Novel Methods for Generating Strains with Increased Copy Number in the Yeast *Pichia pastoris*

Anthony Jay J. Sunga
B.A., University of Guam, 1997
M.S., Oregon Graduate Institute of Science and Technology, 1999

Presented to the Division of Environmental & Biomolecular Systems within
The Department of Science & Engineering
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
in
Biochemistry and Molecular Biology

October 2008

Department of Science & Engineering

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Anthony Jay J. Sunga

has been approved

James M. Cregg, Ph.D., Thesis Advisor

Professor

Geoff Lin Cereghino, Ph.D.

Associate Professor

Matthew S. Sachs, Ph.D.

Professor

James J. Whittaker, Ph.D.

Associate Professor

ACKNOWLEDGEMENTS

I wish to thank my advisor, Jim Cregg, for giving me this great opportunity to jump out of my "island lifestyle" and into the exciting world of molecular biology. This opportunity had provided the education, knowledge and experiences that are critical for a promising career in science. Jim's patience and trust gave me the confidence to thrive in the work I was assigned to do, but more importantly to believe that knowledge gained in failed experiments are as significant as those that are successful. I am very grateful for Jim's open door policy that was always full with helpful suggestions. His passion for *Pichia pastoris*, along with his kindness, generosity, and leadership in the lab make him a great advisor.

I also extend my gratitude to the people who had advised me during my committee meetings. Matthew Sachs, Gebre Woldegiorgis, James Whittaker and Geoff Cereghino, your contribution to my work will always be appreciated. Your suggestions were all insightful and helpful for my cause. I also thank you for the patience and the accommodations you made for our teleconference meetings. Thank you for your encouragement and support.

I would like thank the many people who have contributed in very significant ways. Nancy Christie and Terrie Hadfield, thank you for your help in keeping me updated with the latest requirements for graduation. I always felt secure knowing that you kept watch over me. The Sachs lab (Tony, Christina, etc.)--your friendship has made the late nights in the lab a time of joy and not work.

I am grateful to the wonderful members of the Cregg lab whom I have had the pleasure and honor of meeting during my journey as a Ph.D student. Kim, Russell, Monique Johnson, Shigang Shen, Oleg Stasyk, Deepa Raghavan, Chris Ilgen, Knut Madden and Bernard Dela Cruz, my thanks for giving me a different perspective to look from and for having the patience when I just needed to vent. My special thanks to Geoff and Joan Lin Cereghino, you have always treated me like an equal and accepted me like

one of your family. Ilya and Dina, you were my family away from Guam for the last few years. I would like to thank you also for challenging me to be a better researcher.

To the colleagues that I have met along the way and who have become great friends, I extend my deepest gratitude. Polina Datsova, Kirilynn Svay, Ting-jen Cheng, Nate Freund, Wongrat R., Augustine Velasco, Daniel, Sandip, Joel Oliveros, and the many others that have been important in my quest for this knowledge. You will always be a part of my life wherever it takes me.

My special and most heartfelt thanks go to my family on Guam and abroad. I would like to thank my mom, dad, Michelle, Louie, Isabelle and my wonderful relatives that have continued to support me even as the years began to add up. To Lorraine, without you I don't know where I would be now. Your unwavering support, love and patience with me all these years are truly special gifts that I will always consider as the foundation of my success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

Novel Methods for Generating Strains with Increased Copy Number in the Yeast *Pichia pastoris*

Anthony Jay J. Sunga, B.A., M.S.

Doctor of Philosophy
Division of Environmental and Biomolecular Systems within
The Department of Science and Engineering
and the Oregon Health and Science University

September 2008

Thesis Advisor: Dr. James M. Cregg

This thesis describes the development of a more efficient scheme for multiple copy selection and a new selectable marker in yeast, both of which add to the arsenal of tools used for gene expression in *Pichia pastoris*. Transformation and integration of the expression vector, and then gene expression to produce a high protein yield is a key objective in using *P. pastoris* as a host expression system. Because of its growing popularity for such use, the development of new markers and methods are essential to support this trend.

This thesis describes a novel scheme for the enrichment of multiple copy clones. We demonstrate that a transformant with a single or low copy number of the expression vector can be induced to amplify its vector copy number. The resulting clones selected after this posttransformational vector amplification (PTVA) process have been shown to consist of a tandem arrangement of full vector copies integrated into the genome. The PTVA process proved to be more convenient and efficient than previous methods.

The development of a new marker system for the transformation of *P. pastoris* utilizes the formaldehyde dehydrogenase gene (*FLD1*) as a selectable marker is also described in this thesis. We show that a *P. pastoris fld1* mutant strain can be transformed with normal efficiency using the wild type *FLD1* gene as a selectable marker. There are

two benefits of this marker system which provide researchers with an excellent alternative for the expression of their gene. The first advantage is that *FLD1* is endogenous to *P. pastoris*. This allows for the construction of an expression vector that contains no exogenous sequences, except for the foreign gene insert, upon transformation in the yeast. The lack of bacterial sequences upon transformation is important in the regulation of proteins expressed for therapeutic purposes. The second valuable feature of this marker system is the ability to enhance multiple copies in transformed strains. Studies have shown that increasing copy number of the expression vector is often directly proportional to the increase in protein production. These characteristics can prove to be a significant advantage for researchers in academic and industrial laboratories looking for an alternative marker system.

CHAPTER 1
INTRODUCTION

1.1 Recombinant DNA: The Doorway to Protein Expression

Most biological proteins that intrigue scientists are difficult to study because cells do not naturally synthesize them in great amounts (Tabor, 1988). Molecular biologists interested in obtaining large amounts of prokaryotic proteins for biochemical studies first described this process as overproduction, expression, or over-expression (Ausubel, Brent, et al., 1997). Expression is defined here as the directed synthesis of a recombinant protein by a selected host system (Rai, and Padh, 2001) and genetic engineering has provided that capability. Genetic engineering includes techniques that provide the ability to convert certain organisms into factories capable of producing a desired protein. This ability was first exploited when a collection of interrelated techniques converged to allow scientists to manipulate DNA at the nucleotide level and to introduce those manipulated DNAs back into certain organisms. (Tabor, 1988). By the early 1970s, necessary recombinant DNA (rDNA) techniques for this controlled genetic engineering in bacteria were available: methods for protein synthesis were described (Agarwal, Buchi, et al., 1970), the isolation and use of the first commercial restriction enzyme *EcoRI* was published (Yoshimori, Roulland-Dussoix, and Boyer, 1972), and the enzymatic covalent joining of two DNA fragments from different sources was possible (Jackson, Symons, and Berg, 1972). These advances eventually led to the work conducted by Cohen et al., (1973) that resulted in the first construction of a recombinant DNA (rDNA) molecule containing the genetic material from two different species (Cohen, Chang, et al., 1973). The joint use of these methods proved to be instrumental in producing a landmark study

demonstrating that a transformed *Escherichia coli* strain replicated and transcribed a toad DNA into RNA (Morrow, Cohen, et al., 1974). The hypothesis that a bacterium could express a gene from a higher organism was now established (Wright, 1986) and thus, paved the way to what has become the field of gene expression.

As promising as it seemed, during its infancy the new technology did not come without opposition. Those intimately involved with rDNA technology faced a broad spectrum of reactions from both the public and scientific communities (Wright, 2001). Optimistic predictions that genetic engineering could become a panacea (Cohen, 1977) and yet also a "Pandora's box", were topics that were vigorously debated. Contemplations about the quandaries posed by these new technologies led scientists to question the effects that genetic engineering would have on the environment and to advocate caution towards the advancement of molecular genetics (Cohen, 1977). This concern led to a proposal for a voluntary moratorium on recombinant DNA research in 1974 and to a meeting in 1975 at the Asilomar Conference Center in California (Berg, Baltimore, et al., 1974). This was followed by the first NIH guidelines for recombinant DNA research. The Asilomar and subsequent conferences helped satisfy the public's concern about the safety of rDNA research by offering a solution that seemed to soothe the vocal critics of this science. That solution was a restriction on their research in exchange for self governance (Wright, 2001). With the community at ease and scientists contented, the rapid expansion of the use of these powerful new technologies moved forward and gave rise to the field of biotechnology.

Since the 1940s, an extensive knowledge of bacterial genetics of the organism *Escherichia coli* has accumulated, giving rise to its dominance in the field of gene expression (Dominguez, Ferminan, et al., 1998). The rapid growth of this technology, which provided the tools to easily manipulate *E. coli*, and the ability of this organism to grow rapidly to high cell densities provided the possibility to express rare proteins in quantities that were sufficient and cost effective for academic and industrial use (Gellissen, Melber, et al., 1992). Therefore, it was not surprising that the first attempts to express large amounts of protein from cloned genes were performed in *E. coli*. Initially, these molecular biology studies were conducted only in an academic environment, but that dramatically changed when the potential of genetic engineering by rDNA methods

first showed its practical applications (Wright, 1986). In the 1970s and 1980's, small genetic engineering companies, who depended on universities for expertise, and were heavily funded by large multinational oil, chemical and pharmaceutical corporations formed. This changed the trend of research towards goals set by corporations with capital (Wright, 1986).

In 1976, Genentech was the first firm to use rDNA technology for drug development (Kuntz). In 1977, Somatostatin, a mammalian peptide involved in inhibiting the secretion of human growth hormone, insulin and glucagons, were the first functional polypeptides to be synthesized (Goeddel, Heyneker, et al., 1979; Itakura, Hirose, et al., 1977). Expression of the rat preproinsulin demonstrated the feasibility of directing secretion of a protein along with the removal of the signal sequence (Talmadge, Kaufman, and Gilbert, 1980). In 1982 and 1985, respectively, the FDA gave approval to companies to market and sell human insulin and human growth hormone expressed in *E. coli* (Biotechmedia). This approval validated the growing excitement from academic and privately funded biotechnology companies towards the field of rDNA technology for expression of recombinant proteins. Starting with only a handful of companies in the early 1970s, the number has blossomed to more than 1,200 companies with a market capitalization exceeding over $200 billion and revenues of $18.6 billion (Gwynne, and Page, 2006).

1.2 Selected Expression systems

1.2.1 Characteristics of an expression system

Biochemical analysis of proteins otherwise unavailable naturally in sufficient amounts and product development for biotechnology and pharmaceutical companies are the two main reasons for interest in gene expression (Shatzman, and Rosenberg, 1987). As the number of recombinant proteins increased, their complexity also increased (Dominguez, Ferminan, et al., 1998). *E. coli*, continues to dominate the field of gene expression in organisms but the field now also contains a number of other organisms. The choice of expression system is determined almost totally by the physical characteristics of the protein product. Thus, there is not one expression system that is

best for the expression of all proteins (Kingsman, and Kingsman, 1989). One of the limiting steps in genetic engineering is having a sufficient amount of physiological and genetic knowledge about an organism and molecular tools available for the design of vectors capable of being introduced, maintained and expressed in that organism (Schmidt, 2004)

When *E. coli* was the sole choice, researchers had little choice of expression host organism. Now a number of host systems are available and researchers have to thoroughly consider the disadvantages and advantages of each system when making a decision. Different hosts have different environments for protein synthesis which can significantly affect the functionality of the protein produced or how much effort may be needed during downstream processing events such as, the purification, and refolding of a protein (Eckart, and Bussineau, 1996). Finally, regulatory issues surrounding a particular host can delay a product launch and commercialization (Schmidt, 2004). Choosing the best system for expression involves a careful dissection of the many factors involved.

Some of the factors that should be considered in choosing a system for protein production are: growth characteristics of the host, expression level or protein yield per volume of culture, whether expression is intracellular or extracellular, post-translational modification events such as glycosylation and disulfide bridge formation, biological activity of the protein, process economics (cost of media, etc.) and the ease of manipulating the organism (Eckart, and Bussineau, 1996; Makrides, 1996). After those factors are considered, the general steps and protocols involved for all expression systems should be readily available for the chosen host. They include: 1) gene cloning (the isolation of a specific gene from an organism) and manipulation, 2) vector construction and transformation (returning the gene into a living cell), 3) a selective growth of the organism and, 4) a variety of issues related to the production of the protein (Kingsman, and Kingsman, 1989). The more advanced an organism is with regard to having many of these steps already developed, the better that host is likely to be for expression.

However, even with the best requirements of an expression system fulfilled, the expression of the gene product may not be optimal. Each new protein presents a new set of issues to overcome when optimizing protein production (Goeddel, 1990). In most

cases, it is not possible to recognize these issues prior to actually producing the protein in any one system.

1.2.2 Bacteria

Escherichia coli continue to be by far the best characterized and most frequently employed host system in the field of gene expression. It remains the most valuable host for expression of heterologous proteins and is the most appealing host when protein folding issues are not involved and posttranslational modifications are not required (Rai, and Padh, 2001; Shatzman, and Rosenberg, 1987). A vast amount of information has accumulated through decades of studies relating to protein expression in this bacterium. The comprehensive characterization of its genetics, and physiology has only helped strengthen its position in gene expression (Schmidt, 2004). This gram-negative bacterium's short generation time, ease of manipulation through established molecular tools, ability to grow to high densities in simple, inexpensive and defined media, and a recombinant protein production level that can reach up to 50% of the total protein further enhances its value (Rai, and Padh, 2001). As a result of decades of research, there is now an extensive list of vectors and mutated host strains, inducible promoters, selectable markers (Dominguez, Ferminan, et al., 1998) and other elegant tools to make this expression system more efficient and effective.

Unfortunately, *E. coli's* physiological characteristics do not lend themselves to the production of all recombinant proteins. This is because bacteria have a different internal environment than eukaryotic cells, which results in the misfolding of many eukaryotic proteins and because they cannot perform posttranslational modifications common on eukaryotic proteins. These modifications are often essential to the production of a fully active and functional protein (Gellissen, Melber, et al., 1992; Schmidt, 2004). In addition, other host related issues such as differences in codon usage between the foreign gene and the bacteria can be severe enough to affect mRNA translational efficiency or protein stability and ultimately the yield or activity of the gene product. Even when efficient expression of a eukaryotic protein is achieved in *E. coli*, the protein is often inactive due to its inability to fold into its proper three-dimensional structure in a prokaryotic environment. Such proteins form insoluble inclusion bodies that need to be

recovered from the cells and refolded. Sometimes a protein becomes toxic when produced in high concentrations (Dominguez, Ferminan, et al., 1998). Finally, *E. coli* itself synthesizes toxic proteins that are a regulatory concern when *E. coli* is used in the synthesis of pharmaceutical proteins (Rai, and Padh, 2001).

1.2.3 Mammalian

Animal cell culture is the system with the highest similarity to human cells, with respect to the protein synthetic environment and the pattern and capacity for post-translation modification. However, their cultivation is more complicated and costly and often results in lower product titers (Schmidt, 2004). Chinese Hamster Ovary (CHO) and Baby Hamster Kidney (BHK) cells have the advantage of being recognized as safe with regard to infectious and pathogenic agents and therefore to have a higher acceptance by regulatory bodies (Schmidt, 2004). If product authenticity is absolutely essential for clinical efficacy, then despite the many shortcomings, a mammalian host is commonly the best choice, as it offers the greatest degree of product fidelity. Mammalian expression techniques are time consuming, more costly and much more difficult to perform on a large scale. Complex nutrient requirements and low product concentration have meant that the end product must be highly value-added for this approach to be commercially viable (Rai, and Padh, 2001). Fortunately, most human therapeutic proteins are of high value and worthy of the high cost commitment.

1.2.4 Insects

Insect cells transformed by baculovirus vectors have reached a high level of popularity in the expression world. Baculoviruses are non-infectious to vertebrates and their promoters have been shown to be inactive in mammalians cells (Rai, and Padh, 2001). Compared to mammalian systems, they are considered to be more stress resistant, easily handled and more productive; thus they are frequently utilized for high-throughput protein expression (Schmidt, 2004). They are able to perform many protein modifications but are limited because insect cells lack the necessary functions to produce complex mammalian-type oligosaccharides (Kost, and Condreay, 1999). A well

developed set of transport systems allows proteins to be targeted into the proper cellular compartments; nucleus, mitochondria, membranes or secreted into the medium. A strong promoter from the major viral late gene, polyhedron, and the use of a virus that can be propagated to high levels in insect cells create a system that can produce large amounts of recombinant protein with relative ease (Rai, and Padh, 2001).

Like many systems, insect cells have their disadvantages when employed as an expression system. The deficiencies can be attributed to the use of the baculovirus itself. Having inefficient protein processing, folding and secretion capacity causes many proteins to be insoluble or improperly processed. A high level of protease activity normally directed against native baculovirus proteins requires the use of protease inhibitors or the development of protease deficient strains to alleviate this difficulty. The systems are discontinuous in that expression of the recombinant protein is a result of the fatal infection the host acquires when using the virus (Schmidt, 2004). Co-expression of chaperones, peptidases, foldases and glycosylating enzymes have been effective in some cases to help with these drawbacks, but the insect expression system can still be relatively expensive compared to bacterial and yeast systems (Ailor, and Betenbaugh, 1999).

1.2.5 Yeast Systems

Major breakthroughs in rDNA technology in the early 1970s gave rise to an increase in funding from corporations who were interested in protein expression. Consequently, it was no shock that the search for alternative host systems, aside from *E. coli*, was already underway. In the early 1980s, a mounting interest in yeast as a host system for the expression of eukaryotic proteins began its development as a desirable option for use in the production of important proteins in industrial and academic laboratories (Cregg, and Higgins, 1995). Being neither pathogenic nor pyrogenic, yeasts were considered as Generally Recognized As Safe (GRAS) by the American Food and Drug Administration (FDA). In contrast, toxic pyrogens can be found in the cell walls of bacteria, while mammalian cells presented a concern for oncogenic or viral DNAs (Dominguez, Ferminan, et al., 1998).

Being a true eukaryote, yet also a simple single-celled microorganism, yeast offered favorable advantages to researchers seeking a desirable protein expression system. They

shared the characteristics of rapid microbial growth and ease of genetic manipulation with the mechanisms in place to secrete heterologous proteins and to execute many eukaryotic post-translational modifications (Cereghino, and Cregg, 1999; Gellissen, Melber, et al., 1992). Modifications such as proper protein folding, formation of disulfide bridges or glycosylation usually produce a bioactive eukaryotic product not otherwise attainable in prokaryotic systems (Buckholz, and Gleeson, 1991; Eckart, and Bussineau, 1996). Secretion of recombinant proteins in yeast circumvents toxicity from accumulated product and also simplifies downstream purification steps because yeast typically secretes very low levels of endogenous proteins. Most proteins produced in yeast are soluble and stable, which are characteristics that are sometimes more important than achieving very high levels of expression (Eckart, and Bussineau, 1996). The growth media are inexpensive and enable for rapid growth in shake flasks as well as in large scale fermentations (Schmidt, 2004).

In a fate similar to that of *E. coli*, *Saccharomyces cerevisiae* was the first yeast species to be investigated for the purpose of heterologous gene expression (Faber, Harder, et al., 1995) because there was a vast amount of information already collected about its molecular biology and physiology. It had already been established for use in industry (e.g., bread, beer and wine making), and there were genetic manipulation techniques already developed to make it an obvious choice as an alternative expression system (Dominguez, Ferminan, et al., 1998). Early success in protein production, included human α-interferon (Hitzeman, Leung, et al., 1983), hepatitis B surface antigen (Valenzuela, Medina, et al., 1982), and calf prochymosin (Mellor, Dobson, et al., 1983). These were promising indicators that a yeast system was a good option for use in the biotechnology industry as well as for research purposes, but improvement of the system was still needed.

Like many other host systems, *S. cerevisiae* is not without limitations. The lack of a strong, tightly-regulated promoter, poor secretion efficiency, inability to attain high cell density in fed-batch fermenters, and protein hyperglycosylation were problems that were encountered when using this yeast species (Faber, Harder, et al., 1995). However, with the development of the ability to introduce and integrate exogenous DNA into cells (Gellissen, Melber, et al., 1992) of other yeast species, researchers had the potential to

overcome some of these disadvantages. In the past decade or so, a growing number of non−*Saccharomyces* yeasts have become available as hosts for recombinant protein production (Dominguez, Ferminan, et al., 1998; Schmidt, 2004).

1.3 Methylotrophic yeast

First reported in 1969, methylotrophic yeasts are unique in that they are able to use methanol as a sole carbon source for growth and energy (Ogata, Nishikawa, and Ohsugi, 1969). All methylotrophic yeast strains to date belong to one of four genera: *Hansenula, Pichia, Candida and Torulopsis*. *Candida* and *Torulopsis* are asporogenous and therefore do not have a genetically definable life cycle. However, *Hansenula* and *Pichia* are ascomycetous and can be readily maintained stably as vegetative haploids. Along with being homothallic, the simple and understood life cycle of these yeasts present an advantage in the control of mating by subjecting two strains crossed under nutritional limitation, and the isolation and phenotypic characterization of mutants (Cregg, 1987).

