Applications of yeast in biotechnology: protein production and genetic analysis

Geoffrey P Lin Cereghino* and James M Cregg

Improvements in yeast expression systems, coupled with the development of yeast surface display and refinements in two-hybrid methodology, are expanding the role of yeasts in the process of understanding and engineering eukaryotic proteins.

Addresses
Department of Biochemistry and Molecular Biology, Oregon Graduate Institute of Science and Technology, 20 000 NW Walker Road, Beaverton, Oregon 97006-8921, USA
* e-mail: geoff@bmb.ogi.edu
Correspondence: James M Cregg; e-mail: cregg@bmb.ogi.edu

Current Opinion in Biotechnology 1999, 10:422–427
All rights reserved.

Abbreviations
Agglutinin
AOX1 alcohol oxidase I
BPTI bovine pancreatic trypsin inhibitor
ER endoplasmic reticulum
GAP glyceraldehyde-3-phosphate dehydrogenase
PDI protein disulfide isomerase
sHFc single-chain antibody variable region fragment

Introduction
As we enter the millennium, the so-called ‘Biotech Century’, scientists are continuing to engineer yeasts to produce eukaryotic proteins. Although these creations are not as appreciated as bread and beer, foreign proteins expressed in yeasts are being used to synthesize life-saving drugs for the pharmaceutical industry and unravel the complex regulatory phenomena at the heart of basic research. This review will describe recent advances in two broad areas: firstly, the use of yeasts to make large quantities of foreign proteins for research and therapeutic applications; secondly, the use of yeasts to determine the functional and regulatory dynamics of a recombinant protein, such as its interaction partners or its affinity for ligands.

Yeasts are suitable for these uses for several reasons. Foremost, yeasts offer the ease of microbial growth and gene manipulation found in bacteria along with the eukaryotic environment and ability to perform many eukaryotic-specific post-translational modifications, such as proteolytic processing, folding, disulfide bridge formation, and glycosylation [1]. Bacteria lack these capabilities and often produce eukaryotic proteins that are misfolded, insoluble, or inactive. Relative to more complex eukaryotic expression systems, such as Chinese hamster ovary cells and baculovirus-infected cell lines, yeasts are economical, usually give higher yields, and are less demanding in terms of time and effort [2••]. Nevertheless, there are disadvantages to using yeasts for expression of some heterologous proteins, mostly related to their inability to perform certain complex post-translational modifications, such as prolyl hydroxylation and amidation, as well as some types of phosphorylation and glycosylation [3]. Recent findings, however, should help to alleviate some of these problems and broaden the scope of future applications for yeasts in biotechnology.

Yeasts as protein factories
Yeasts have been used since the early 1980s for the large-scale production of intracellular and extracellular proteins of human, animal, and plant origin [4,5]. The expression of a foreign protein in yeasts consists of four steps: firstly, cloning of a foreign protein-coding DNA sequence within an expression cassette containing a yeast promoter and transcriptional termination sequences; secondly, transformation and stable maintenance of this DNA fusion in the host; thirdly, synthesis of the foreign protein under specified culture conditions; and finally, purification of the heterologous protein and comparison with its native counterpart. Usually, a regulatable promoter is used to drive foreign protein expression because, prior to induction, the ability to maintain cultures in an ‘expression off’ mode minimizes selection for non-expressing mutant cells during the cell growth phase. Such a selection can occur as a result of the added metabolic burden placed on cells expressing high levels of a foreign protein or the toxic effect of a foreign protein on the cells. Several yeast species have been engineered as systems for heterologous protein expression [6–9]. This review, however, will focus on the methylotrophic yeast Pichia pastoris and the baker’s yeast Saccharomyces cerevisiae, the organisms most commonly used for this purpose.

P. pastoris and other nonconventional yeasts
P. pastoris has been utilized to produce ~300 foreign proteins since 1984 [10,11,12••]. There are several factors that account for this system’s popularity: firstly, the use of the alcohol oxidase I (AOX1) promoter, one of the strongest, most regulated promoters known; secondly, the ability to stably integrate expression plasmids at specific sites in the P. pastoris genome in either single or multicopy; thirdly, the ability to culture strains in high density fermenters; and finally, its ready availability as a kit from Invitrogen Corporation (Carlsbad, CA, USA). The AOX1 promoter is tightly repressed by glucose and most other carbon sources but is induced >1000-fold in cells shifted to methanol as a sole carbon source [13]. With this promoter, expression of recombinant proteins is highly repressed while cultures are grown to high density in glucose or glycerol, which prevents selection for non-expressing mutant cells. Cultures are then shifted to a methanol medium to induce rapid high-level expression [14].

