidase Y activities, and partial reduction in proteinase B activity. In the *prb1* mutant, only proteinase B activity is eliminated, while *pep4 prb1* double mutants show a substantial reduction or elimination in all three of these protease activities.

5. Expression Vectors

Plasmid vectors designed for heterologous protein expression in *P. pastoris* have several common features. The foreign gene expression cassette is one of those and is composed of DNA sequences containing the *P. pastoris AOX1* promoter, followed by one or more unique restriction sites for insertion of the foreign gene, followed by the transcriptional termination sequence from the *P. pastoris AOX1* gene that directs efficient 3' processing and polyadenylation of the mRNAs. Some vectors also contain *AOX1* 3' flanking sequences that are derived from a region of the *P. pastoris* genome that lies immediately 3' of the *AOX1* gene and can be used to direct fragments containing a foreign gene expression cassette to integration at the *AOX1* locus by gene replacement (or gene insertion 3' to *AOX1* gene).

Additional features that are present in certain *P. pastoris* expression vectors serve as tools for specialized functions. For secretion of foreign proteins, vectors have been constructed that contain a DNA sequence immediately following the *AOX1* promoter that encodes a secretion signal. The most frequently used of these is the *S. cerevisiae* α factor prepro signal sequence. However, vectors containing the signal sequence derived from the *P. pastoris* acid phosphatase gene (*PHO1*) are also available.

Vectors with dominant drug resistance markers that allow for enrichment of strains that receive multiple copies of foreign gene expression cassettes during transformations have been developed. One set of vectors (pPIC3K and pPIC9K) contains the bacterial kanamycin-resistance gene and confers resistance to high levels of G418 upon strains that contain multiple copies of these vectors. Another set of vectors, the pPICZ series, contains the *Sh ble* gene from *Streptoalloteichus hindustanus*. This gene is small (375 bp) and confers resistance to the drug Zeocin in *E. coli*, yeasts (including *P. pastoris*) and other eukaryotes. Because the *ble* gene serves as the selectable marker for both *E. coli* and *P. pastoris*, the Zeo^R vectors are much

smaller (~3 kb) and easier to manipulate than other *P. pastoris* expression vectors. These vectors also contain a multiple cloning site (MCS) with several unique restriction sites for convenience of foreign gene insertion and sequences encoding the His_6 and *myc* epitopes so that foreign proteins can be easily epitope tagged at their carboxyl termini, if desired.

Another feature present on certain vectors (e.g., pAO815 and the pPICZ vector series) is designed to facilitate the construction of expression vectors with multiple expression cassette copies. Multiple copies of an expression cassette are introduced in these vectors by inserting an expression cassette bounded by a BamHI and a BglII site into the BamHI site of a vector already containing a single expression cassette copy. The resulting BamHI/BglII junction between the two cassettes can no longer be cleaved by either enzyme allowing for the insertion of another BamHI-BglII-bounded cassette into the same vector to generate a vector with three cassette copies. The process of addition is repeated until six to eight copies of a cassette are present in a single final vector, which is then transformed into the *P. pastoris* host strain.

Finally, vectors containing a constitutive *P. Pastoris* promoter derived from the *P. pastoris* glyceraldehyde-3-phosphate dehydrogenase gene (*GAP*) have recently become available. The *GAP* promoter is a convenient alternative to the *AOX1* promoter for expression of genes whose products are not toxic to *P. pastoris*. In addition, its use does not involve the use of methanol, which may be problematic in some instances.

6. Post-translational Modifications

P. pastoris has the potential of performing many of the posttranslational modifications typically associated with higher eukaryotes. These include processing of signal sequences (both pre- and prepro-type), folding, disulfide bridge formation, and both O- and N-linked glycosylation.

Glycosylation of secreted foreign (higher) eukaryotic proteins by *P. pastoris* and other fungi can be problematic. In mammals, O-linked oligosaccharides are composed of a variety of sugars including N-acetylgalactosamine, galactose and sialic acid. In contrast, lower eukaryotes, including *P. pastoris*, add O-oligosaccharides solely composed of mannose (Man) residues. The number of Man residues per chain, their manner of linkage and the frequency and specificity of O-glycosylation in *P. pastoris* have yet to be determined. One should not assume that because a protein is not O-glycosylated by its native host, *P. pastoris* will not glycosylate it. *P. pastoris*

added O-linked mannose to approximately 15 percent of human IGF-1 protein, although this protein is not glycosylated at all in humans. Furthermore, one should not assume that the specific Ser and Thr residue (s) selected for O-glycosylation by *P. pastoris* will be the same as the native host.

N-glycosylation in *P. pastoris* and other fungi is also different than in higher eukaryots. In all eukaryotes, it begins in the ER with the transfer of a lipid-linked oligosaccharide unit, $Glc_3Man_9GlcNAc_2$ (Glc = glucose; GlcNAc = N-acetylglucosamine), to asparagine at the recognition sequence Asn-X-Ser/Thr. This oligosaccharide core unit is subsequently trimmed to $Man_8GlcNAc_2$. It is at this point that lower and higher eukaryotic glycosylation patterns begin to differ. The mammalian Golgi apparatus performs a series of trimming and addition reactions that generates oligosaccharides composed of either $Man_{5-6}GlcNAc_2$ (high-mannose type), a mixture of several different sugars (complex type) or a combination of both (hybrid type). Two distinct patterns of N-glycosylation have been observed on foreign proteins secreted by *P. pastoris*. Some proteins such as *S. cerevisiae* invertase are secreted with carbohydrate structures similar in size and structure to the core unit ($Man_{8-11}GlcNAc_2$).

Others foreign proteins secreted from *P. pastoris* receive much more carbohydrate and appear by SDS-PAGE and western blotting to be hyperglycosylated. Interestingly, *P. pastoris* does not appear to be capable of adding α1,3-terminal mannose to oligosaccharides. This contrasts with *S. cerevisiae* oligosaccharides where α1,3-linked terminal mannose is common. Aside from the probable absence of α1,3-linked mannose, little is known regarding the structure of *P. pastoris* outer-chain oligosaccharides. Furthermore, it is also not clear why outer chains are added to some *P. pastoris* secreted proteins and not others nor how outer chain addition may be prevented.

N-linked high mannose oligosaccharides added to proteins by yeasts represent a significant problem in the use of foreign secreted proteins by the pharmaceutical industry. They can be exceedingly antigenic when introduced intravenously into mammals and are rapidly cleared from the blood by the liver. Recent work by the groups of Roland Contreras in Ghent, Belgium, and Tillman Gerngross at Glycofi, Boston, USA, have resulted in the construction of *P. pastoris* host strains that produce proteins with mannose patterns identical to those on human glycoproteins. With continued work, it is hoped that these groups may be able to produce *P.*

pastoris-synthesized recombinant proteins that are fully humanized with the complex sugars typical of higher eukaryotic glycoproteins.