Extensive biochemical and physiological analysis of the yeast growing on methanol have demonstrated one of its key advantages as an expression system, regulated expression. The metabolism of methanol in this yeast requires the expression of several enzymes that are only present when methanol is used as a sole carbon source. During growth on methanol, enzymes such as alcohol oxidase (*AOX*) and dihydroxyacetone synthase (DHAS) and catalase constitute about 60-80% of the total cellular protein and are sequestered into organelles called peroxisomes. This compartmentalization serves to protect the cytosol from toxic the effects of hydrogen peroxide, which is metabolized within the peroxisome (Cregg, Digan, et al., 1989). Understanding the mechanisms involved in the biogenesis and function of peroxisomes found in methanol-grown cells and investigating the possible commercial applications for these yeasts (Faber, Harder, et al., 1995) were the driving force for the continued studies done in laboratories worldwide.

The presence of methanol-induced enzymes in these yeast suggested to early researchers that there is a tightly controlled regulatory system involved in the metabolism of methanol, which eventually led to the discovery of the tightly-regulated promoter for the alcohol oxidase gene (*AOX*). Using in vivo techniques, studies in 1983 and 1984 demonstrated that mRNA of *AOX*, dihydroxyacetone synthases (*DAS*) and catalase genes

of *Hansenula polymorpha* and *Pichia pastoris* were under transcriptional control (Cregg, Digan, et al., 1989). Further studies have now elucidated the biochemical pathway for methanol utilization which appears to be similar in all methylotrophic yeasts. The first step is the oxidization of methanol to formaldehyde and hydrogen peroxide by the enzyme *AOX*. Hydrogen peroxide is broken into water and oxygen by the enzyme catalase. Formaldehyde on the other hand follows two possible pathways. In the first, formaldehyde is oxidized further into formate and eventually to carbon dioxide by a dissimilatory pathway that generates energy. Alternatively, formaldehyde can be assimilated to form cellular constituents by condensing it with xylulose 5-monophosphate, a reaction catalyzed by DHAS. The product of this pathway, glyceraldehyde 3-phosphate (GAP), continues into the metabolic pathway of the cells (Cregg, Digan, et al., 1989) (Figure 1.1). With the cloning and characterization of the promoter elements controlling the high level expression of these enzymes involved in methanol metabolism, methylotrophic yeast have become attractive hosts for the expression of foreign genes.

1.3.1 *Pichia pastoris*

Pichia pastoris was originally utilized in the 1970s by Phillips Petroleum Co. for the production of single-cell protein (SCP), but development ceased when the economics of using methanol for the manufacture of SCP were no longer favorable. *P. pastoris* was then optimized by SIBIA as a second generation expression system to help overcome problems encountered when using *S. cerevisiae* as a host system (Cregg, Digan, et al., 1989). Researchers observed that methylotrophic yeast like *P. pastoris* had a methanol-inducing mechanism that expressed *AOX* when cells were grown in methanol, whereas *AOX* was not detected when grown with other carbon sources. Thus they expected that the promoter driving the expression of the *AOX* enzyme was regulated by methanol and postulated that it would provide an excellent choice for the regulation of protein expression. Therefore, a significant discovery that assisted the development of this yeast was the cloning and characterization of the *AOX* gene and its promoter (Lin Cereghino, Sunga, et al., 2001). Combined with the fact that most molecular genetic techniques for *P. pastoris* have turned out to be similar to the well-developed yeast *S. cerevisiae*, the development of vectors, strains and modified protocols have established this

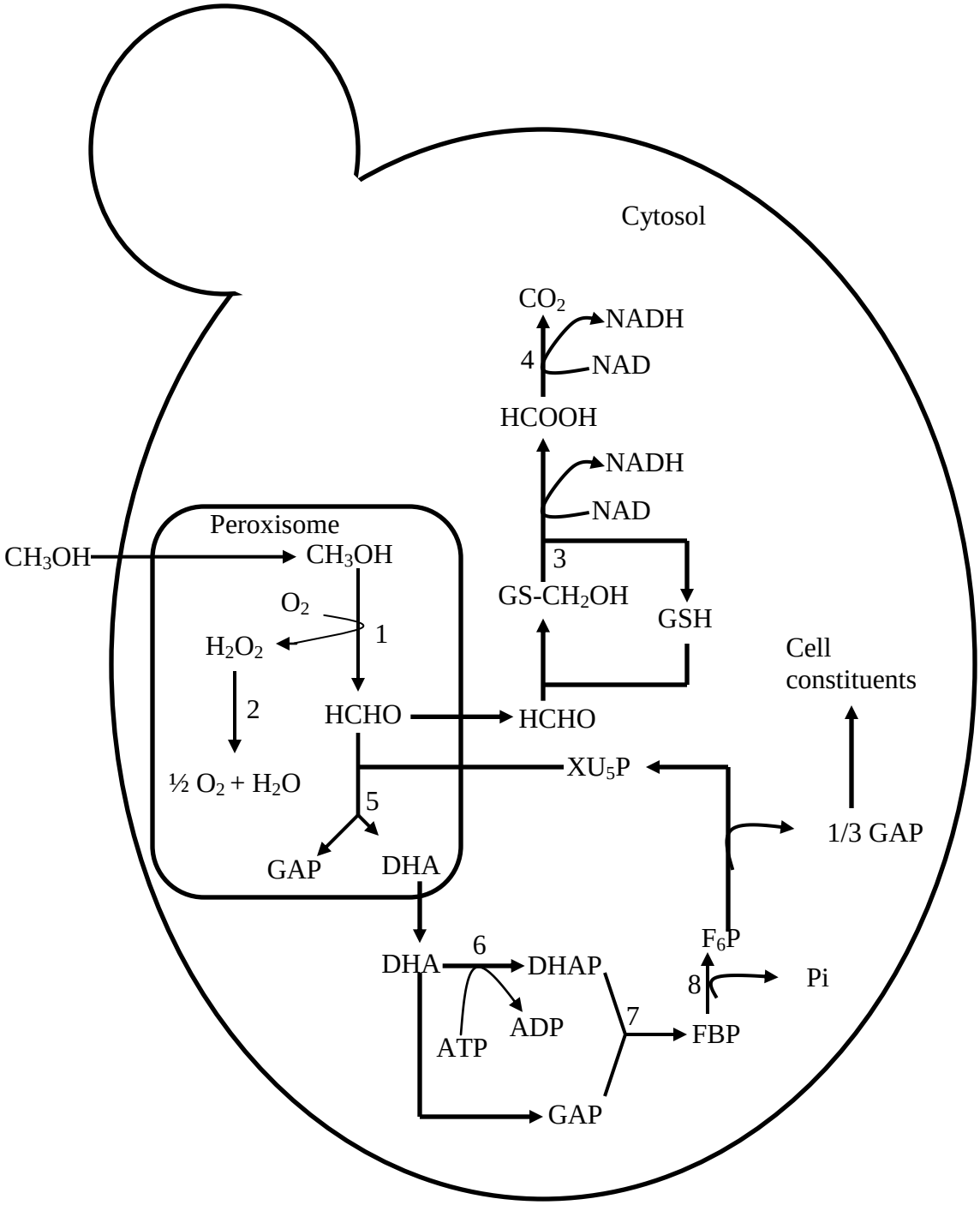


Figure 1.1 The methanol pathway in yeasts. Enzymes involved: 1, alcohol oxidase; 2, catalase; 3, formaldehyde dehydrogenase; 4, formate dehydrogenase; 5, dihydroxyacetone synthase; 6, dihydroxyacetone kinase; 7, fructose 1,6-bisphosphate aldolase; 8, fructose 1,6-bisphosphate. Acronyms: DHA, dihydroxyacetone; DHAP, dihydroxyacetone phosphate; FBP, fructose 1,6-bisphosphate; FGP, fructose 6-phosphate; GAP, glyceraldehyde-3-phosphate; GSH, reduced glutathione; Xu_5P, xylulose 5-phosphate.

methylotroph as an exceptional choice for recombinant protein production (Lin Cereghino, Sunga, et al., 2001).

The advantages of this new host system attracted interest from many researchers involved in heterologous protein production. The well known growth characteristics and inexpensive culture media are common advantages of yeast systems. However, one subtle advantage as a result of growth conditions observed with *P. pastoris* is the below optimum pH of the medium not ideal for contamination by many organisms and the presence of methanol which also makes cultures less susceptible to contamination (Cregg, Digan, et al., 1989). The ability to express heterologous proteins intracellularly at high levels was observed, but more important was *P. pastoris'* facility to secrete proteins into the medium at high levels and carry out posttranslational modifications such as glycosylation, disulfide bridge formation, and proteolytic maturations (Cregg, Vedvick, and Raschke, 1993). Unlike *S. cerevisiae*, which prefers to ferment and as a consequence succumbs to the toxic effects of ethanol, *P. pastoris* is a poor fermentor, which in turn permits cultures to grow to high cell densities. Because the amount of heterologous protein found in the medium is generally proportional to the amount of cells in the culture, this is an important characteristic. In addition, yeast cells secrete very little endogenous protein which allows for easier downstream purification methods since a secreted protein product is virtually the only protein in the culture medium (Cereghino, and Cregg, 2000). Genetic stability in the form of integrative plasmids and easy scale-up from shake flask to large-scale fermentation makes scale up straight forward and smooth in most situations (Romanos, 1995). Benefiting from the availability and marketing of expression kits sold by Invitrogen Corporation and licensed by Research Corporation Technology (RCT), the *P. pastoris* expression system has evolved into a major system for the production of heterologous proteins.

1.4 Promoters

The cloning and characterization of promoter elements from *Pichia pastoris*, particularly the *AOX1* promoter, was a key element responsible for the increased popularity of the yeast as a suitable expression system host. Promoters contain specific DNA sequences that are recognized by a set of regulatory proteins called transcription

factors. These factors bind to the promoter sequences and work together with a transcriptional complex, which includes RNA polymerase, to direct the level of transcription of a particular gene (Smale, and Kadonaga, 2003; Watson, Gilman, et al., 1992).

Yeast promoters are composed of at least two distinct elements, the TATA element and Upstream Activating Sequences (UASs), both of which are required for transcription. The highly conserved TATA element (TATAAA) is believed to be involved in the mechanism for the accurate initiation of transcription by RNA polymerase II. This sequence is usually located about 35 to 55 nucleotides upstream from the start site (Chen, and Struhl, 1985; Struhl, 1985). The second set of regulatory elements, UASs, appear to play a more defined role in determining a promoter's specificity. Usually located hundreds to thousands of nucleotides upstream of the region of transcription, the UASs respond to specific physiological signals and act as "enhancers" in transcription. Even though their regulatory functions are important in transcription, they are not always part of the promoter as defined by the RNA polymerase binding site (Guarente, 1984).

1.4.1 A strong and tightly regulated promoter: P_{AOX1}

The ability of the methylotrophic yeast *P. pastoris* to utilize methanol as a sole carbon source for growth has been discussed in section 1.3, in this manuscript. The cloning (Ellis, Brust, et al., 1985) and complete characterization of this tightly-regulated promoter (P_{AOX1}) (Cregg, Madden, et al., 1989) gave rise to initial expression vectors using the promoter for regulated foreign gene expression. Its regulation is based on a repression/induction mechanism that is dependent on a carbon source (glucose/glycerol for repression) and methanol for induction (Tschopp, Brust, et al., 1987). This regulation is a great advantage when the foreign gene of interest is toxic to the cells. Yeast cultures containing the expression vector can be grown to high cell densities on glucose without the expression of a toxic protein (repression) and then induced on methanol when the cells have reached a desired OD for expression. However, there are situations when tight regulation of a promoter is not necessary and the use of methanol induction poses more of a problem rather than a solution to gene expression. Such a problem can be

anticipated during a large scale fermentation production where methanol poses a potential fire hazard. In this case, an actively strong and constitutive promoter driven by an inert carbon source would be the ideal alternative for recombinant gene expression.

1.4.2 Strong and constitutive promoters: P_{GAP}, P_{TEF1}

Glyceraldehyde-3-phosphate dehydrogenase is an enzyme that is highly expressed in both non-repressive and repressive media (Waterham, Digan, et al., 1997). Its function involves glycolysis and gluconeogenesis in yeast. These characteristics indicate that the promoter driving expression of the enzyme is constitutive and highly active. Cloning and characterization of the GAP promoter by Waterham et al., 1997, demonstrated this to be the case. Expression studies, using β-lactamase as a reporter protein, confirmed the activity of the constitutive promoter as comparable in strength to P_{AOX1}. The expression of β-lactamase was higher in the P_{GAP}-controlled, glucose-grown cells compared to the expression in the P_{AOX1}-controlled, methanol-induced cells. When glycerol was the carbon source, the P_{GAP}-controlled expression of β-lactamase was similar to the methanol-induced cells under control of the *AOX1* promoter (Waterham, Digan, et al., 1997).

The translocation elongation factor-1 (*TEF1*) gene codes for the most abundant soluble protein in eukaryotic cells. Therefore, the probability of a strong promoter driving the expression led to the cloning and characterization of the *P. pastoris TEF1* promoter (Ahn, Hong, et al., 2007). Expression studies in glucose conditions revealed that P_{TEF1} showed similar expression levels of a reporter protein when compared to P_{GAP}. Interestingly, the activity of the reporter protein was found earlier in a *TEF1*-driven construct as opposed to a construct making reporter protein from mRNA products from the *GAP* promoter. This suggested that P_{TEF1} was a "more actively constitutive" promoter than P_{GAP} (Ahn, Hong, et al., 2007). Either promoter provides an excellent alternative for the high level expression of a foreign protein under constitutive conditions.

1.4.3 Strong and regulated promoter: P_{FLD}

The glutathione-dependent formaldehyde dehydrogenase (*FLD1*) gene encodes an essential enzyme that is required for the metabolism of methanol as a carbon source and

certain alkylated amines as a nitrogen source. The major function seems to be the conversion of formaldehyde to nontoxic intermediates, during methanol or methylamine metabolism (Shen, Sulter, et al., 1998). Genes that are under control of the *FLD1* promoter (P_{FLD}) can be induced under conditions similar to P_{AOX1}. In fact, the expression of the reporter gene β-lactamase showed similar levels when compared to P_{AOX1} to express the protein. This promoter provides another alternative with the use of methylamine, a nitrogen source, as an inducing agent.

1.4.4 Moderate and regulated promoter: P_{PEX8} and P_{YPT1}

In situations where the expression of genes at a high level induces poor secretion, mislocalization or improper folding of the recombinant protein (Thill, Davis, et al., 1990), a moderate or low expressing promoter may be more favorable. This is when the *PEX8* or *YPT1* promoters are indispensable tools for gene expression. The *PEX8* protein is an essential component in the peroxisome biogenesis pathway for the import of matrix proteins. The low level of *PEX8* RNA transcripts found in glucose–grown cells, as well as a moderately induced level in methanol or oleate-grown cells, indicates that P_{PEX8} is a weaker alternative to P_{AOX1}(Liu, Tan, et al., 1995). If low and constitutive expression is preferred, then the promoter for *YPT1* may fulfill the criteria. The *YPT1* gene encodes a GTPase involved in the secretory pathway. It is suitable for expression in media containing glucose, methanol, or mannitol as carbon sources (Sears, O'Connor, et al., 1998).

1.5 Integration of plasmid DNA

The transformation of yeast with plasmid DNA was developed by Hinnen et. al., (1978) with the theory that integration occurred by homologous recombination. Supporting research show that plasmid integration stimulates mitotic recombination through a DSB or mismatch repair (MMR) mechanism such as at a single-gap strand or denatured DNA region (Silberman, and Kupiec, 1994; Zgaga, Chanet, et al., 1991). A plasmid with two regions homologous to the yeast genome was shown to integrate at either locus when circular molecules were used for transformation. Transformation with

circular plasmid DNA resulted in 0.1-1.0 transformants per microgram of DNA per kilobase of homology per 10^8 cells (Orr-Weaver, Szostak, and Rothstein, 1981). Once a DSB was introduce by restriction enzyme digestion into a homologous region on a transforming plasmid, all integration occurred at that genomic locus via crossing over. Stimulation of integration at the site of the DBSs increased by as much as 3000-fold with a concomitant increase in the frequency of transformation (Hinnen, Hicks, and Fink, 1978; Orr-Weaver, Szostak, and Rothstein, 1981). In addition, when a double-strand gap was created using two restriction enzymes that cut within the homologous yeast sequences on the plasmid, the gap was always found to be repaired during integration (Szostak, Orr-Weaver, et al., 1983) implying that the genomic homolog sequence was used as a template for repair. Therefore, plasmid integration results from a reciprocal recombination event between the homologous regions on the plasmid and chromosome (Orr-Weaver, and Szostak, 1985). These landmark studies promoted the characterization of the mechanism involved in homologous recombination.

1.6 Recombination

1.6.1 Homologous recombination

Homologous recombination (HR) is a mechanism that controls the exchange of genetic information between homologous DNA sequences in an organism (Aguilera, Chavez, and Malagon, 2000). HR is essential during meiosis for proper chromosome segregation and preservation of genetic diversity (Aylon, and Kupiec, 2004b). It is necessary during mitosis for the repair of different types of DNA lesions such as mismatches (Prado, Cortes-Ledesma, et al., 2003) or double-strand breaks (Resnick, 1976; Szostak, Orr-Weaver, et al., 1983) generated either by exposure to free radicals, chemicals, radiation, or by action of nucleases(Aguilera, Chavez, and Malagon, 2000),or by plasmid integration(Hinnen, Hicks, and Fink, 1978).

1.6.2 Holliday, Meselson and Radding models

Robin Holliday (1964) proposed a molecular mechanism that described DNA-strand exchange to account for gene conversion and crossing-over events during meiosis

in fungi (Meselson, and Radding, 1975). The model predicted that crossing-over begins with a coordinated pair of single-strand nicks on homologous chromosomes followed by an invasion and exchanged of single strands (Figure 1.2). This model helped explain the presence of a Holliday structure, which contains a heteroduplex DNA and Holliday junctions (HJ). It also provided a reason why gene conversion events almost always had the flanking region of the homologous marker gene (Haber, Ira, et al., 2004). Holliday's model accounted for the formation of a symmetric heteroduplex DNA, whereas in the revised double strand break model proposed by Meselson and Radding, the formation of an asymmetric heteroduplex DNA better explained what researchers observed during the aberrant segregation found in yeast (Orr-Weaver, and Szostak, 1985) (Figure 1.3).

1.6.3 Double Strand Break

Homologous recombination is an important mechanism in DNA repair and the growing number of potential models of recombination that have been proposed have a common denominator in that recombination is initiated via a double-strand break (DSB) (Cromie, Connelly, and Leach, 2001; Orr-Weaver, and Szostak, 1983b; Orr-Weaver, and Szostak, 1983b; Paques, and Haber, 1999; Resnick, 1976; Szostak, Orr-Weaver, et al., 1983; Wang, Ira, et al., 2004). A DSB is a simple and frequently occurring break in duplex DNA and if left unrepaired can cause severe consequences to the organism. It can interrupt the coding sequence of a gene, disrupt the linkage between coding and regulatory sequences, alter chromosome segregation, and it can result in illegitimate and mutagenic DNA rearrangements. Repair of DSBs is essential to the preservation of genome integrity (Cromie, Connelly, and Leach, 2001).