Nevertheless, there are limitations to this system, some of which have been remedied recently. Although essential for
maximum induction of the AOX1 promoter, methanol (a petroleum byproduct) is a potential fire hazard and may not be appropriate for the production of food products. Thus, strong promoters that do not require methanol for induction are needed. The glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter provides a constitutively high level of expression on glucose, glycerol, and methanol media [15]. With the GAP promoter one cannot repress expression of recombinant proteins, which limits its use to foreign genes the products of which are not a burden to the cells. A second promoter derived from the P. pastoris FLD1 gene, whose product is a glutathione-dependent formaldehyde dehydrogenase, can be induced either by methanol or methylamine (a nontoxic nitrogen source) in glucose-containing media. Expression levels from the methylamine-induced FLD1 promoter are comparable to those obtained with the AOX1 promoter in methanol [16].

A second limitation of the P. pastoris system has been a lack of moderately expressed promoters. The high level of expression provided by the AOX1, FLD1 and GAP promoters is toxic in some cases and may overwhelm the protein-handling machinery of the cell, causing a significant portion of the protein to be misfolded or unprocessed [17]. The availability of a variety of promoters would also facilitate the simultaneous expression of multiple genes, each at an optimal level, which may be important for the production of multisubunit proteins [18]. The promoter of the PEX8 gene, which encodes a peroxisomal biogenesis protein, gives low-level expression on glucose and is induced modestly (about 10-fold) when cells are shifted to methanol [19]. Another moderate promoter, derived from the P. pastoris YPT1 gene, provides a low but constitutive expression level in either glucose, methanol, or mannitol media [20].

A third limitation has been the existence of only a few selectable markers for P. pastoris transformation. Until recently, only three selectable marker genes, HIS4, ARG4, and Sh ble (for Zeocin antibiotic resistance), were available. To alleviate this problem, a series of expression vectors with new biosynthetic marker genes, ADE1 and URA3, have been constructed along with the respective auxotrophically marked P. pastoris strains (GP Lin Cerghino, unpublished data). The auxotrophically marked strains of P. pastoris are defective in one or more biosynthetic genes and thus cannot grow unless supplemented with the required nutrient or transformed with a vector containing a wild-type copy of the appropriate gene.

There have been improvements in other yeast systems as well (Table 1), such as the identification of new, strong promoters for foreign protein expression in Yarrowia lipolytica and Hansenula polymorpha [7,23]. Several new yeast expression systems have been reported, including Pichia methanolica, which touts many attributes of P. pastoris and Hansenula polymorpha, including the ability to be readily grown to very high cell densities and the availability of expression vectors that contain the tightly regulated alcohol oxidase promoter to control expression of foreign genes [22].

**Table 1**

<table>
<thead>
<tr>
<th>Species name</th>
<th>Promoter source</th>
<th>Regulation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol utilizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candida boidini</td>
<td>AOX1</td>
<td>Methanol induced</td>
<td>[38]</td>
</tr>
<tr>
<td>Hansenula polymorpha</td>
<td>MXK</td>
<td>Methanol induced</td>
<td>[9]</td>
</tr>
<tr>
<td>Pichia methanolica</td>
<td>AUG1</td>
<td>Methanol induced</td>
<td>[22]</td>
</tr>
<tr>
<td>Pichia pastoris</td>
<td>AOR1</td>
<td>Methanol induced</td>
<td>[24]</td>
</tr>
<tr>
<td>GAP</td>
<td>Strong constitutive</td>
<td>[16]</td>
<td></td>
</tr>
<tr>
<td>FLD1</td>
<td>Methanol or methylamine induced</td>
<td>[14]</td>
<td></td>
</tr>
<tr>
<td>PEX8</td>
<td>Moderate</td>
<td>[19]</td>
<td></td>
</tr>
<tr>
<td>YPT1</td>
<td>Methanol induced</td>
<td>[20]</td>
<td></td>
</tr>
<tr>
<td>Lactose-utilizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kluyveromyces lactis</td>
<td>LAC4</td>
<td>Lactose induced</td>
<td>[29]</td>
</tr>
<tr>
<td>PGK</td>
<td>Strong constitutive</td>
<td>[40]</td>
<td></td>
</tr>
<tr>
<td>ASH4</td>
<td>Ethanol induced</td>
<td>[21]</td>
<td></td>
</tr>
<tr>
<td>Starch-utilizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwanniomyces occidentalis</td>
<td>AMY1</td>
<td>Malate or starch induced</td>
<td>[9]</td>
</tr>
<tr>
<td>GAM1</td>
<td>Malate or starch induced</td>
<td>[8]</td>
<td></td>
</tr>
<tr>
<td>Xylose utilizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pichia stipitis</td>
<td>XYL1</td>
<td>Xylose induced</td>
<td>[9]</td>
</tr>
<tr>
<td>Afkan and fatty acid utilizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yarrowia lipolytica</td>
<td>XPR2</td>
<td>Peptrine induced</td>
<td>[41]</td>
</tr>
<tr>
<td>TEF</td>
<td>Strong constitutive</td>
<td>[7]</td>
<td></td>
</tr>
<tr>
<td>RPS7</td>
<td>Strong constitutive</td>
<td>[7]</td>
<td></td>
</tr>
</tbody>
</table>

Gene nomenclature: AOX1, alcohol dehydrogenase; AOX1, α-amylase; AOX1, AUG1, AOX1 and MXK, alcohol oxidase in species shown; FLD1, formaldehyde dehydrogenase; GAP, glyceraldehyde-3-phosphate dehydrogenase; GAM1, glucoamylase; LAC4; β-glactosidase; PEX8, peroxin 8; PGK, phosphoglycerate kinase from Saccharomyces cerevisiae; RPS7, ribosomal protein S7; TEF, translation elongation factor-1a; XPR2, extracellular protease; YPT1, GTPase involved in secretion.

**S. cerevisiae: the conventional alternative**

Despite a wealth of information on its genetics and molecular biology, S. cerevisiae is sometimes not viewed favorably as a host for recombinant protein expression because of the perception that it has a lower secretory capacity relative to P. pastoris and other yeasts [7]. The commonly used 3a multicopy vectors may, however, actually be the real culprits. When bovine pancreatic trypsin inhibitor (BPTI) was expressed from a 3a multicopy vector, expression levels of BPTI varied among the heterogeneous population of transformed cells, apparently because of the plasmid’s inherent instability. Cells with high copy numbers of the BPTI expression cassette accumulated unfolded protein in the endoplasmic reticulum (ER), which aggregated, overwhelmed, and essentially shut down the secretory pathway [23,24]. To remedy this problem, two approaches were tested: firstly, the construction of strains with selected numbers of integrated expression cassettes; and secondly,
the creation of strains that overexpress certain components of the ER folding machinery.

Strains containing one stably integrated copy of the expression cassette secreted more BPTI than strains with the same expression cassette on a 2\µ multicopy vector. Optimal expression was reached with 10 integrated copies [25]. Also, strains were created that overexpressed the Hsp70 chaperone BiP, which binds polypeptides in the ER during translocation, and protein disulfide isomerase (PDI), which catalyzes the formation and isomerization of disulfide bridges in the ER. Overexpression of either ER resident did not significantly increase BPTI secretion, even under conditions where the compartment was saturated [26]. Overexpression of BiP or PDI did, however, raise the secretion efficiency of a recombinant single-chain antibody variable region fragment (scFv) over twofold, whereas co-overexpression produced an eightfold increase [27]. The authors hypothesized that because scFvs are more aggregation prone than BPTI, these polypeptides benefited from the increased concentration of the chaperone protein and PDI in the ER. Thus, the unique biophysical properties of individual proteins determine whether modifications to the secretory apparatus will improve the secretion rate of that protein.

Using yeast to understand and improve protein function

*S. cerevisiae* can be used to elucidate and dissect the function of a protein in a manner similar to phage-display systems. Either can be used to detect protein-ligand interaction and to select mutant proteins with altered binding capacity [28]. Phage systems, however, typically cannot display secreted eukaryotic proteins in their native functional conformations. The yeast surface-display system alleviates this problem [29]. It displays eukaryotic proteins as fusions to the carboxyl terminus of an α-agglutinin (Agα) subunit (Agα2), a mating adhesion receptor (Figure 1). Agα2 fusion is disulfide bonded to the Agα1 peptide, which in turn is covalently linked to the yeast cell wall by phosphatidyl inositol glycan linkages. Thus, each product of a DNA library is tethered to the surface of the yeast cell wall in a manner that makes it accessible to macromolecular recognition without steric hindrance from cell-wall components. Ligand-binding kinetics and equilibrium can be measured using flow cytometry and other methods [30].

Originally the system was used to select for an antifluorescein scFv with improved binding from a randomly mutated library of the antibody fragment [29]. After the Agα2-scFv gene fusions were passed through an *E. coli* mutator strain, they were expressed on the yeast cell surface and subjected to three rounds of competitive binding to fluorescein isothiocyanate-dextran. This procedure successfully isolated clones with a threefold decreased antigen dissociation rate. The technique has also been applied to select scFv mutants with a higher affinity for a specific domain on a soluble biotinylated T-cell receptor [31]. These scFvs, presented on yeast, can stimulate native T cells to trigger responses that are associated with T-cell activation [32]. This approach may be utilized to generate antagonists that interact with a variety of other cell-surface receptors. Its availability as a kit from Invitrogen should enhance the accessibility of this method. The major limitation is that one must begin with a library restricted to variations of a specific protein or class of proteins. One cannot yet use yeast surface display to screen a library of total cellular protein.