1.6.4 Possible mechanism of recombination

Exactly how homologous recombination occurs has yet to be elucidated. However, it is well understood that HR requires one intact homologous DNA sequence to serve as a repair template partner (Aylon, and Kupiec, 2004a). Studies have shown that the mechanism involved in searching for homology encompasses a large region within the chromosome that interacts with the broken ends (Inbar, and Kupiec, 1999). Partner choice and type of recombination is dependent on the length and sequence of the region

adjacent to the break, the genomic location of the substrates (Kadyk, and Hartwell, 1992), the effect of mutations in genes involved in recombination and the mismatch repair

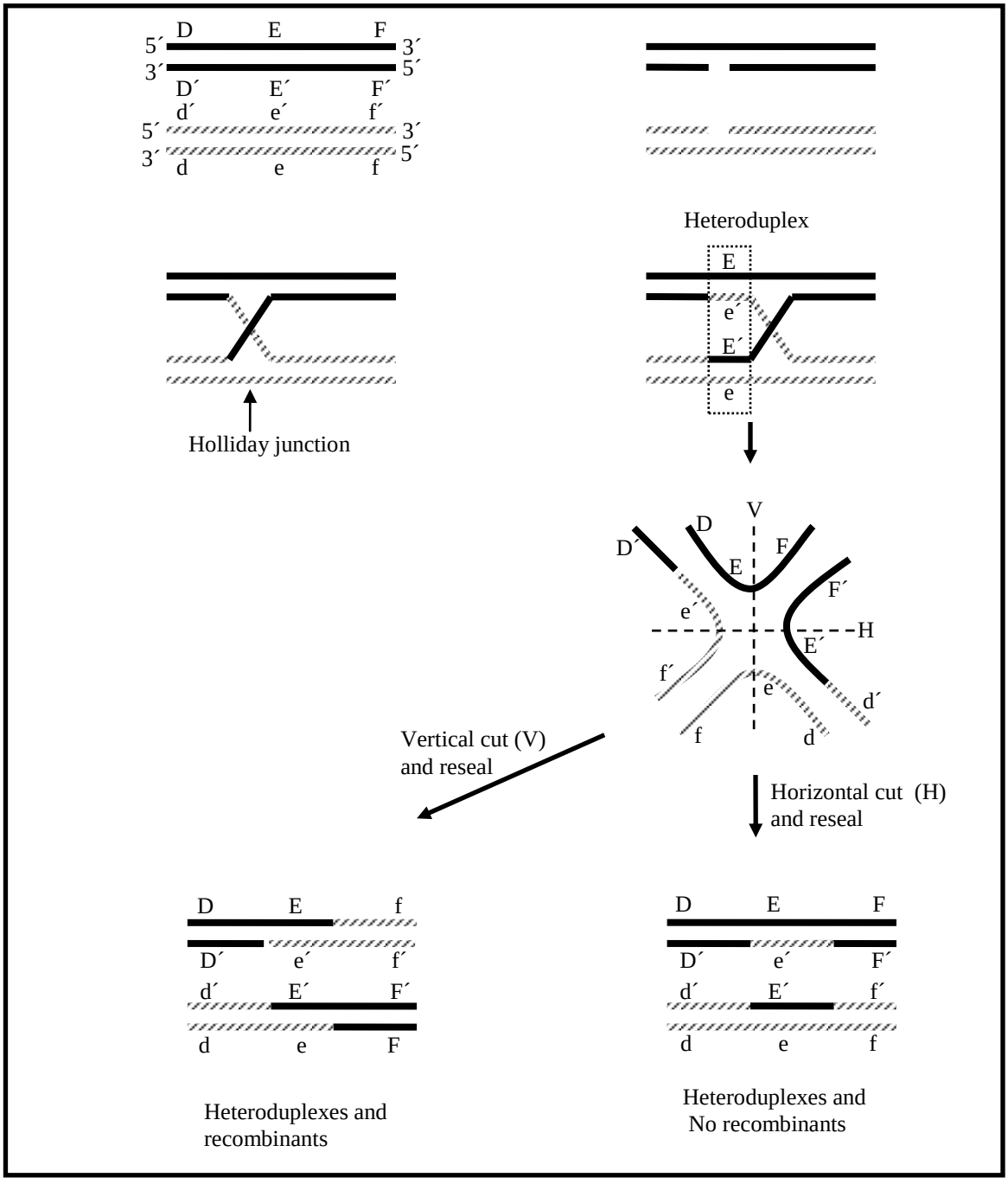


Figure 1.2 Holliday Model

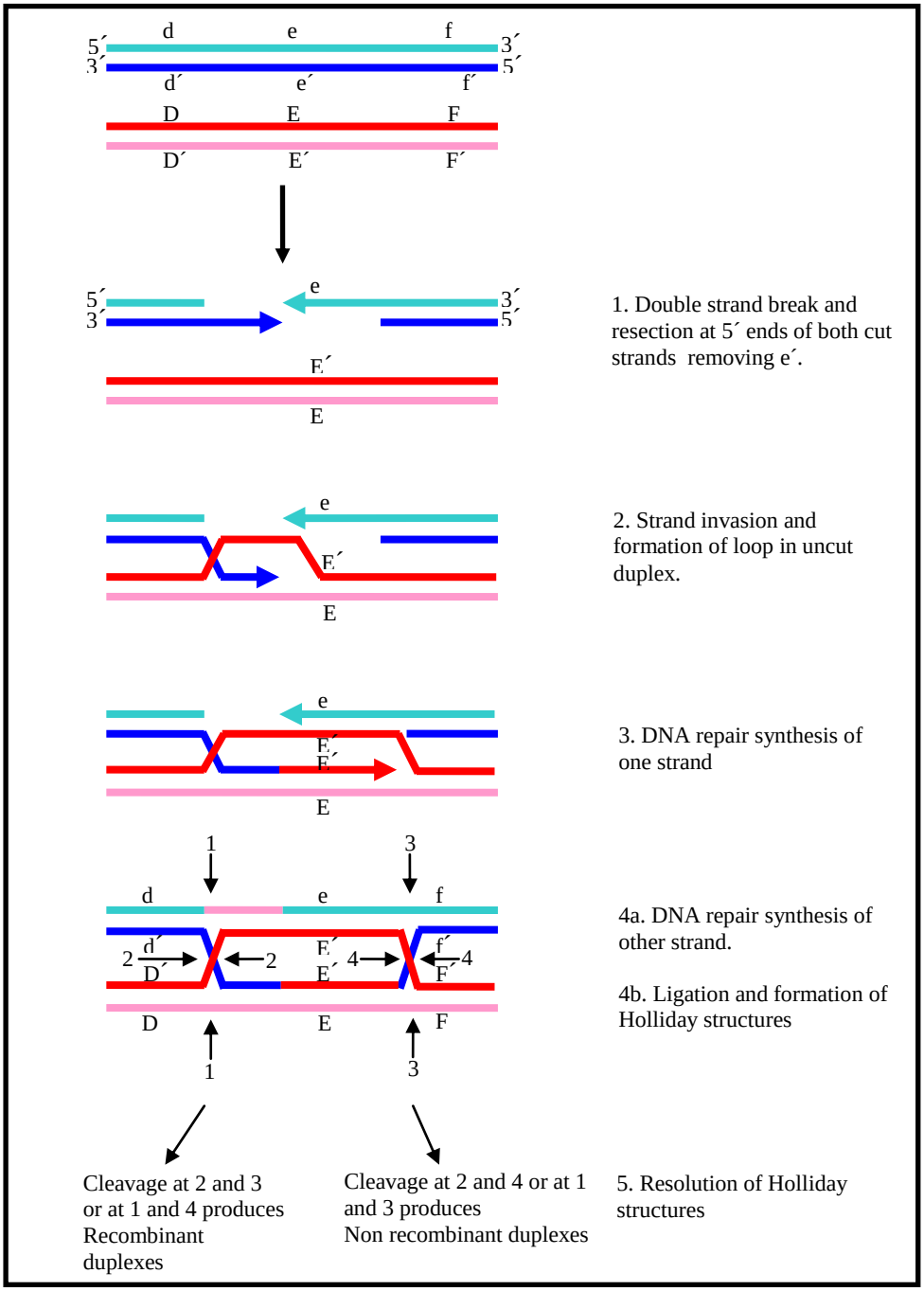


Figure 1.3 Double strand model

(MMR) system (Fasullo, Giallanza, et al., 2001), the type of lesion to be repaired; other factors such as chromatin structure may also play a role(Prado, Cortes-Ledesma, et al., 2003). To elucidate the mechanism of homologous recombination may first require a comprehensive understanding of the different elements involved. It is by no means a straight forward matching of sequences, but a variety of possible interactions that form the mechanism of homologous recombination.

Mitotic recombination can occur between different homologous partners; either allelic or ectopic (Kupiec, and Petes, 1988), direct-repeat or inverted-repeat (Prado, and Aguilera, 1995), and sister chromatid exchanges (Gonzalez-Barrera, Cortes-Ledesma, et al., 2003; Kadyk, and Hartwell, 1992). A dynamic mechanism for homology search is required to protect the integrity of the genome by preventing recombination between the many repeated sequences found in a genome (Inbar, and Kupiec, 1999). Once the partner is chosen, mitotic recombination can proceed through multiple mechanisms. These include double-strand break repair (Paques, and Haber, 1999) (Aylon, and Kupiec, 2004a; Szostak, Orr-Weaver, et al., 1983), synthesis-dependent strand annealing (Hastings, 1988; Nassif, Penney, et al., 1994), break-induced replication (Voelkel-Meiman, and Roeder, 1990) and single-strand annealing (SSA) (Prado, and Aguilera, 1995) methods. The occurrence of one recombination mechanism versus another depends on different elements, including the position of the homologous partners, the initiation event, and the length of homology between the molecules (Prado, Cortes-Ledesma, et al., 2003)

1.7 Gene Amplification

Gene amplification is the process in which a cell duplicates a gene or region in the genome, either in a tandem array or palindromic fashion (Haber, and Debatisse, 2006; Rattray, Shafer, et al., 2005). It is considered as an important component during genome evolution because it is associated with normal processes such as eggshell formation in Drosophila (Claycomb, and Orr-Weaver, 2005) as well as abnormal processes, such as

increased drug resistance (Alt, Kellems, et al., 1978; Paquin, Dorsey, et al., 1992; Schimke, 1986), and tumorigenesis (Lengauer, Kinzler, and Vogelstein, 1998). Depending upon the resulting product of gene amplification, it can either be viewed as important or detrimental to the organism. In such case as gene expression for recombinant protein production, the common view would be that gene amplification is an advantageous process. Thus it is of great interest to researchers to further understand this process.

Several models have been proposed for gene amplification; unequal sister chromatid exchange (Yandeau-Nelson, Xia, et al., 2006), localized over-replication (Hoy, Rice, et al., 1987), rolling circle replication (Khan, 1997), double rolling-circle replication (Watanabe, and Horiuchi, 2005), extrachromosomal amplification and reintegration (Liu, and Chang, 1994) and breakage-fusion-bridge cycles (Kaye, Melo, et al., 2004; Moore, Martin, et al., 2000). Unfortunately the molecular mechanism continues to elude scientists because there seems to be more than one mechanism involved in gene amplification to account for the many complex products formed. However, there is a consensus that supports the conclusion that DSBs initiate recombination and induce gene amplification (Lin, Lukacsovich, and Waldman, 1999; Paulson, Almasan, et al., 1998). Evidence also shows that a site-specific DSB promotes amplification near the break site (Pipiras, Coquelle, et al., 1998). This provides support to the belief that a DSB is a site with increase recombinogenic properties and will continue to recombine homologous sequences until the repair is made (Szostak, Orr-Weaver, et al., 1983).

When cells are transformed with a plasmid containing a DSB or gap within sequences homologous to the yeast genome, the frequency of transformation increases, homologous recombination into the genome occurs at the region of the DSB, and the gap is repaired from chromosomal information during plasmid integration (Orr-Weaver, and Szostak, 1983b; Orr-Weaver, and Szostak, 1985). Plasmid integration is vital in transformation because it provides the highest stability of a transforming DNA, but it is usually associated with the presence of low copy numbers within the genome. The use of targets such as transposable elements (ty) (Boeke, Xu, and Fink, 1988), ribosomal DNA (Lopes, Hakkaart, et al., 1991), 2µm circular plasmid (Morlino, Tizzani, et al., 1999)and

autonomous replication sequences (ARS) (Williamson, 1985), existing as multicopy elements in the genome have been successful approaches to generate high copy numbers during transformation. In rare instances, integration of up to 100 tandemly repeated copies has been reported (Gellissen, Weydemann, et al., 1992). While mammalian cells show a mixture of direct and indirect repeats as products of amplification, *E. coli* and yeast are normally associated with having direct tandem copies due to gene amplification (Stark, and Wahl, 1984; Windle, and Wahl, 1992).

As with many processes involved in homologous recombination and gene repair, the mechanism involved in creating multiple tandem repeats of linear plasmids integrated into the genome has been of great interest. Tandem repeats could result from the ligation of the plasmid molecules into a multimer before integration. However, this was excluded by Orr-Weaver and Szostak, 1983, in the transformation of a plasmid with an 800 bp gap. If multimers were first formed, then the 800 bp gap would still exist as part of the multimer. However they observed that every integrated copy was repaired and full length (Orr-Weaver, and Szostak, 1983a). Two hypotheses suggest that there are other alternative models for multimer formation. This could be explained in the sequential integration of plasmid molecules into the same site or a DNA replication error during integration.

The first hypothesis suggests that after first strand invasion, polymerization, and resolution and then plasmid integration, a recombinogenic lesion may be generated within the same site to stimulate the subsequent integration of another plasmid molecule (Orr-Weaver, and Szostak, 1983a; Silberman, and Kupiec, 1994). The second hypothesis suggests that a replication fork pauses and collapses generating a chromosome break. The DSB can then be processed into a new replication fork (Schacherer, de Montigny, et al., 2005). DSB formation is followed by resection of the ends (Sugawara, and Haber, 1992). The resulting 3'ends can then invade the template and provide a primer for new DNA synthesis, resulting in the restoration of the degraded single strands. Unfortunately, the mechanism for the generation of tandem repeats has still yet to be determined.

The DSB repair model has been the modified and accepted model for recombination and thus gene amplification. The DSB repair model postulates that resolution of the gene conversion occurs through the cutting of two Holliday junctions

which can be resolved either in crossover or with no crossover (Schwacha, and Kleckner, 1995). However, subsequent studies have found that this model did not explain why the rate of crossover associated with gene conversion was less than 50% (Orr-Weaver, and Szostak, 1983b; Paques, Leung, and Haber, 1998). Thus a series of models which do not require Holliday junctions have been proposed. One such model that supports current studies is called the synthesis-dependent strand annealing (SDSA) model (Nassif, Penney, et al., 1994). The basic assumptions in this model is that after strand invasion and new DNA synthesis, the newly synthesized DNA strands are unwound from the template, allowing the annealing of the two free 3' ends surrounding the DSB (Paques, Leung, and Haber, 1998) (Figure 1.4). This model suggests that DNA synthesis resulting from a DSB is highly activated to reproduce tandem copies and that large amplification is possible as a result of repairing a single DSB (Paques, Leung, and Haber, 1998).

1.8 *Pichia pastoris* is an excellent host system for the expression of recombinant proteins

First described in 1969 as a yeast capable of utilizing methanol as a sole carbon source for growth and energy (Ogata, Nishikawa, and Ohsugi, 1969), the fate of *Pichia pastoris* can be credited to the collaboration of Phillips Petroleum and the Salk Institute Biotechnology/Industrial Associates Inc. (SIBIA, La Jolla, CA, USA). This collaboration initiated the development of *P. pastoris* as a heterologous protein expression system (Cereghino, and Cregg, 2000). Within three decades, the use of this host for the production of heterologous proteins proved to be exceedingly successful. It can be seen in the increasing number of publications and proteins produced using *P. pastoris* as the system of choice (Cereghino, and Cregg, 1999; Cereghino, and Cregg, 2000; Cregg, Cereghino, et al., 2000; Lin Cereghino, Sunga, et al., 2001). With its popularity comes the necessity to expand the availability of promoters, selectable markers, and methods for the improvement of the expression system (Lin Cereghino, Lin Cereghino, et al., 2001; Sunga, and Cregg, 2004; Sunga, Tolstorukov, and Cregg, 2008).

P. pastoris' ability to grow to high cell densities, integrate vectors for genetic stability, secrete proteins into a relatively clean media, perform higher eukaryotic protein modification, and be easily manipulated with genetic techniques makes this yeast a

system of choice for heterologous protein production(Cereghino, and Cregg, 2000). The value of the many proteins produced in *P. pastoris* in laboratories world-wide give rise to an assortment of studies that focus on improving protein production. It is logical that new methods and tools that help promote a better and versatile host system will constantly grow as DNA technologies improve.

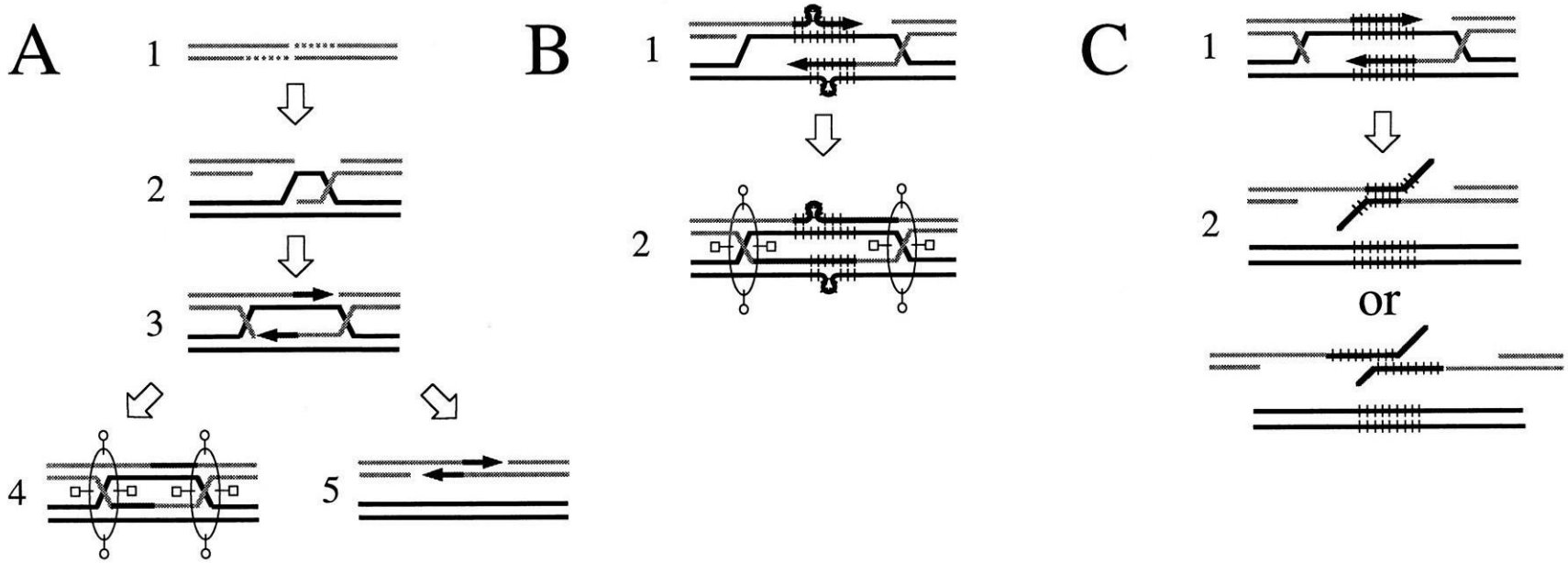


Figure 1.4 SDSA model

CHAPTER 2
Post-Transformational Vector Amplification in the Yeast *Pichia Pastoris*

2.1 Introduction

The methylotrophic yeast *Pichia pastoris* is the best known of a select number of yeast species developed for the synthesis of recombinant proteins (Ilgen, Cereghino, and Cregg, 2004). Others include baker's yeast Saccharomyces cerevisiae, Kluyveromyces lactis, and two other methylotrophic yeasts, *Hansenula polymorpha* and *Pichia methanolica* (Gellissen, 2005). As single-celled microorganisms, they are well suited for foreign gene expression. Like prokaryotes, they rapidly grow on inexpensive media and are easily manipulated at the molecular level. However, because they are eukaryotes, recombinant proteins that are expressed in them are exposed to a proper eukaryotic folding environment as well as certain eukaryotic posttranslational modification events such as proteolytic processing, disulfide bridge formation and glycosylation (Cereghino, and Cregg, 1999).

Methylotrophic yeasts such as *P. pastoris* have several advantages over other yeast systems for expression (Cereghino, Cereghino, et al., 2002). One important advantage is the existence of a tightly regulated and highly productive promoter. This promoter is derived from the alcohol oxidase I (*AOX1*) gene of *P. pastoris* and has been thoroughly described in other reports (Cregg, Madden, et al., 1989; Ellis, Brust, et al., 1985). Another major advantage of *P. pastoris* for expression is its capacity to reach high-cell densities (>100 g L^{-1} dry cell weight) in fermentor cultures (Brierley, 1998; Cereghino, and Cregg, 1999). A third feature that makes this expression system an attractive choice is the relatively low levels of endogenous protein that are secreted into the medium (Cereghino, and Cregg, 2000). One needs to merely centrifuge out the cells and the recombinant product is already purified to a significant extent. These

characteristics have facilitated the popularity of the system for the production of recombinant proteins (Cereghino, Cereghino, et al., 2002).

However, these features do not ensure that all or even most genes will be expressed at a high level, initially. A common technique to increase expression levels is to enrich for strains with multiple copies of an expression vector by selecting for transformants that are resistant to high levels of a selectable marker compound. Generally, transformed strains contain only one or a few integrated copies of an expression vector, thus limiting transcription of an expression cassette (Clare, Rayment, et al., 1991; Mansur, Cabello, et al., 2005; Vassileva, Chugh, et al., 2001). Two methods have evolved to further increase copy numbers of vector (Sunga, and Cregg, 2004). The first involves performing multiple transformations of the host using a different selectable marker for each sequential transformation. Although reliable, this approach is relatively slow and laborious, and the number of copies is limited to the number of selectable markers available. The second and most commonly utilized method takes advantage of a selectable marker system that allows the user to screen for strains containing multiple copies of an expression vector. The product of the selectable marker gene typically confers resistance to an antibiotic. Furthermore, the resistance gene product is of a type that is sensitive to the concentration of the drug, so that host strains that contain more of the resistance product are resistant to higher levels of the drug (e.g., Zeocin, G418). Thus, transformant populations that are resistant to high levels of the drug are enriched in strains that contain an increased number of copies of the vector. This method, although effective, is still laborious and inefficient. Only a small percentage (<5%) of highly drug-resistant colonies are the result of an increased number of vector copies in a transformed colony. The majority of transformants are resistant to the drug for other (unknown) reasons. Therefore, to succeed with this approach, there is a need to screen large numbers of high-drug-resistant colonies to find a few that are resistant due to increased vector copy number. Fifty to one hundred of transformants are typically screened to have a reasonable chance of finding the 1–2%'jackpot' (>10 copies) clones (Romanos, Scorer, et al., 1998).

We have discovered that one can select for strains with multiple copies of an expression vector from a single-transformed cell line long after the original

transformation event. With an appropriate drug-resistant selectable marker vector, one can readily enrich for strains with an increased number of vector copies by plating cells on agar medium containing a high concentration of the selection drug. Relative to identifying multicopy strains immediately after transformation, this method is much easier because it generates many more high copy number colonies and a greater proportion of these highly drug-resistant colonies are resistant due to increased copy number.

We describe here our scheme to amplify the number of vector copies post transformation and demonstrate this posttransformational vector amplification (PTVA) method using the selectable marker gene for Zeocin resistance.

2.2 Materials and methods

2.2.1 Strains, media and growth conditions

The wild-type strain was *P. pastoris* NRRL Y-11430 (Northern Regional Research Laboratories, US Department of Agriculture, Peoria, IL). The mutant strain GS115 (his4) is described (Cregg, Barringer, et al., 1985). Classical genetic manipulations of *P. pastoris* strains were performed as described (Cregg, Shen, et al., 1998). Bacterial recombinant DNA manipulations were carried out in an Escherichia coli strain, Top 10 (Invitrogen, Carlsbad, CA). Yeast strains were cultured in a rich YPD medium (1% yeast extract, 2% peptone, and 0.4% glucose) or a minimal medium (YNB), which consisted of 0.17% yeast nitrogen base without ammonium sulfate and amino acids, a carbon source (0.4% glucose or 0.5% methanol) and a nitrogen source (0.5% ammonium sulfate). YPD plates made for Zeocin selection were adjusted to pH 7.5. All yeast cells were grown and induced at 30 °C in shake flask (200 r.p.m.) cultures or on plates. Amino acids were added to 50 μg mL^{-1} as required. Escherichia coli strains were cultured in Luria broth medium supplemented with either 100 μg mL^{-1} ampicillin (Sigma, St. Louis, MO), 50 μg mL^{-1} kanamycin (Sigma), or 50 μg mL^{-1} Zeocin (Invitrogen), as required. The vectors pK321 (Shen, Sulter, et al., 1998), and pPIC9K (Invitrogen) were used in this study.

2.2.2 *Pichia pastoris* transformation

Up to 5 μg of SacI-digested plasmid DNA was used for electroporation of *P. pastoris* host strains using a BTX ECM630 electrocellmanipulator. Conditions for the pulsing included a cuvette with a 2-mm gap and a sample volume of 40 μL, a charging voltage of 1500 V, a resistance of 129 Ω, a field strength of 7500 kV cm^{-1} and a pulse length of 5 ms (Cregg, 2007). Immediately after pulsing, 500 μL of cold 1 M sorbitol was added directly into the cuvette, followed by the addition of 500 μL of YPD medium. The entire suspension was then transferred into a 1.5-mL mini centrifuge tube. Cells were then allowed to recover for about 4 h in a shaking incubator at 30 °C at 100 r.p.m. and then plated on selection plates.

2.2.3 Single colony enrichments

After electroporation, transformants were plated on either Zeocin (100 μg mL^{-1}) or another selective plate and incubated at 30 °C. Single colonies were then streaked onto a YPD master plate. This streaking process of single colonies was repeated three times.

2.2.4 Biochemical assays

Protocols for the β-lactamase assay were from EMD Biosciences (San Diego, CA). In the plate assay, yeast strains initially grown on YPD plates were spotted onto YNB plates with 0.5% methanol as carbon source and incubated overnight at 30 °C. The substrate for β-lactamase was Nitrocefin (EMD Biosciences). A working solution was prepared by dissolving 1 mg of Nitrocefin in 100 μL dimethylsulfoxide (DMSO) and 1.9 mL phosphate buffer (100 mM phosphate buffer, pH 7) to produce a 2 mL total volume of 500 mg mL^{-1} (1 mM) solution. The working solution of Nitrocefin was diluted fivefold for the plate assay or 10-fold for the liquid assay in buffer (0.1 M phosphate; 1 mM EDTA, pH 7.0). For the plate assay, 4 μL of substrate was spotted on each colony to initiate the reaction. The colony was observed for 5 min and a substrate color change from yellow to red indicated β-lactamase activity. For liquid β-lactamase activity assays, yeast strains were grown in shake flasks at 30 °C in YPD and then induced in YNB medium containing 0.5% methanol. Cultures were harvested and cell-

free extracts were prepared using the glass bead disruption method as described (Waterham, Digan, et al., 1997). Protein concentrations were determined using the Pierce BCA protein assay kit (Rockford, IL) with bovine serum albumin as a standard. β-Lactamase activities were measured in cell-free extracts. A working concentration of 2 ng mL^{-1} total protein was used for the enzyme assays. A change in $A_{486\ nm}$ was measured for a total of 10–15 min in 30-s intervals. As a positive control and standard, purified β-lactamase (Sigma) was used.

2.2.5 Vector copy number analysis

Chromosomal DNA from *P. pastoris* transformants was prepared using the YeaStar™ Genomic DNA kit (Zymo Research, Orange, CA). PCR fragments were generated using oligonucleotides that amplified specific targets in the expression cassette. A 935-bp PCR fragment of the *AOX1* gene promoter was produced using primer AOX3'-935 (5'-cgaataattagttgtttttt-3') and primer AOX5'-1a (5'-agatctaacatccaaagacg-3'). A 374-bp Zeocin resistance gene PCR fragment was produced using primer Zeo 5' (5'-atggccaagttgaccagtgcc-3') and primer Zeo 3' (5'-tcagtcctgctcctcggccac-3'). The BioPrime DNA-labeling kit (Invitrogen) was used to prepare biotinylated probes for Southern blots. Up to 300 ng of probe was used per hybridization. Southern blotting techniques were as described (Ausubel, Brent, et al., 2001b). Hybridization and detection techniques were performed as described (Sunga, and Cregg, 2004).

For copy number estimation, chromosomal BglI-digested DNA (2 μg) extracted from each initial low copy number predrug selection strain, as well as from high copy number strains were electrophoresed through agarose. Serially diluted aliquots of the chromosomal DNA from the putative high copy number strains were also electrophoresed through agarose for Southern blot analysis.

2.2.6 Miscellaneous methods

Screening for strains with increased drug resistance was performed by spotting (10^4–10^5) yeast cells suspended in sterile water onto agar plates containing YPD plus Zeocin (0.1, 0.5, 1, 1.5, 2.0, 2.5 or 3.0 mg mL^{-1}) or G418 (0.5, 1.0, 1.5 or 2.0 mg mL^{-1}), or by streaking the strains onto these plates using sterile toothpicks. Recombinant DNA

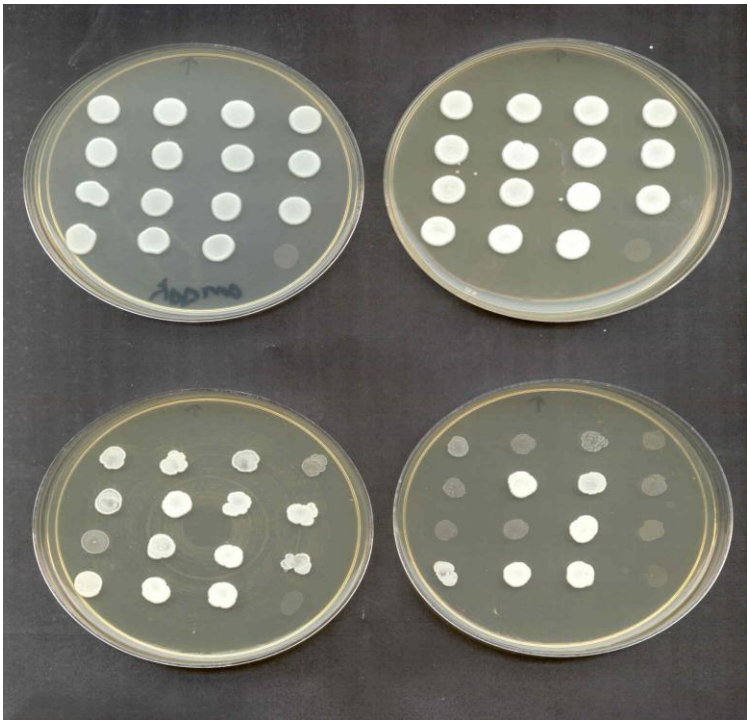
methods were performed essentially as described (Ausubel, Brent, et al., 2001a). Transformations of *P. pastoris* were performed using electroporation as described (Cregg, and Russell, 1998). DNA sequencing was performed by the Davis Sequencing Center (Davis, CA). PCR amplifications were performed as described (Kramer, and Coen, 2000). Restriction enzyme digestions were performed as recommended by the supplier (New England Biolabs, Beverly, MA).

2.3 Results

2.3.1 Construction of the test strain

To test the PTVA method, we constructed a *P. pastoris* strain transformed with the P. pastoris–E. coli shuttle vector pK321 (Shen, Sulter, et al., 1998). This vector, a derivative of the pPICZ series, contains the Zeocin resistance gene as the selectable marker for transformations of both *P. pastoris* and *E. coli* hosts and an expression cassette composed of the *AOX1* promoter and a β-lactamase gene. The presence of the β-lactamase expression cassette made evaluation of expression rapid and straightforward either qualitatively on plates or quantitatively in liquid samples. Before transformation, the vector was linearized with SacI, which cleaves the vector within the *AOX1* promoter sequences and aliquots of 5 μg of the linearized vector were electroporated into standard samples of competent cells of *P. pastoris* wild type. Transformants were selected at a low Zeocin concentration (100 μg mL^{-1}), which typically results in transformants that have one copy of a vector. Several Zeocin-resistant colonies were selected from the transformation plate for further use. Each selected colony was first streaked onto a YPD plate for single colonies. This single colony selection process was repeated three times for each transformant to be sure that each was the progeny of a single cell.

Samples of each single-copy strain were spotted on YPD plates containing selected concentrations of Zeocin (0.2, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 mg mL^{-1}). None of the spots showed colonies resistant to 2 mg mL^{-1} of Zeocin or higher. However, most colonies were resistant to 500 μg mL^{-1} Zeocin. These colonies were selected and spotted a second time on the Zeocin selection plates. This process was repeated two to three



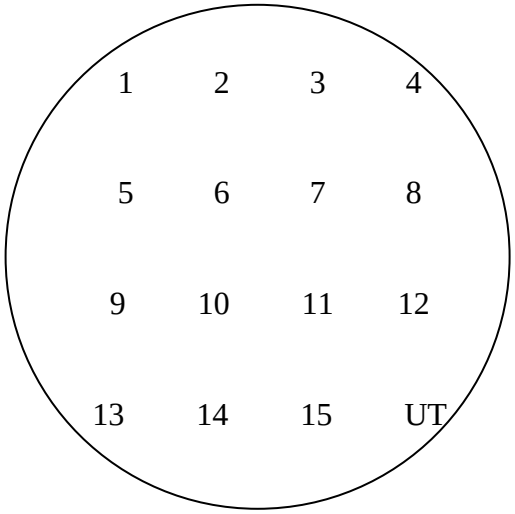


Figure 2.1 Selected *Pichia pastoris* strains after undergoing the PTVA-enrichment process. The key is shown on the right. Strains shown are: UT, untransformed WT strain (NRRL-Y11430) (not resistant to Zeocin); 6, 7, 11, 13, 14, 15, are highly enriched strains and therefore highly resistant to high concentrations of Zeocin.

times until strains showing resistance to Zeocin at 2 mg mL^{-1} were apparent (Figure. 2.1). Five hundred of these PTVA-enriched strains were collected for further analysis.

To determine which PTVA-selected colonies may have multiple vector copies, samples of each strain were spotted on YNB methanol plates for about 14 h and then β-lactamase activity solution was spotted on each yeast colony. With this solution, colonies expressing β-lactamase turn a pink to red color with colonies expressing higher levels of β-lactamase showing a darker more intense color (Figure 2.2). Of our PTVA-subjected clones, 40% showed a three- to fivefold increase in vector copy number and recombinant protein over the parent strain. So called 'jackpot' clones with >10 copies of the expression vector represented 5–6% of selected clones after PTVA and showed an increase in recombinant protein level proportional to their copy number.

2.3.2 Examination of structure of vectors in putative multicopy strains.

A major concern of this method of amplification was that a significant proportion of the amplification events would result in vector copies that were partially deleted or that were integrated into the *P. pastoris* genome at locations other than the original vector. To examine this possibility, genomic DNA was extracted from several high-level Zeocin-resistant clones as well as the pre-PTVA parent controls, digested with the restriction enzyme BglI and probed via Southern blot using defined regions of the pK321 test vector (the *AOX1* promoter and the Zeocin resistance gene) as labeled probes. For the control (DNA from the untransformed strain) probed with the *AOX1* promoter fragment, two bands of the expected size (0.9 and 4 kb) were observed (Figure 2.3a). Integration of a single copy of the pK321 vector resulted in four bands: the 0.9- and 4-kb bands seen in the untransformed strain and two additional bands with sizes of 1.3 and 1.5 kb (Figures 2.3b and 2.4a). The two additional bands result from the integration of the vector, which contains a copy of the *AOX1* promoter fragment. Multiple direct head-to-tail integrations of the vector at the same locus were predicted to show the same four bands in these Southern blot profiles; however, the two bands of vector origin would increase in intensity relative to the genomic fragments as the number of vector copies increased (Figures 2.4a and b). A deletion or addition to any of the vector copies, or integration of a vector copy into the *P. pastoris* genome at a location other than the *AOX1* promoter, or

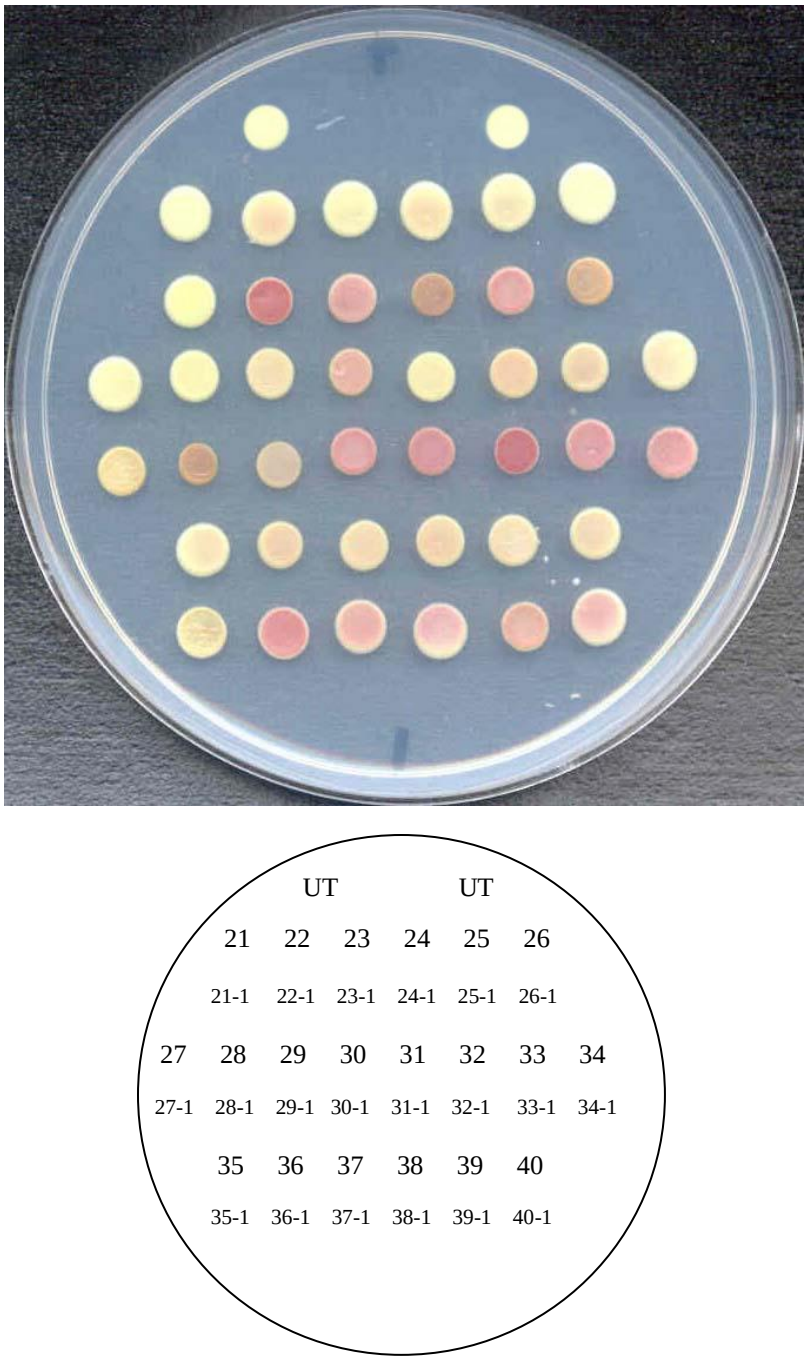


Figure 2.2 β-Lactamase plate assay. Photographs of pre-enriched and enriched/selected strains verified by β-lactamase plate assay. The key is shown on the right. Strains shown are: UT, untransformed WT strain (NRRL-Y11430); 1–40, PTVA-enriched strains denoted by '−1'. Yellow strains have no or very little expression of β-lactamase, while dark red strains have high concentrations of β-lactamase expressed.

integration of a copy of the vector at sequences other than the *AOX1* promoter was expected to result in the generation of bands different from these original four bands.

As can be seen in Figure 2.4a, the DNAs from high-level Zeocin-resistant colonies all showed exactly the same four hybridization bands in each sample except that the *AOX1* vector bands were often significantly more intense than the bands from the genomic copy. In Figure 2.4b, when DNAs from the same Zeocin-resistant clones were probed using Zeocin resistance gene sequences, a similar pattern of increased intensity in the enriched strain relative to the parent strain was seen. These results suggested that the amplification mechanism resulted in additional copies of the entire vector in every case and not just a portion of the vector. In addition, the Southern pattern indicted that the additional vector copies must be integrated into the *P. pastoris* genome at the original *AOX1* promoter locus and in a direct repeat head-to-tail fashion (Figures 2.3b and 2.4).

To determine the number of vector copies present in genomic DNAs of our high-level Zeocin-resistant colonies quantitatively, we performed a series of Southern blot experiments in which we compared the intensity of a band present in the single-copy parent strain to that from the multicopy strains. Before electrophoresis, each genomic DNA sample was digested with the restriction enzyme BglI. To estimate the number of vector copies in each strain, we prepared a series of known dilutions of each genomic DNA digest and subjected them to Southern blot analysis, utilizing the labeled *AOX1* promoter fragment as probe. Figure 2.5a, lane 1 is DNA from 2 μg of initial clone 31. Lane 6 is DNA from 2 μg of PTVA-enriched clone 31-1. Lanes 2–5 contain the dilution series of the same DNA. From these Southern blots, sample 31-1 appeared to contain c. 30 (+ or −10) vector copies and sample 55-1 appeared to contain c. 25 (+ or −10) copies. Enzyme assays for β-lactamase on these same clones produced comparable results, i.e., that β-lactamase levels were proportionately higher in each PVTA clone (Table 2.1). Analysis of these clones by the dilution test and the β-lactamase enzyme assay confirmed that this method results in gene amplification.