Using yeast to detect protein–protein interaction

The yeast two-hybrid system is a well-established method to pluck an interaction partner of a protein from a library of random cellular proteins. Several recent reviews in this and other journals have superbly outlined the refinements and extensions of this system [33,34]. The basic two-hybrid system has three principal components: firstly, a vector directing the synthesis of a specific protein fused to the DNA-binding moiety of a transcription factor, termed the ‘bait’; secondly, a second vector directing synthesis of a library of proteins fused to the activation domain of a transcription factor, termed the ‘prey’; and finally, one or more reporter genes placed downstream of the DNA-binding sites recognized by the DNA-binding moiety of the bait. If the bait interacts with the prey, the two moieties of the transcription factor are joined and the downstream reporter genes are activated. False positive signals result when reporter activity is observed even though the bait and prey do not interact in nature. Recent improvements have helped to eliminate the number of false positives generated by this method, a significant problem in its use. The review by Colas and Brent [34] describes various scenarios.
that explain how these false positives arise and how they can be avoided.

More recently, the Golemis group investigated the effect of overexpressing several false positive clones in *S. cerevisiae*, both in the presence and absence of the original bait protein (EA Golemis, personal communication). Overexpression induced a variety of biological effects, including altered cell growth rate and permeability, which skewed the perceived activity of a *LacZ* reporter. They also found that the perceived reporter results were influenced by the assay method.

Recognizing these problems, the Golemis group has developed a novel dual bait system designed to simultaneously assay for protein interaction with two related or unrelated partners in a single cell [35••]. The general strategy consists of expressing two bait protein hybrids in a yeast cell, the authentic bait protein fused to the DNA-binding domain of LexA and a filtering protein for detecting false positives fused to the binding domain of λ cI. Along with a prey, the cell contains four reporter genes, two controlled by LexA and two controlled by λ cI. A true positive activates only the two LexA promoters but not the λ cI-regulated genes. In a model system assaying the interactions of two related GTPases as baits, the dual bait system was able to differentiate high affinity versus low affinity interactions in one step. The advantages of the system, which is available as a kit from Invitrogen, make it a valuable tool in high-throughput drug screening strategies, which are aimed at identifying agents that regulate the activity of biologically important target molecules.

Despite the many applications of the two-hybrid system, it is largely limited to the analysis of soluble proteins or the soluble domains of proteins. To investigate interactions between integral membrane proteins, a system based on split-ubiquitin has been developed [36,37•]. In this system, two membrane proteins (X and Y) are fused to the amino- and carboxy-terminal parts of the ubiquitin protein (Figure 2). The carboxy-terminal portion is also fused to a transcription factor (TF) to detect a ubiquitin heterodimer complex and is detected by a ubiquitin-specific protease. The protease liberates the transcription factor, which subsequently activates reporter genes in the nucleus. The system confirmed the interaction between Wbp1p and Ost1p, two known subunits of an ER membrane protein complex. This method has a serious limitation in that it requires knowledge of the topology of the protein of interest, as it works only if the ubiquitin components are fused to the cytoplasmic region of the proteins; however, it possesses the exciting potential to become a useful extension of the two-hybrid technique.

Conclusions

Yeast methodology is playing an increasingly crucial role in the study of higher eukaryotic proteins. Interacting partners of a protein that could not be identified with the original two-hybrid technique, such as membrane proteins, can be pursued with the split-ubiquitin method. The physical interactions of a protein can be characterized and modified with the use of such technology as yeast surface display. Finally, the original or improved protein can be produced in large quantities in one of several yeast host systems. Thus, yeasts have become theatrical stages where complex molecular phenomena of eukaryotes can be recreated and reengineered.

Acknowledgements

This work was supported by Grant DK43698 from the National Institutes of Health and Grant ER20334 from the Department of Energy, Office of Basic Energy Sciences.
Expression vectors and delivery systems

Expression Sciences to James M Cregg, and Postdoctoral Fellowship EWM199 from the National Institute of Environmental Health Sciences to Geoffrey P Las Casellas. We thank K Kiew Wittrup and Erica A Feliciano for sharing their unpublished findings with us, and to Nancy Chronic and Terri Hartfield for their help in the preparation of this manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest

**of outstanding interest


A good introduction to the P. pastoris and other expression systems. The advantages and disadvantages for the production of a specific class of proteins are described.


The split-ubiquitin system is modified for use in the detection of interactions between membrane proteins. The potential applications and limitations of this system are discussed.