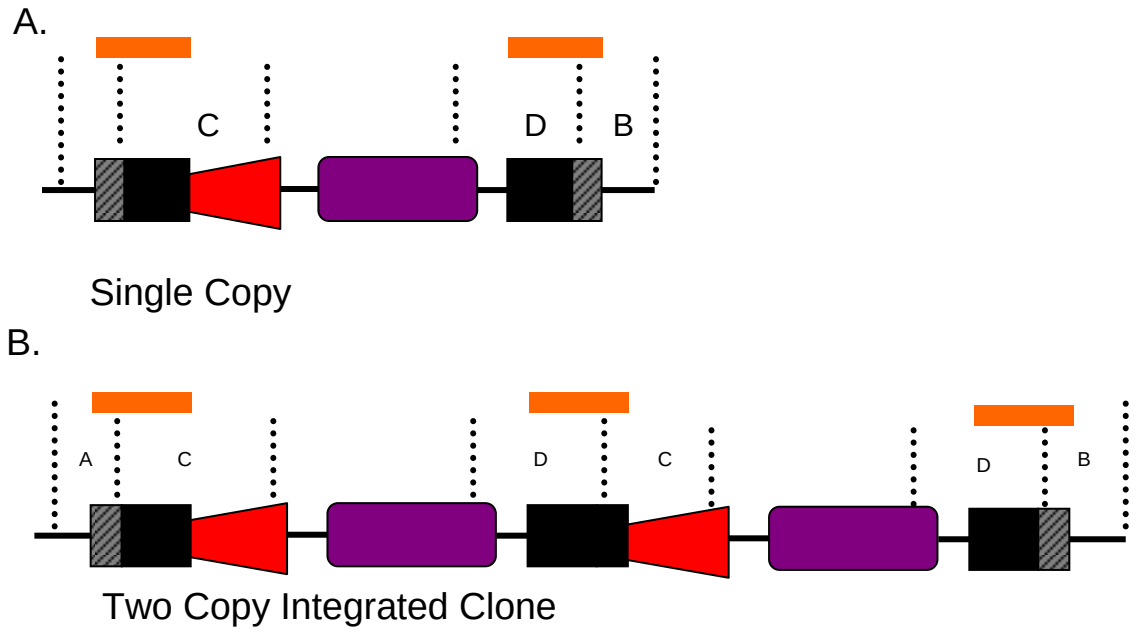


Figure 2.3 Diagram of products of integration of one copy (a) or two copies (b) of pPICZB-β-lactamase into the *AOX1* promoter locus of the *Pichia pastoris* genome.

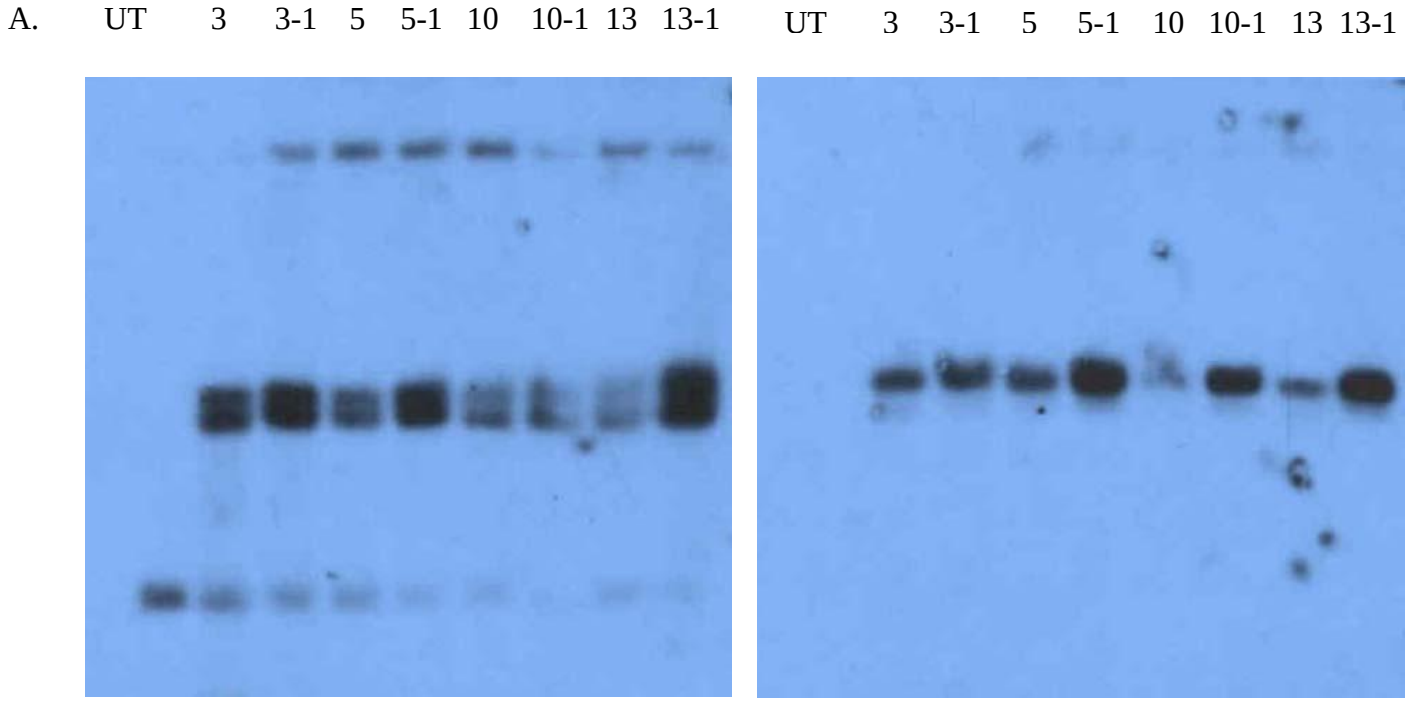


Figure 2.4 Southern blot showing genomic DNA from selected *Pichia pastoris* strains transformed with vector pPICZB-β-lactamase. Labeled DNA fragments containing sequences from the 5'-*AOX1* promoter (a) and labeled DNA fragments containing sequences from the Zeocin gene (b) were used as probes to compare and validate whole plasmid amplification. Lanes contain genomic DNA digested with BglI from each strain. Strains shown are: untransformed control (UT); 'enriched' strains are denoted by the parent number followed by '−1', while 'pre-enriched' equivalent strains are denoted just by a number.

2.3.3 Selection of multicopy strain with a G418 resistance marker by PTVA

Finally, to determine if this selection scheme was somehow specific for the Zeocin resistance marker gene selection or could be applied to other drug resistance marker systems, a PTVA-enrichment experiment was conducted on a *P. pastoris* strain transformed with the G418-resistance vector pPIC9K (data not shown). As observed with the Zeocin-based vectors, the G418 resistant strains showed a similar amplification after PTVA selection. Thus, it appears that the PTVA method is not specific to the Zeocin resistance marker system but is likely applicable to any selectable marker system in which a higher copy number can be selected by increased drug resistance.

2.4 Discussion

Analogous to methotrexate selection in mammalian cells, the PTVA method works best by generating resistance in a stepwise approach (Schimke, 1986). An initially high-concentration of drug applied immediately after transformation usually results in a low frequency of resistant strains with low copy numbers of the selected sequences, while the posttransformation multiple-step selection tends to induce the recovery of highly resistant cells with multiple copies of these sequences.

While we do not know the molecular details of the process(es) by which these amplification events occur, we have made key observations about the process in this work. The first observation is that the amplification process appears to occur naturally in a small percentage of cells in virtually any vector containing strain. The second observation is that the PTVA process leads to a considerable increase in copy number of the entire vector and not just portions of it. This is clearly important if a uniform recombinant product is desired. Thirdly, the Southern blot data demonstrate that the copies must be inserted into the *P. pastoris* genome in the same location as the original copy and in a head-to-tail configuration. As shown here, the result of the PTVA process can be a considerable increase in the level of recombinant protein production.

A variation of the homologous recombination mechanism is believed to elicit the generation of multiple copies of the transforming vector into the yeast genome (Plessis, and Dujon, 1993; Saffran, Smith, and Chan, 1991). It has been shown by several groups

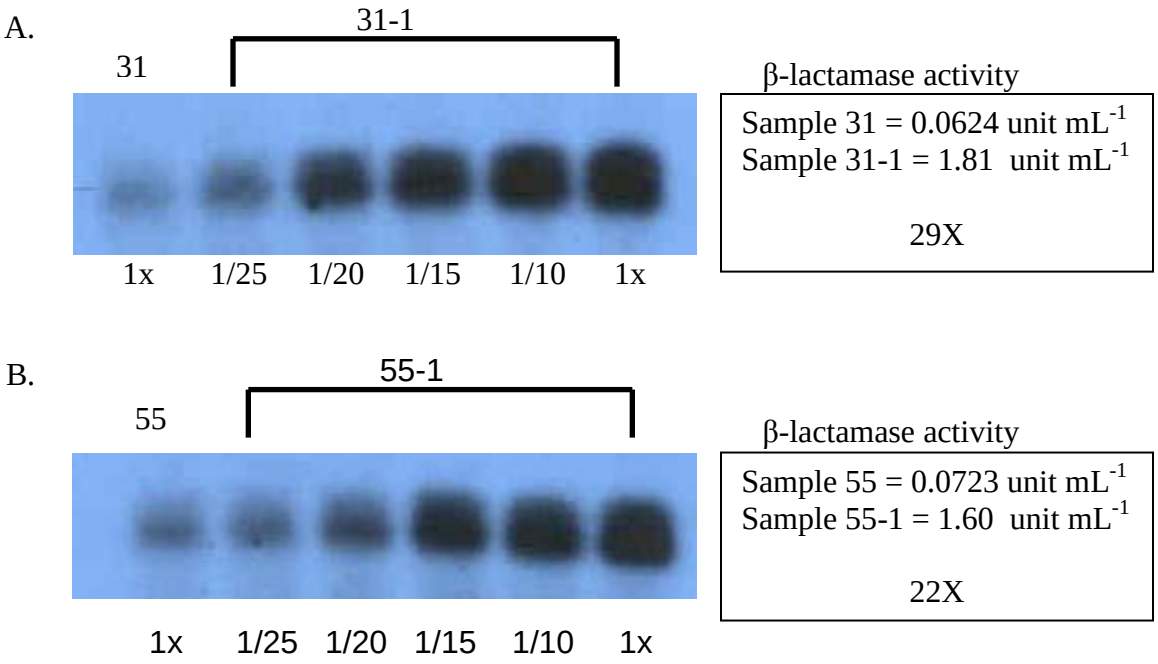


Figure 2.5 Southern blots showing a dilution test to estimate increased copy number. Probe used was a labeled DNA fragments containing sequences from the 5'-*AOX1* promoter. Lanes contain 2 μg of genomic DNA digested with BglI from each strain. Strains shown are: (a) *P. pastoris* strain 31; (b) *P. pastoris* strain 55; lane 1, 'single colony' clone; lanes 2–6, diluted 'enriched' strains.

Table 2.1 Comparison of β-lactamase activity levels in selected *Pichia pastoris* strains between 'single copy' and 'enriched' PTVA clones containing the pPICZB-β-lactamase vector.

Strain Designation	Single Copy Lactamase units mL^{-1}	Enriched Copy Lactamase units mL^{-1}	Fold Increased in expression
3	0.0481	0.3877	8.1x
4	0.0365	0.4886	13.4x
5	0.0275	0.1174	4.3x
10	0.0326	0.1021	3.1x
13	0.0273	0.1029	3.8x
22	0.0258	0.3142	12.2x
25	0.0255	0.2008	7.9x
28	0.0170	0.1242	7.3x
31	0.0624	1.81	29x
32	0.0307	0.2318	7.6x
34	0.0157	0.1451	9.2x
36	0.0394	0.1296	3.3x

that one of the causes of increased drug resistance in yeast can be the spontaneous amplification of that resistance gene (Dorsey, Peterson, et al., 1992; Fogel, and Welch, 1982; Stark, and Wahl, 1984). Two mechanisms are hypothesized as allowing this to occur. Free plasmids, linear or circular, first undergo homologous recombination with each other and then with the yeast chromosome. The second mechanism involves an initial event of integration between the plasmid and the yeast chromosome, which is subsequently followed by a second round of integration with more free plasmids. Strong evidence supports the second mechanism as responsible for multicopy integration in yeast. In fact, this event can be readily stimulated in vivo (Orr-Weaver, Szostak, and Rothstein, 1981; Orr-Weaver, and Szostak, 1983a).

Relative to the standard practice of selecting for transformants with high drug resistance levels directly during transformation, the PTVA method has two strong advantages. First, selection for high-level drug resistance during transformation often results in few or no highly drug-resistant strains. In contrast, selecting for high-level drug-resistant strains by the PTVA method readily generates unlimited numbers of strains. Second, a higher percentage of strains resistant to high levels of drug contain multiple vector copies with the PTVA method. Although it is still necessary to screen the resulting high level drug resistant strains for ones with multiple copies, we found that by the PTVA method, c. 6% of enriched strains contained 10 or greater copies of vector each, as opposed to <1% of strains selected during transformations.

Finally, we showed that the PTVA method works with at least one other drug marker selection system, the kanamycin/G418 system, and therefore is not specific to Zeocin but most likely will work with any drug marker system that allows one to select for higher copy numbers by higher levels of resistance to the marker gene substrate (e.g., blasticidin, formaldehyde).

CHAPTER 3
THE *PICHIA PASTORIS* FORMALDEHYDE DEHYDROGENASE GENE (FLD1) AS A MARKER FOR SELECTION OF MULTICOPY EXPRESSION STRAINS OF *PICHIA PASTORIS*

3.1 Introduction

3.1.1 *Pichia pastoris* as a recombinant protein production system

Pichia pastoris is a major system for the efficient production of recombinant proteins(Cereghino, and Cregg, 1999; Cereghino, and Cregg, 2000) http://faculty.kgi.edu/cregg/index.htm). As a single-cell microorganism, it grows rapidly on inexpensive substrates and is easily manipulated. As a eukaryote, it can often correctly fold and process other eukaryotic proteins. *P. pastoris* has several key advantages over other yeast expression systems (Higgins, and Cregg, 1998). One advantage is that it has a tightly regulated and highly efficient promoter, obtained from the *P. pastoris* alcohol oxidase I gene (*AOX1*). This promoter stimulates high levels of transcription in cells cultured on methanol but is strongly repressed in cells cultured on other carbon sources (Cregg, Madden, et al., 1989; Tschopp, Brust, et al., 1987). A second important advantage is that *P. pastoris* readily grows to unusually high-cell densities (>100 g/l dry cell weight). This is possible because *P. pastoris* strongly favors a respiratory mode of metabolism over a fermentative mode (Cereghino, Cereghino, et al., 2002; Lin Cereghino, Sunga, et al., 2001). As a result, *P. pastoris* does not have a tendency to produce ethanol, an inhibitor of cell growth at high concentrations. A third advantage of *P. pastoris* for certain recombinant proteins is its capacity for secreting large amounts of these proteins free into the culture medium. Since *P. pastoris* secretes only low levels of its own proteins, secretion provides an effective method to separate recombinant proteins from the bulk of intracellular host proteins and other cellular materials (Romanos, 1995).

Despite the strength of the *P. pastoris AOX1* promoter, levels of most foreign gene products remain limited by transcription, primarily because these strains typically contain only one integrated copy of an expression vector. A number of methods have evolved to increase the number of expression vectors per strain. One method is to simply perform multiple rounds of transformation on a *P. pastoris* strain, utilizing a vector with a different selectable marker each round (Lin Cereghino, Lin Cereghino, et al., 2001); http://faculty.kgi.edu/cregg/index.htm). A second method is to introduce a vector that has been constructed with multiple copies of an expression cassette (Brierley, 1998). A third method is to utilize certain bacterial antibiotic resistance genes as markers and select for strains that show increased resistance to the antibiotic (Miles, Busser, et al., 1998; Romanos, Clare, et al., 1991). For *P. pastoris*, two such selectable marker genes have been described; the *Sh ble* gene (Zeo^R), from the bacterium *Streptoalloteichus hindustanus* which confers resistance to the bleomycin-related drug Zeocin (Higgins, Busser, et al., 1998) and the bacterial kanamycin resistance (Kan^R) gene, which confers resistance to the eukaryotic antibiotic G418 (Clare, Rayment, et al., 1991; Scorer, Clare, et al., 1994). With either of these markers, selection for *P. pastoris* transformants resistant to high concentrations of drug enriches for populations of transformed strains carrying multiple integrated copies of an expression vector. In many reported instances, these strains have been shown to yield dramatically higher levels of recombinant protein (Higgins, and Cregg, 1998; Romanos, 1995).

A disadvantage of these multicopy selection vectors is the presence of bacterial DNA sequences (one or more antibiotic resistance genes and a bacterial origin of replication) in each vector. These bacterial sequences are integrated into the *P. pastoris* along with the rest of the expression vector and represent a potential recombinant DNA hazard since the possibility exists that they may escape from the expression strain and spread back into the environment. To minimize this possibility, strict and expensive containment controls are placed on bioprocess facilities culturing recombinant expression strains. Clearly, a *P. pastoris* vector system that allows for the enrichment of multicopy expression vector strains but avoids the simultaneous introduction of functional bacterial sequences would be advantageous.

3.1.2 The formaldehyde dehydrogenase gene

P. pastoris is capable of utilizing methanol as sole carbon and energy source (Veenhuis, Van Dijken, and Harder, 1983) and certain alkylated amines (e.g., methylamine and choline) as sole nitrogen source (Zwart, Veenhuis, et al., 1980). A common intermediate in both pathways is formaldehyde, the toxic product of alcohol oxidase in the methanol pathway and amine oxidase in the methylamine pathway (Figure 3.1). The formaldehyde generated by either pathway can be further oxidized to formate by formaldehyde dehydrogenase (Fld1p) and then carbon dioxide by formate dehydrogenase. In the process, this oxidative pathway provides energy for the cell in the form of NADH and may also help the cell protect itself from the toxic effects of excess formaldehyde (Sibirny, Ubiyvovk, et al., 1990). Recently, we described the isolation and characterization of the Fld1p gene (*FLD1*) from *P. pastoris* (Shen et al., 1998). In addition, we described *P. pastoris fld1* mutants as defective in ability to grow on methanol as a carbon source or methylamine as a nitrogen source.

In this report, we show that *P. pastoris fld1* mutants have increased sensitivity to formaldehyde relative to wild-type cells and that the level of resistance to formaldehyde is generally proportional to the number of *FLD1* genes present in a strain. Based on this observation, we developed the *FLD1* gene as a novel selectable marker. Like the bacterial antibiotic resistant genes described above, the *FLD1* marker gene allows for multi-copy expression strain enrichment by selection for strains with increased levels of resistance to formaldehyde. However, unlike the bacterial genes, *FLD1* is a native *P. pastoris* gene and therefore, does not pose a potential biohazard problem. In addition, we constructed expression vectors containing the *FLD1* marker in which all functional bacterial sequences are removed prior to introduction of the expression vector in to the *P. pastoris* genome.

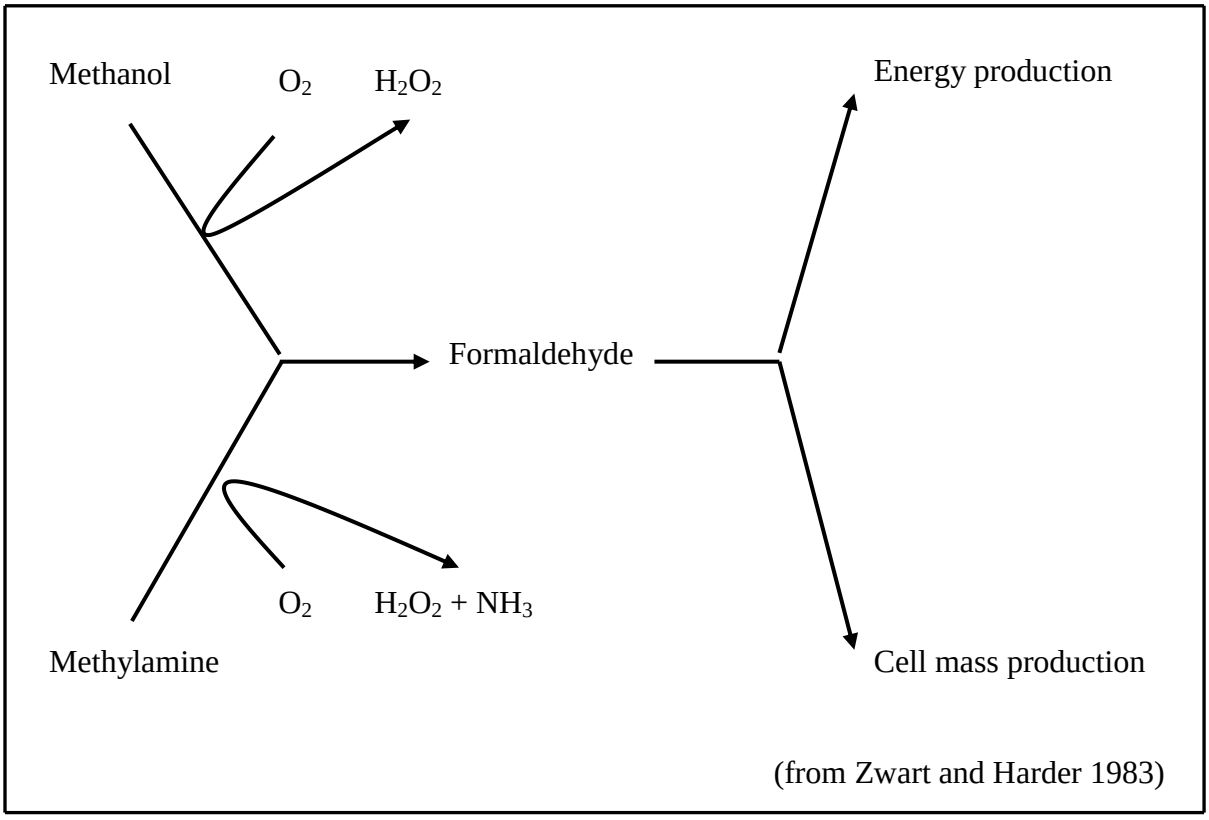


Figure 3.1 Metabolism of methanol and methylamine in yeast.

3.2 Materials and methods

3.2.1 Strains and media

The wild-type strain was *P. pastoris* NRRL Y-11430 (Northern Regional Research Laboratories, US Department of Agriculture, Peoria, IL). The mutant strains GS115 (*his4*) and MS105 (*fld1 his4*) were described previously (Cregg, Barringer, et al., 1985; Shen, Sulter, et al., 1998). Classical genetic manipulation of *P. pastoris* strains was performed as described (Cregg, and Russell, 1998). Bacterial recombinant DNA manipulations were carried out in *Escherichia coli* strains DH5α, GM2163 (New England Biolabs, Beverly MA), a *dam* methylase defective host, or Top 10 (Invitrogen, Carlsbad, CA). Yeast strains were cultured in a rich YPD medium (1% yeast extract, 2% peptone, and 0.4% glucose) or a minimal medium (YNB), which consisted of 0.17% yeast nitrogen base without ammonium sulfate and amino acids, a carbon source (0.4% glucose or 0.5% methanol) and a nitrogen source (0.5% ammonium sulfate or 0.25% methylamine chloride). Amino acids were added to 50μg/mL as required. *E. coli* strains were cultured in Luria broth medium supplemented with either 100 μg/mL ampicillin (Sigma Chemical, St. Louis, MO) or 50 μg/mL Zeocin (Invitrogen, Carlsbad, CA) as required.

3.2.2 Plasmid constructions

The construction of pJL-IX was initiated by a polymerase chain reaction (PCR) amplification of a DNA fragment containing the *P. pastoris FLD1* gene from plasmid pYG1 (Shen, Sulter, et al., 1998) using oligonucleotide primers AJS1 and AJS2 and inserting the resulting MfeI digested fragment into the EcoRI site of vector pGAPZ-B (Invitrogen, Carlsbad, CA) (see Table 3.1 for sequences of all oligonucleotides used in this study). The resulting vector, pJS2, contained the *FLD1* gene under the control of the glyceraldehydes-3-phosphate dehydrogenase gene (*GAP*) promoter (P_{GAP}). This vector was then used as template in a PCR reaction with primers AJS10 and AJS 21. These primers amplified a small fragment containing a portion of *FLD1* and simultaneously introduced a mutation (A→G) that destroyed an EcoRI site present in *FLD1*. The mutated fragment was digested with BanII and SphI and inserted back into pJS2 cut with

the same enzymes to create pJS2-RI. This plasmid was then used as a template in a PCR reaction with primers AJS8 and AJS9. The product of this reaction was a DNA fragment of 1735 bp that contained *FLD1* under control of P_{GAP} (P_{GAP}-*FLD1*) with restriction sites for BamHI and NcoI at its termini. For the next step, plasmid pHILD2 (Invitrogen) was digested with AfeI, which removed a DNA fragment containing the *P. pastoris HIS4* gene from the vector to create pHILD2-HIS. This vector was digested with NheI and a synthesized adaptor fragment containing restriction sites for BamHI, XhoI, and NcoI was introduced via insertion of two complementary oligonucleotides AJS24 and AJS25. The P_{GAP}-*FLD1* PCR fragment product was ligated into the BamHI and NcoI sites of this vector to create pJL-IX (Figure 3.2A). pJL-lacZ was created by inserting an EcoRI fragment containing the β-galactosidase gene (*lacZ*) gene into the EcoRI site of pJL-IX. The *lacZ* gene fragment was generated by PCR from the plasmid pGC180 using AJS30 and AJS31.

For the creation of pJL-SX, a NotI site present in the multiple cloning site (MCS) of pPICZα-B (Invitrogen) was first destroyed by NotI digestion followed by filling in the single-stranded termini using large fragment of Klenow polymerase (New England Biolabs, Beverly, MA) and dNTPs. A DNA fragment of 815 bp containing a portion of the *AOX1* 5' sequences, the *Saccharomyces cerevisiae* α mating factor leader sequence, MCS, c-myc and polyhistidine epitope tags, and a portion of the *AOX1* terminator fragment was then removed by cleavage with BclI and inserted into BclI digested pJL-IX to create pJL-SX (Figure 3.2B).

Other promoters used in this study were P_{AOX1}, P_{PEX8}, and P_{FLD1} as described in Lin Cereghino and Cregg (2000)(Cereghino, and Cregg, 2000). Test vectors containing the *FLD1* gene under control of these different promoters were constructed to study which promoter would best suit our purpose in the development of the selectable marker system. The initial vector pPICZB (Invitrogen) was digested with EcoRI and BglII to remove P_{AOX1}. The *PEX8* promoter fragment with BamHI and EcoRI sites was generated by PCR using oligos AJS5 and AJS 6. This fragment was then ligated into BamHI and EcoRI digested pPICZB vector to create pJS1. pJS1 was then digested with EcoRI and a DNA fragment containing the *FLD1* gene, generated by PCR using oligonucleotide primers AJS1 and AJS2, was inserted. The resulting vector pJS3

Table 3.1 Oligonucleotides used in this study

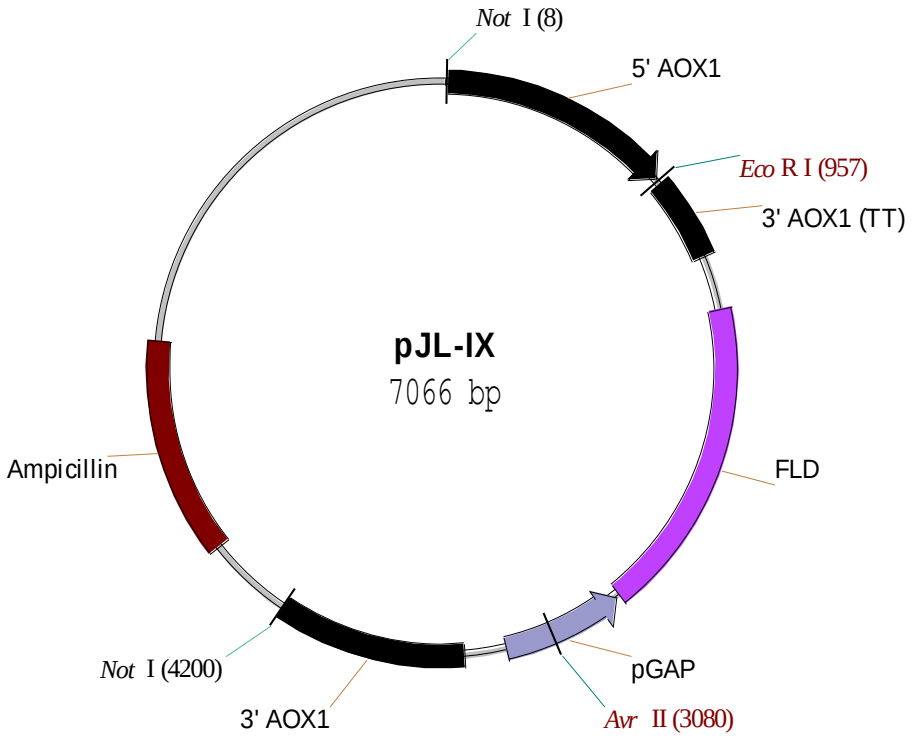
Name	Sequence 5'→ 3'
AJS 1	Cgcaattgatgtctaccgaaggtcaagt
AJS 2	Cgcaattgttagtgcatagtaatcacag
AJS3	Cgggatccgcatgcaggaatctctg
AJS 5	Gcacagggaataatgttgaa
AJS 6	Gccttatcgaatcctaagtt
AJS 8	Cgggatccagatctttttgtagaaatg
AJS 9	Catgccatggttagtgcatagtaatc
AJS10	Ggaattccatatgttagtgcatagtaatc
AJS21	cctccaagagcccatgaagttagagtgaaagtggagttcactggtg
AJS24	Ctaggggatccctcgagccatggc
AJS25	Ctaggccatggctcgagggatccc
AJS30	ccggaattcatgggggatcccgtcgttttac
AJS31	ccggaattcatggatttccttacgcgaaatacggg
AJS82	Atgtctaccgaaggtcaagt
AJS83	Aatcaagtagtcaacaatattgg

contained the *FLD1* gene under the *PEX8* promoter (P_{PEX8}-*FLD1*). To construct a vector with the *FLD1* gene under control of its native promoter P_{FLD1}, pPICZB was digested with BglII and EcoRI to remove the *AOX1* promoter fragment. Using primers AJS2 and AJS3, a PCR fragment containing P_{FLD1}-*FLD1* was generated and ligated into pPICZB to create pJS4. A final test vector, called pJS5, was created by inserting the *FLD1* gene into pPICZB digested only with EcoRI. This placed the *FLD1* gene under control of the *AOX1* promoter (P_{AOX1}-*FLD1*). These vectors were then used in initial studies to examine the ability of each promoter to express *FLD1* in a range useful for multicopy selection in response to formaldehyde concentration.

3.2.3 Biochemical methods

For enzyme assays, yeast strains were grown in shake-flask cultures at 30^{o}C in YPD or YNB medium supplemented with methanol to late logarithmic phase or until fully induced for growth on methanol (typically at least 6 hours after shift from YPD medium). Cultures were harvested and cell-free extracts were prepared using the glass bead disruption method as described in Waterham et al. (1992)(Waterham, Keizer-Gunnink, et al., 1992). Protein concentrations were determined using the Pierce BCA protein assay kit (Rockford, IL) with bovine serum albumin as standard. Fld1p activity was measured spectrophotometrically by following the rate of NADH formation at 340 nm in the presence of saturating amounts of formaldehyde, glutathione, and NAD (Schutte et al., 1976; Shen et al., 1998). ß-galactosidase levels were measured in cell-free extracts as described in Sambrook et al.(1989)(Sambrook, Fritsch, and Maniatis, 1989). As a positive control, purified β-galactosidase (Sigma Aldrich, St Louis, MO) was used at a concentration of 700 units/mg protein and 1.7 mg protein/mL. A working concentration of 2 ng/μL total protein was used for the assay. The optical density of the reactions was measured at a wavelength of 420 nm. Other details are as described in Sambrook et al. (1989).

A.



B.

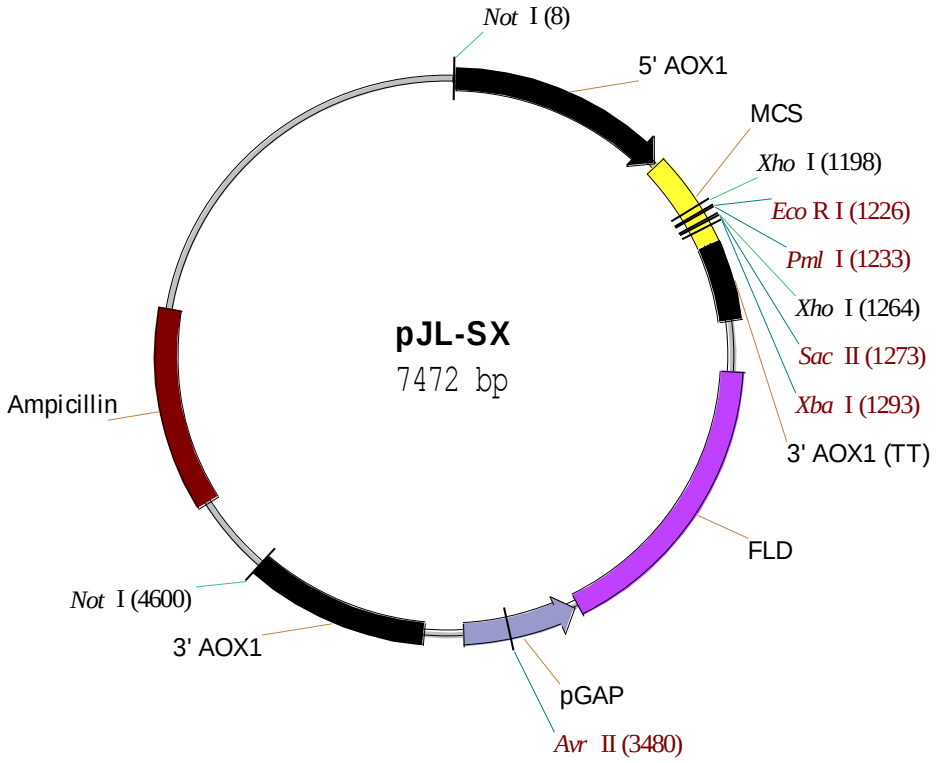


Figure 3.2 Diagram of *FLD1*-marker expression vectors pJL-IX and pJL-SX.

3.2.4 DNA analysis

Chromosomal DNA from *P. pastoris* transformants was prepared as described by Ausubel et al. (2000)(Ausubel, Brent, et al., 2001a). The labeled DNA probe used for the Southern blot is an 893-bp PCR fragment that was amplified using primers AJS82 and AJS 83 (Table 3.1). Biotin-labeled dUTPs to 1 mM (Molecular Probes Inc. Eugene, Oregon), unlabeled dATPs, dCTPs, dGTPs to 10 mM and dTTPs to 2 mM (New England Biolabs, Beverly, MA) were added for the amplification to label the probe. Preparation of *P. pastoris* genomic DNA samples and transfer onto nylon membranes was as described in Ausubel et al. (2000). A XL-1000 UV crosslinker from Spectronics Corporation (Westbury, New York) was used to crosslink the DNA onto the nylon membrane. After crosslinking, the membrane was rinsed with 0.25 M disodium phosphate pH 7.2. Warm (65^{o}C) hybridization solution was added to the blot and the blot was incubated at 65^{o}C in a hybridization chamber for one hour to overnight. The hybridization solution was prepared fresh before each experiment and contained 100 μL of 0.5 M EDTA, 17.5 mL of 20% SDS, 25 mL of 0.5 M disodium phosphate and 7.4 mL dH_2O for a final volume of 50 mL. The solution was filter sterilized prior to use. Following prehybridization, the membrane was further hybridized overnight in the hybridization chamber with 300-500 ng of a biotin labeled probe. Just prior to initiating hybridization, the probe was denatured by boiling in a water bath for ten minutes and then held on ice for three minutes. The probe was then added to hybridization solution (10 mL) and the mixture added to the membrane and incubated overnight in the incubation chamber. After hybridization, the membrane was washed as per instructions for the Tropix Southern Light Detection kit (Bedford, MA). Images on membranes were visualized following instructions provided with the Bright Star Bio Detect Nonisotropic Detection Kit (Ambion, Austin, TX) using Kodak X-Omat film.

3.2.5 Miscellaneous methods

Screening for multicopy strains was performed either by placing serial dilutions of potential multicopy strains onto YPD plates plus (5mM, 10mM, 15mM, 20mM, 25mM and 30mM) formaldehyde plates or by streaking the strains onto these plates using sterile toothpicks. Recombinant DNA methods were performed essentially as described in

Ausubel et al. (2000). Transformations of *P. pastoris* were performed by electroporation as described in Cregg and Russell (1998). DNA sequencing was performed by the Oregon Regional Primate Research Center, Molecular Biology Core Facility (Beaverton, OR) or the Davis Sequencing Center (Davis, CA). PCR amplifications were performed as described by Kramer and Coen (2000)(Kramer, and Coen, 2000).

3.3 Results

3.3.1 Preliminary testing of the FLD1 marker gene concept

To examine the potential of using the *P. pastoris FLD1* gene as a selectable marker for *P. pastoris* transformations, we first conducted a series of preliminary tests. The first test was to define conditions under which *P. pastoris* was sensitive to the presence of formaldehyde in an agar medium. For this, two strains were examined: GS115 (*his4*) (wild-type with respect to *FLD1*) and MS105 (*his4 fld1*) (a chemically induced mutant previously used by our laboratory to clone the *P. pastoris FLD1* gene; Shen et al., 1998). Fresh culture samples from these two strains were spotted onto sets of YPD plates containing formaldehyde at selected concentrations (0, 3, 5, 10, 15, 25, and 30 mM). After incubation, results showed that the wild-type strain was resistant to formaldehyde up to 5 mM but was sensitive to 10 mM or higher concentrations (Figure 3.3, spot 16). In contrast, the *fld1* mutant strain was sensitive to formaldehyde at concentrations of 3-5 mM or higher (Figure 3.3, spot 1). These results suggested that *P. pastoris* strains are sensitive to formaldehyde and, that a functional *FLD1* gene (present in the wild-type strain) confers a limited degree of resistance to formaldehyde.

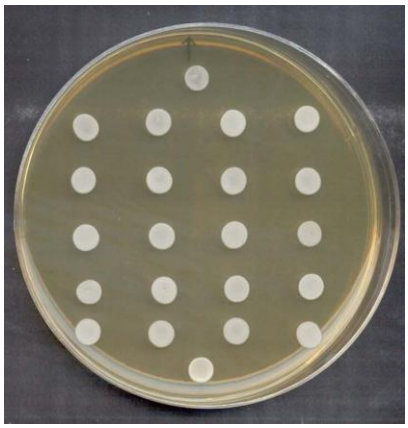
In a second set of tests, we examined whether an increasing number of copies of the *FLD1* gene conferred increasing resistance to formaldehyde in these same *P. pastoris* strains. For these studies, we inserted a DNA fragment with the *FLD1* gene (including enough 5' sequence to contain the promoter) into the *Sh ble* (Zeocin resistance)-based marker vector pPICZ-B to create pPICZ-*FLD1*. We transformed this vector (linearized by cleavage within the *AOX1* promoter region) into the two strains by selection for resistance to Zeocin. As we hypothesized, we observed a general correlation between

Zeocin and formaldehyde resistance levels in the transformants (data not shown). We then attempted to transform the two strains by direct selection for resistance to formaldehyde. Colonies resistant to selected levels formaldehyde were individually collected and examined for the presence of the vector as judged by their concomitant resistance to Zeocin. Although some strains were Zeocin resistant, most were not, presumably due to a high rate of spontaneous resistance to formaldehyde.

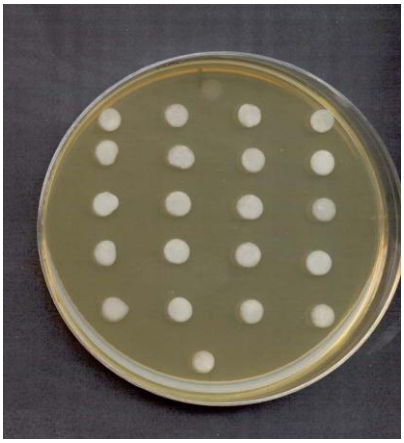
To improve the efficiency of transformation using the *FLD1* gene as a marker, the pPICZ-*FLD1* vector was transformed into the *fld1* strain by selection for restored ability to grow on minimal medium plates with 0.25% methylamine as sole nitrogen source (Mta^+ phenotype). With this selection method, 60% - 80% of Mta^+ colonies were also resistant to Zeocin, the same frequency observed for *P. pastoris* transformations with auxotrophic mutant hosts and vectors containing the complementing *P. pastoris* biosynthetic gene (Cregg, Barringer, et al., 1985; Lin Cereghino, Lin Cereghino, et al., 2001). The remaining Mta+ strains that do not have the vector appear to be the product of gene conversion events in which the vector-born wild-type copy of *FLD1* recombines into the *fld1* genomic locus without other vector sequences (Cregg, Barringer, et al., 1985).

A third and final set of preliminary studies were aimed at identifying a suitable promoter to control the *FLD1* gene, that is a promoter that expressed *FLD1* at a high enough level such that cells containing a single copy of the gene displayed measurable resistance to formaldehyde but low enough such that formaldehyde resistance levels were approximately proportional to the number of *FLD1* gene copies per cell. The suitability of promoters from each of the following four *P. pastoris* genes was examined: 1) P_{FLD1} (Shen, Sulter, et al., 1998); 2) the strong constitutive glyceraldehydes-3-phosphate dehydrogenase gene (*GAP*) promoter, P_{GAP} (Waterham, Digan, et al., 1997); 3) the strong methanol-inducible alcohol oxidase 1 gene (*AOX1*) promoter, P_{AOX1} (Ellis, Brust, et al., 1985; Tschopp, Brust, et al., 1987); and the weak methanol-inducible peroxin 8 (*PEX8*) promoter, P_{PEX8} (Liu, Tan, et al., 1995).

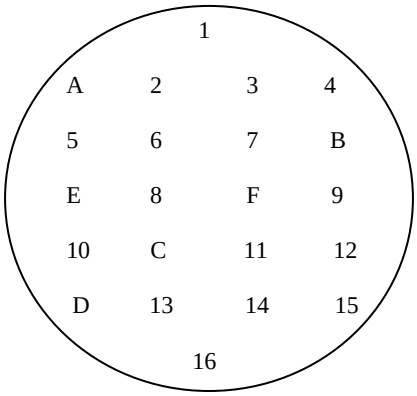
Each of these promoter-*FLD1* gene constructs was inserted into the Zeocin resistance gene vector pPICZ and the resulting vectors were linearized and transformed into the *fld1* mutant strain by selection for resistance to selected concentrations of Zeocin

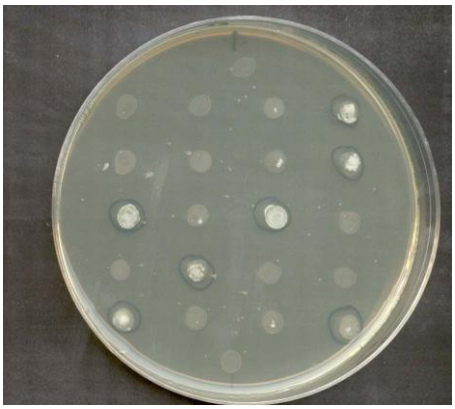


YPD + 0mM Formaldehyde



YPD + 5mM Formaldehyde





YPD + 20 mM Formaldehyde

Figure 3.3 Photographs of selected *P. pastoris* strains containing multiple copies of pJL lacZ on formaldehyde containing YPD plates. Key is shown at bottom left. Strains shown are: 1, MS105 (*fld1 his4*); A through F, selected multicopy strains also shown in Figure 3.5; 2-15 other pJL-lacZ-transformed strains not resistant to high formaldehyde levels; 16, GS115 (*his4*).

(2 mg/ml). Typically, the level of Zeocin resistance of a strain is approximately proportional to vector copy number (Miles, Busser, et al., 1998). Based on this, we examined our collection of Zeocin resistant transformants for level of formaldehyde resistance. We observed that strains transformed with vectors containing either the P_{AOX1}-*FLD1* or P_{PEX8}-*FLD1* showed little resistance to formaldehyde even in strains resistant to high levels (2 mg/ml) of Zeocin. However, Zeocin resistant strains that had been transformed with either the P_{FLD1}-*FLD1* or P_{GAP}-*FLD1* vectors showed levels of formaldehyde resistance that were approximately proportional to their levels of Zeocin resistance. Based on these results, we arbitrarily selected the P_{GAP}-*FLD1* marker for incorporation into our *FLD1* marker expression vectors.

3.3.2 Construction of FLD1-based expression vectors

As shown in Figures 3.2A and 3.2B, we designed and constructed two *FLD1*-based expression vectors: one, pJL-IX for intracellular expression and the other, pJL-SX, which included sequences encoding the *Saccharomyces cerevisiae* α mating factor pre pro signal for secretion of recombinant proteins. The overall design of these vectors is similar to a vector first described by our lab for the expression of the hepatitis B surface antigen gene in *P. pastoris* (Cregg, Tschopp, et al., 1987). pJL-IX is composed of sequences from 5' of the *AOX1* gene (P_{AOX1}), followed by a unique EcoRI site for insertion of foreign genes, followed by sequences from just 3' of the *AOX1* gene (the transcriptional terminator), followed by the P_{GAP}-*FLD1* selectable marker, followed by a DNA fragment from further 3' of the *AOX1* gene. This cluster of contiguous *P. pastoris* DNA fragments is attached by NotI sites to the bacterial plasmid pBR322 so the vector can be maintained and amplified in *E. coli*. The secretion vector pJL-SX is identical to pJL-IX except for the addition of the α mating factor signal sequences. Prior to transformation, the bacterial sequences in both vectors are removed by NotI digestion, leaving mostly *P. pastoris* sequences.

The fate of vectors of this general design upon introduction into *P. pastoris* was previously described in detail by Clare et al. (1991)(Clare, Rayment, et al., 1991). Briefly, 20 to 40 % of transformants are the result of gene conversion events in which just the marker gene has been inserted into the *P. pastoris* genome but without any

additional vector sequences. Another 10 to 20% of transformants have undergone a gene replacement event in which the vector sequences replace the *AOX1* gene (Figure 3.4A). These transformants must rely on the transcriptionally weak *AOX2* gene for Alcohol oxidase and as a result grow slowly on methanol relative to wild-type strains and are easily identified by this slow growth (Mut^s phenotype). The remaining transformants appear to be the result of a recircularization of the *P. pastoris* vector sequences via the NotI sites and then a simple single cross-over type insertion of the vector into the genome, most often at the *AOX1* 5' sequences but occasionally also at the *AOX1* 3' sequences (e.g., Figure 3.4B). Finally, most multicopy strains resulting from transformation with this type of vector contain head-to-tail concatemers of the *P. pastoris* vector all inserted at the same genomic locus.

3.3.3 Selection of multicopy expression cassette strains using a FLD1 marker based vector

We performed a series of studies on a collection of *P. pastoris fld1* strains transformed with a derivative of pJL-IX (pJL-lacZ). pJL-lacZ contained the bacterial β-galactosidase (β-gal) gene (*lacZ*) as a reporter inserted at the EcoRI site of pJL-IX under the control of P_{AOX1}. Prior to transformation, the vector was digested with NotI to remove bacterial vector sequences. As before, transformants were initially selected on minimal medium plates containing methylamine as sole nitrogen source. Individual colonies were collected and a portion of the cells from each colony was spotted onto plates containing formaldehyde at selected concentrations (0, 5, 10 20 mM). As shown in Figure 3.3 (spots B-F), some transformant colonies grew on 20 mM formaldehyde or higher (1-2% of Mta^+ colonies) and were collected for further study.

The number of copies of pJL-lacZ was estimated directly from total DNA extracted from cultures of each strain. For this, genomic DNAs were digested with EcoRV, separated by agarose gel electrophoresis, blotted to charged nylon membranes and hybridized with a labeled probe composed of an 883-bp fragment of the *P. pastoris FLD1* gene. In preliminary studies with genomic DNA from the untransformed parent strain, we determined that a band of 2.2 kb corresponded to the single-copy wild-type genomic *FLD1* gene (Figure 3.5, lane UT). Upon examination of pJL-lacZ transformed

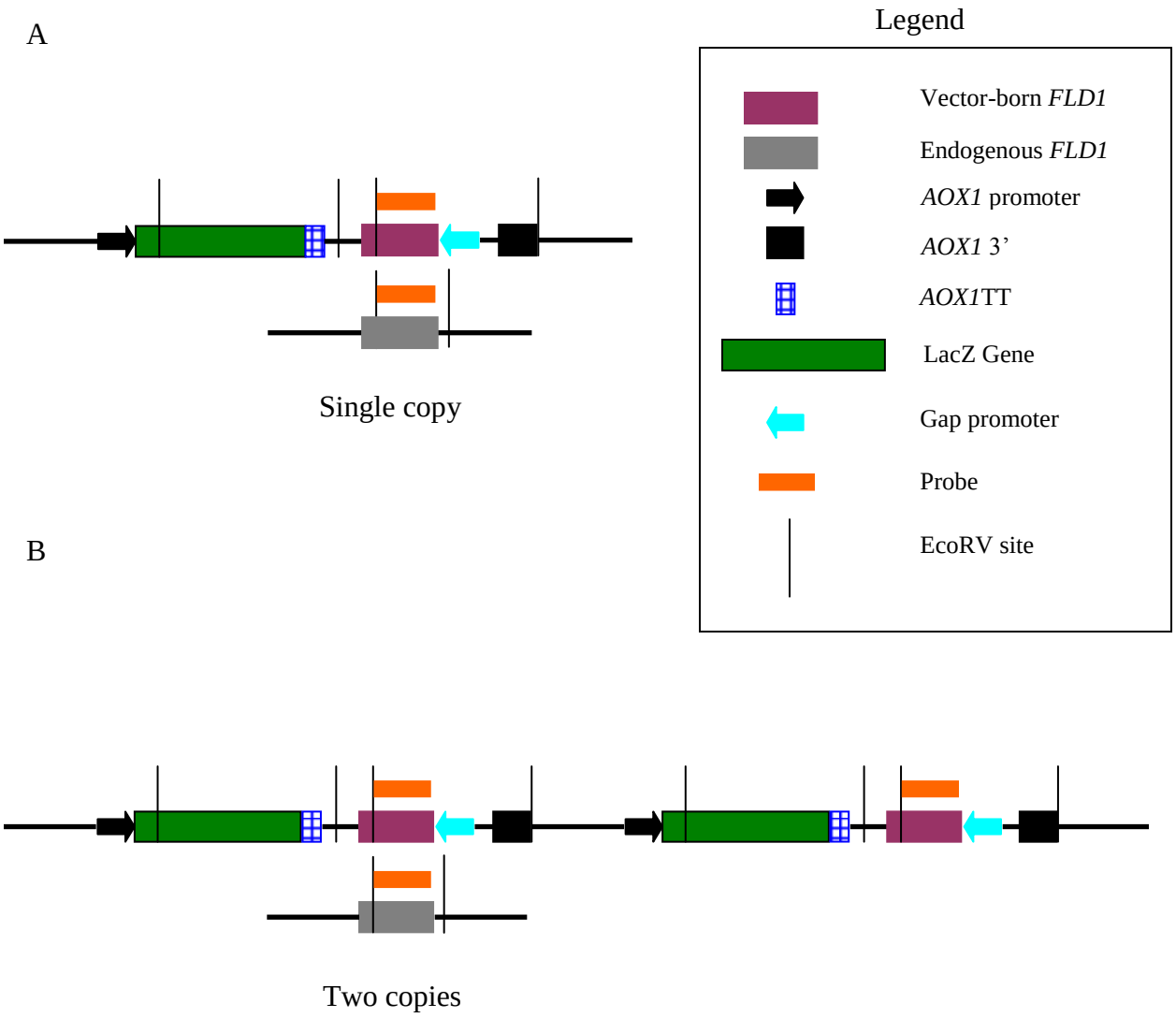


Figure 3.4 Diagram of result of integration of one (A) or more (B) copies of pJL-lacZ into the *AOX1* promoter locus of the *P. pastoris* genome.

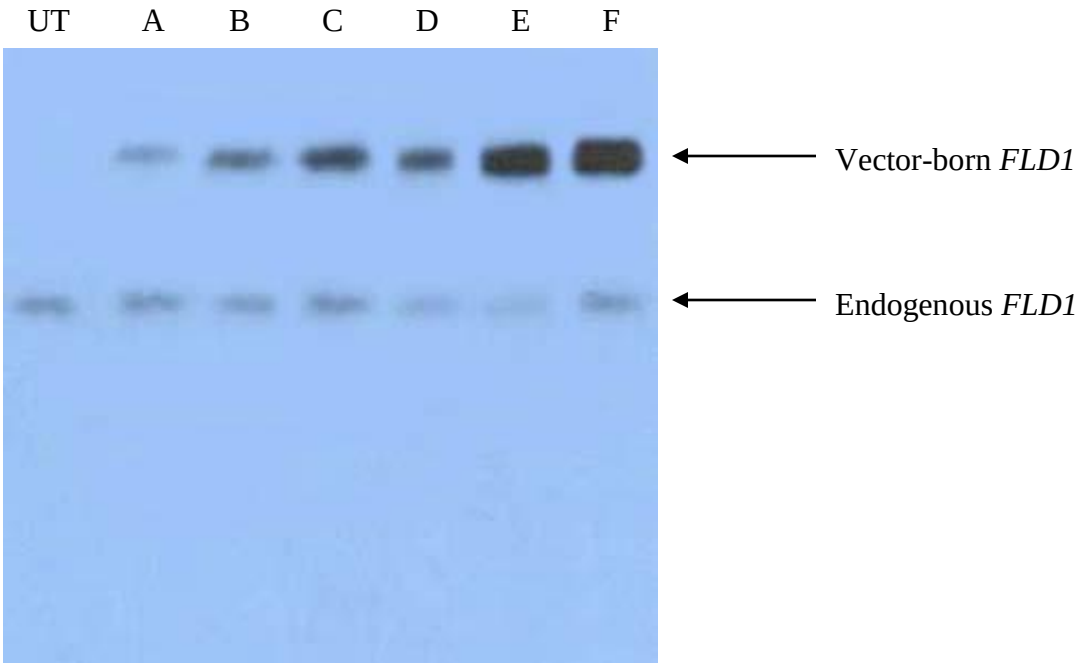


Figure 3.5 Southern blot showing FLD1 genes in selected *P. pastoris* strains transformed with vector pJL-lacZ. Probe used was a labeled DNA fragment containing sequences from *FLD1* gene. Lanes contain genomic DNA digested with EcoRV from each strain. Strains shown are: untransformed control, UT; a single copy vector control, A; selected multicopy strains, B-F.

strains, we observed the same band plus a second larger (3.5 kb**)** band that is the vector copy of the *FLD1* gene (Figure 3.5, lanes A-F). By comparing the relative intensities of these two bands, we obtained a reasonable estimate of the number of copies of the vector-born *FLD1* gene and therefore, the number of expression vector copies present in each transformed strain. Similar to strains transformed with Zeocin and G418 resistance gene vectors, most high-level formaldehyde resistant strains appeared to contain only one copy of the plasmid-born *FLD1* gene (e.g., Figure 3.5, lane A). Also similar to the antibiotic selection systems, approximately 1%-2% of Mta+ strains appeared to contain more than one copy of pJL-lacZ (Figure 3.5, lanes B-F). The copy numbers estimated by densitometry analysis for the strains shown in Figure 3.5 are displayed in Table 3.2. Out of approximately 500 Mta+ and formaldehyde resistant strains examined, the highest copy number observed was approximately 22 per haploid genome. These same formaldehyde resistance and copy number studies were repeated with the secretion vector pJL-SX digested with NotI. Results were similar to those shown above for the intracellular expression vector pJL-lacZ with regard to transformation efficiency, proportion of transformants that were resistant to elevated levels of formaldehyde and percentage of transformants harboring multiple copies of the vector (data not shown).

As a further confirmation that our pJL-lacZ strains actually contained multiple copies of the vector, we examined methanol-grown cultures of each strain for levels of formaldehyde dehydrogenase (Fld1p) and β-gal activities. If these strains actually contain multiple copies of this vector, they should also produce levels of Fld1p and β-gal that are approximately proportional to that copy number. To examine this, cultures of each strain were induced on methanol, prepared as cell-free extracts, and the specific activities of Fld1p and β-gal in each determined. As shown in Table 3.2, there was an obvious strong correlation between estimated vector-born *FLD1* copy number and levels of Fld1p and β-gal activity (relative to single-copy control strain A) in each strain.

Table 3.2 Comparison of vector copy number to β-galactosidase and formaldehyde dehydrogenase activity levels in selected *P. pastoris* strains containing multiple copies of pJL-lacZ.

Strain	Copy Number[a]	β-Gal Activity[b]	Relative Activity[c]	Fld Activity[d]	Relative Activity[e]
A	1	2×10^{-3}	1	1.2×10^{-5}	1
B	2.7	5×10^{-3}	3	4.2×10^{-5}	4
C	6.5	7×10^{-3}	4	7.6×10^{-5}	6
D	9.8	14×10^{-3}	7	1.0×10^{-4}	8
E	13	29×10^{-3}	15	1.6×10^{-4}	13
F	22	34×10^{-3}	17	3.0×10^{-4}	25

[a]Number of copies of vector pJL-lacZ present in each strain estimated by Southern blot comparison of signal intensity from the single-copy endogenous *FLD1* gene to the vector-born copies of the *FLD1* gene (see Figure 3.5).
[b]Specific activity of β-gal in Units/ng in methanol-grown cells of each strain.
[c]β-gal activity levels relative to levels in single-copy strain A.
[d]Specific activity of formaldehyde dehydrogenase (Fld) in Units/μg in methanol-grown cells of each strain.
[e]Fld activity levels relative to levels in single-copy strain A.

3.4 Discussion

P. pastoris has become an important host organism for the production of recombinant proteins for both commercial and academic organizations (Cereghino, and Cregg, 2000). Because transcription of the system is a limiting factor in the expression of most foreign genes, even when they are under the control of the highly efficient methanol-inducible *AOX1* promoter, methods to increase the number of copies of recombinant protein expression cassettes have been developed and proven effective (e.g., Clare, Rayment, et al., 1991; Romanos, Clare, et al., 1991; Zhu, Shi, et al., 1997).

In this report, we describe a new selectable marker and scheme for identifying such multicopy expression strains. The marker gene is the *P. pastoris* formaldehyde dehydrogenase gene (*FLD1*), whose product is required for growth of *P. pastoris* on either methanol as a sole carbon source or alkylated amines such as methylamine as a nitrogen source. We discovered that *FLD1* can be used as a selectable marker in DNA-mediated transformations of *P. pastoris fld1* mutant strains by selection for utilization of methylamine as nitrogen source (Mta^+ phenotype). Furthermore, we discovered that it is possible to enrich populations of Mta^+ transformants for ones that contain multiple integrated copies of an *FLD1*-containing vector by screening Mta^+ colonies for ones that are also resistant to high levels **(**>20 mM) of formaldehyde, a toxic compound that is the substrate for formaldehyde dehydrogenase (Shen, Sulter, et al., 1998). The *FLD1* marker system is comparable to the previously described *P. pastoris* G418- and Zeocin-resistance marker systems with regard to ease of use of the system, the frequency of multicopy strains recovered from transformant populations, and the range of copy numbers observed in multicopy strains after application of the enrichment method (Clare, Rayment, et al., 1991; Higgins, and Cregg, 1998). As with the antibiotic-resistance marker systems, the selectable compound, in this case formaldehyde, need only be added to transformation plates, and to a second set of plates used to single colony purify transformed strains (Higgins, Busser, et al., 1998). Once isolated, *P. pastoris* strains with integrated expression vectors need not be subjected to further formaldehyde selection and are stable with regard to the presence and copy number of the vector when maintained on a standard repressing growth medium such as YPD.

The *FLD1* marker system has advantages that are uniquely suited to certain circumstances. Since *FLD1* is a native *P. pastoris* gene, we were able to incorporate this gene into expression vectors composed almost entirely of *P. pastoris* DNA sequences (except for the foreign expressed gene) and devoid of functional bacterial sequences when transformed into *P. pastoris*. In particular, the resulting *P. pastoris* expression strains do not contain bacterial antibiotic resistance genes or replication origins that might be considered a biological hazard. Thus, the *FLD1* marker system should be particularly useful to commercial organizations that wish to produce a recombinant protein at a large industrial scale and are looking to minimize potential hazards associated with the strain and process. For academic labs, the *FLD1* marker system is significantly less expensive to use than the G418 and Zeocin marker systems, since formaldehyde, which is inexpensive, substitutes for these expensive antibiotics.

CHAPTER 4
CONCLUSION AND FUTURE DIRECTION

4.1 Summary of Research

This thesis describes a novel method for the enrichment of strains containing multiple copies of an expression vector for increased protein production. It also details the development of a new selectable marker for use in the transformation of the yeast *Pichia pastoris.*

Chapter 3 describes a novel method for the enrichment of *Pichia pastoris* yeast cells for multicopy integration called Post Transformational Vector Amplification (PTVA) and validates that the vector is amplified in a head-to-tail arrangement. We demonstrate this scheme using a Zeocin-based vector for the transformation of a *Pichia pastoris* wild type strain. Upon initial transformation, "single copy" or low copy strains are isolated to define the initial parent strain. These parent strains then undergo the PTVA method for enrichment of high copy clones using zeocin containing plates as our selection media. The PTVA process continues until high copy clones are isolated and tested for positive enrichment.

The data demonstrate that the enrichment process works more efficiently than the previous method screening for multiple copy strains. We show that 40% of our clones have three to five fold higher expression of our reporter protein after transformation, than the parent strain. Also among the group of PTVA clones, 5-6% contained greater than ten copies of our expression vector, while previous methods usually generated only about 1% from a typical screen. This new method has been shown be applicable to our new *FLD* marker system as well as the G418-based and Zeocin-based vector systems.

Chapter 3 describes the use of the *Pichia pastoris* formaldehyde dehydrogenase (*FLD1*) gene as a new selectable marker for the selection of multicopy expression strains in *Pichia pastoris*. This system is based on the observation that *FLD* is required to protect cells from the toxic levels of formaldehyde formed when methanol or methylamine is metabolized from the growth media.

The selection of transformants with the expression vector occurs in two steps. The first step is the transformation of a *P. pastoris* fld1 mutant strain with the expression vector containing the *FLD1* marker gene. Transformants are selected on methylamine containing plates. This first selection screens for positive clones that have integrated the expression vector, since the mutant strains require the *FLD* gene for growth on methylamine. Once the positive clones are chosen, the second step involves the selection of these strains on YPD plates containing formaldehyde. The higher concentration of formaldehyde would theoretically select for clones with a higher copy number of our expression/marker vector. This would generally translate to the higher expression of the "gene of interest".

The development of a new selectable marker that is endogenous to *Pichia pastoris* allows for the integration of an expression vector that is devoid of any bacterial sequences. This is advantageous in the production of therapeutic proteins that are highly regulated by the Federal Drug Administration. The absence of bacterial genes in the production strain avoids the concern of introducing resistant genes into the environment during production. Since the only exogenous gene integrated with the vector is "your favorite gene", the initial purification of the synthesized protein would be less complicated.

Previous screening methods for multiple copy selection was believe to occur only during transformation, thus the percentage of "high copy" clones were at a low frequency. We had addressed this problem by developing an optimized method and by showing that enrichment is possible after transformation. The PTVA method allows for the enrichment of "high copy" clones after transformation and thus makes the efficiency of generating these highly sought after clones more reasonable.

The development of more molecular tools to enhance the production of proteins using the *Pichia pastoris* expression system continues to be a growing field of research. The popularity of this system has fueled many researchers to come up with better techniques, more selectable markers, different promoters or secretion signals, optimized fermentation protocols and better strains that would handle the complex characteristics of the wide range of proteins being produced in *Pichia pastoris*.

4.2 Future Directions

Pichia pastoris is widely accepted as a host system for the production of recombinant proteins and as a valuable tool in cell biology research. In order to develop new strategies for both research fields, a comprehensive understanding of the organism would provide new insights and approaches that may help improve the yeast system. With the sequence of the *Pichia pastoris* genome completed, the use of bioinformatics would provide this avenue for researchers. Microarray technology is an expansion of the basic principle of Southern blotting with high throughput capability. DNA or RNA sequences are covalently attached to a substrate and then probed with known genes or fragments (Schena, Shalon, et al., 1995). The mRNA or gene expression profiling experiments can observe expression levels of thousands of genes simultaneously to study the effects of certain conditions on the genome (Lashkari, DeRisi, et al., 1997). This technology can identify genes whose expression levels have changed due to specific substrates added or removed to the growth medium. Since yeast arrays can contain thousands of probes, one experiment can compare gene profiles between cells under many different conditions and hasten studies that would otherwise take years. Figure 4.1 show the many changes in gene expression when comparing RNA from *Pichia pastoris* cells grown in glycerol or methanol for 12h hours.

The use of the available *Pichia pastoris* genome and the microarray expression profile could identify significant genes and their promoters that may help scientist develop new selectable markers or promoters used for expression. For cell biologist, the molecular players involved in specific processes, such as peroxisome biogenesis, may make elucidation more efficient and thorough. This exciting technology is promising, but the complexity of gene expression makes the experimental design and thus analysis (data standardization and normalization) critical for any conclusion to be statistically or biologically valid. Figure 4.2 show the excellent reproducibility of the *Pichia* Affymetrix Gene Chip when comparing the hybridization of biological replicates. This technology provides the opportunity for faster characterization and discovery of new promoters and genes in the *Pichia pastoris* genome. This will provide researchers the platform to develop new tools for the yeast expression system as well as help in the study of genes involved in important pathways that are relevant in *Pichia pastoris*.

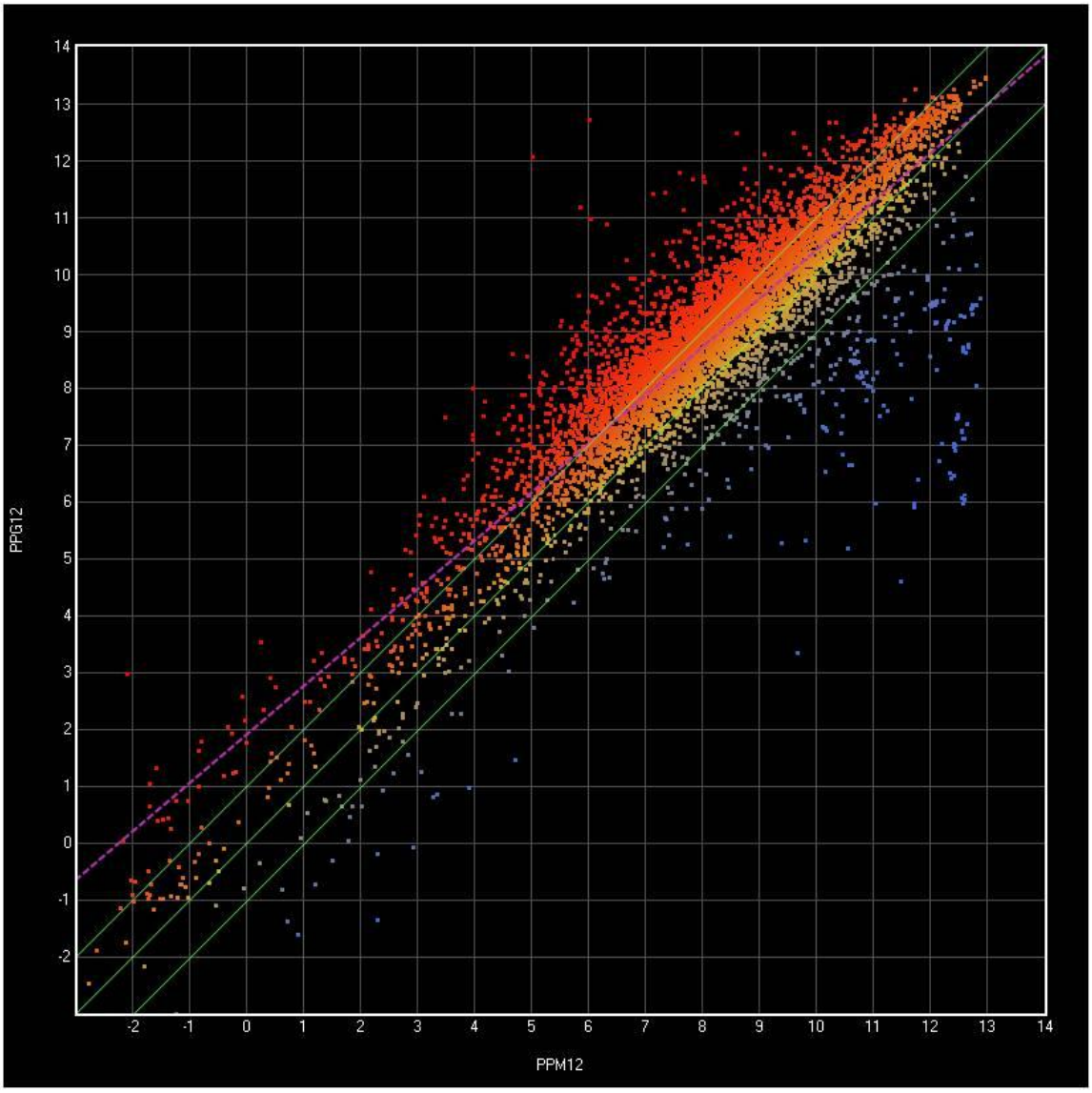


Figure 4.1 Comparison of RNA from *Pichia pastoris* cells grown in glycerol or methanol for 12 hours show many changes in gene expression. Graphs provided my Joel Kreps of Verenium Corporation (San Diego, CA).

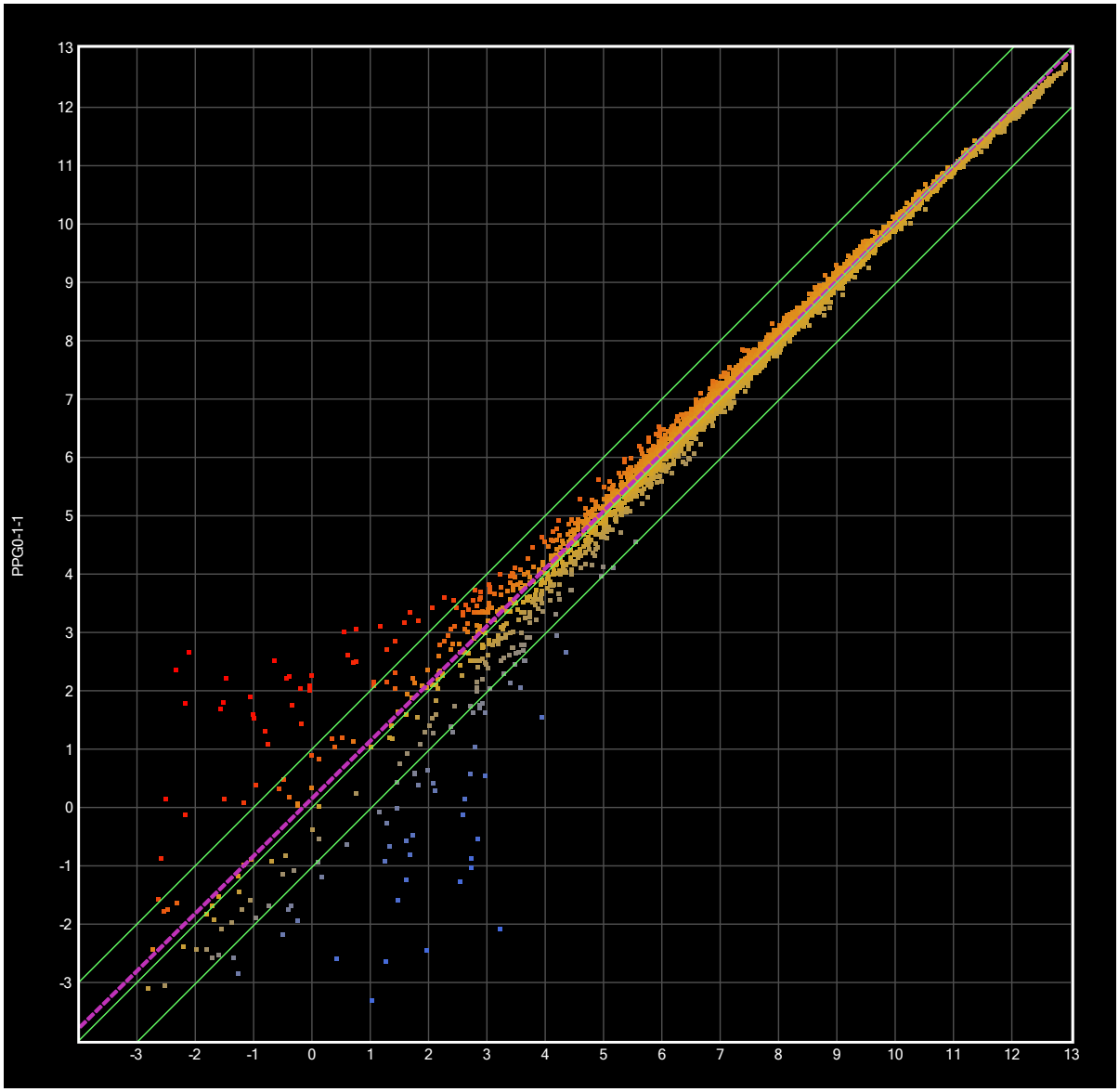


Figure 4.2 Comparison of RNA expression levels from biological replicates show an excellent sample to sample reproducibility.

LITERATURE CITED

Agarwal, K.L., Buchi, H., Caruthers, M.H., Gupta, N., Khorana, H.G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U.L., Van de Sande, J.H., Sgaramella, V., Weber, H., and Yamada, T. (1970). Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. Nature *227,* 27-34.

Aguilera, A., Chavez, S., and Malagon, F. (2000). Mitotic recombination in yeast: elements controlling its incidence. Yeast *16,* 731-754.

Ahn, J., Hong, J., Lee, H., Park, M., Lee, E., Kim, C., Choi, E., Jung, J., and 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.

Ailor, E., and Betenbaugh, M.J. (1999). Modifying secretion and post-translational processing in insect cells. Curr. Opin. Biotechnol. *10,* 142-145.

Alt, F.W., Kellems, R.E., Bertino, J.R., and Schimke, R.T. (1978). Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J. Biol. Chem. *253,* 1357-1370.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G., and Struhl, K. (2001a). Current Protocols in Molecular Biology. (New York: John Wiley and Sons, Inc.) pp. 1.3.1-3.19.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G., and Struhl, K. (2001b). Current Protocols in Molecular Biology. (New York: John Wiley and Sons, Inc.) pp. 1.2.1-2.9.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1997). Current Protocols in Molecular Biology.16.0.1-16.4.3-16.0.3.

Aylon, Y., and Kupiec, M. (2004a). DSB repair: the yeast paradigm. DNA Repair (Amst) *3,* 797-815.

Aylon, Y., and Kupiec, M. (2004b). New insights into the mechanism of homologous recombination in yeast. Mutat. Res. *566,* 231-248.

Berg, P., Baltimore, D., Boyer, H.W., Cohen, S.N., Davis, R.W., Hogness, D.S., Nathans, D., Roblin, R., Watson, J.D., Weissman, S., and Zinder, N.D. (1974). Letter: Potential biohazards of recombinant DNA molecules. Science *185,* 303.

Biotechmedia. The History of Biotechnology.*2006,* 1.

Boeke, J.D., Xu, H., and Fink, G.R. (1988). A general method for the chromosomal amplification of genes in yeast. Science *239,* 280-282.

Brierley, R.A. (1998). Secretion of recombinant human insulin-like growth factor I (IGF-I). Methods Mol. Biol. *103,* 149-177.

Buckholz, R.G., and Gleeson, M.A. (1991). Yeast systems for the commercial production of heterologous proteins. Biotechnology (N. Y) *9,* 1067-1072.

Cereghino, G.P., Cereghino, J.L., Ilgen, C., and Cregg, J.M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr. Opin. Biotechnol. *13,* 329-332.

Cereghino, G.P., and Cregg, J.M. (1999). Applications of yeast in biotechnology: protein production and genetic analysis. Curr. Opin. Biotechnol. *10,* 422-427.

Cereghino, J.L., and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. *24,* 45-66.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J. *4,* 3273-3280.

Clare, J.J., Rayment, F.B., Ballantine, S.P., Sreekrishna, K., and Romanos, M.A. (1991). High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Biotechnology (N. Y) *9,* 455-460.

Claycomb, J.M., and Orr-Weaver, T.L. (2005). Developmental gene amplification: insights into DNA replication and gene expression. Trends Genet. *21,* 149-162.

Cohen, S.N. (1977). Recombinant DNA: fact and fiction. Science *195,* 654-657.

Cohen, S.N., Chang, A.C., Boyer, H.W., and Helling, R.B. (1973). Construction of biologically functional bacterial plasmids in vitro. Proc. Natl. Acad. Sci. U. S. A. *70,* 3240-3244.

Cregg, J.M. (2007). DNA-Mediated Transformation. In Pichia Protocols: Methods in Molecular Biology, pp. 27-42.

Cregg, J.M., Tschopp, J.F., Stillman, C., Siegel, R., Akong, M., Craig, W.S., Buckholz, R.G., Madden, K.R., Kellaris, P.A., Davis, G.R. *et al*. (1987). High-level Expression and Efficient Assemble of Hepatitis B Surface Antigen in the Methylotrophic Yeast, *Pichia pastoris*. BIO/TECHNOLOGY *5,* 479-485.

Cregg, J.M. (1987). Genetics of Methylotrophic Yeasts. In Proceedings of the Fifth Internatinal Symposium on Microbial Growth on C_1 Compounds, Duine, J. A., and van Verseveld, H. W. eds., (Dordrecht: Martinus Nijhoff Publishers) pp. 158-167.

Cregg, J.M., Digan, M.E., Tschopp, J.A., Brierley, R.A., Craig, W.S., Velicelebi, G., Siegel, R.S., and Thill, G.P. (1989). Expression of Foreign Genes in *Pichia pastoris*. In Genetics and Molecular Biology of Industrial Microorganisms, Hershberger, C. L., Queener, S. W. and Hegeman, G. eds., (Washington, DC: American Society for Microbiology) pp. 343-352.

Cregg, J.M., and Higgins, D.R. (1995). Production of foreign proteins in the yeast *Pichia pastoris*. Can. J. Bot *73,* S891-S897.

Cregg, J.M., Barringer, K.J., Hessler, A.Y., and Madden, K.R. (1985). *Pichia pastoris* as a host system for transformations. Mol. Cell. Biol. *5,* 3376-3385.

Cregg, J.M., Cereghino, J.L., Shi, J., and Higgins, D.R. (2000). Recombinant protein expression in *Pichia pastoris*. Mol. Biotechnol. *16,* 23-52.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P., and Stillman, C.A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol. Cell. Biol. *9,* 1316-1323.

Cregg, J.M., and Russell, K.A. (1998). Transformation. Methods Mol. Biol. *103,* 27-39.

Cregg, J.M., Shen, S., Johnson, M., and Waterham, H.R. (1998). Classical genetic manipulation. Methods Mol. Biol. *103,* 17-26.

Cregg, J.M., Vedvick, T.S., and Raschke, W.C. (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. Biotechnology (N. Y) *11,* 905-910.

Cromie, G.A., Connelly, J.C., and Leach, D.R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol. Cell *8,* 1163-1174.

Dominguez, A., Ferminan, E., Sanchez, M., Gonzalez, F.J., Perez-Campo, F.M., Garcia, S., Herrero, A.B., San Vicente, A., Cabello, J., Prado, M., Iglesias, F.J., Choupina, A., Burguillo, F.J., Fernandez-Lago, L., and Lopez, M.C. (1998). Non-conventional yeasts as hosts for heterologous protein production. Int. Microbiol. *1,* 131-142.

Dorsey, M., Peterson, C., Bray, K., and Paquin, C.E. (1992). Spontaneous amplification of the ADH4 gene in *Saccharomyces cerevisiae*. Genetics *132,* 943-950.

Eckart, M.R., and Bussineau, C.M. (1996). Quality and authenticity of heterologous proteins synthesized in yeast. Curr. Opin. Biotechnol. *7,* 525-530.

Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M., and Gingeras, T.R. (1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. Mol. Cell. Biol. *5,* 1111-1121.

Faber, K.N., Harder, W., Ab, G., and Veenhuis, M. (1995). Review: methylotrophic yeasts as factories for the production of foreign proteins. Yeast *11,* 1331-1344.

Fasullo, M., Giallanza, P., Dong, Z., Cera, C., and Bennett, T. (2001). *Saccharomyces cerevisiae* rad51 mutants are defective in DNA damage-associated sister chromatid exchanges but exhibit increased rates of homology-directed translocations. Genetics *158,* 959-972.

Fogel, S., and Welch, J.W. (1982). Tandem gene amplification mediates copper resistance in yeast. Proc. Natl. Acad. Sci. U. S. A. *79,* 5342-5346.

Gellissen, G. (2005). Yeast Systems. In Production of Recombinant Proteins: Microbial and Eukaryotic Expression Systems, (Weinheim, Germany: Wiley-VCH Verlag) pp. 111-185.

Gellissen, G., Melber, K., Janowicz, Z.A., Dahlems, U.M., Weydemann, U., Piontek, M., Strasser, A.W.M., and Hollenberg, C.P. (1992). Heterologous protein production in yeast. Antonie Van Leeuwenhoek *62,* 79-93.

Gellissen, G., Weydemann, U., Strasser, A.W., Piontek, M., Janowicz, Z.A., and Hollenberg, C.P. (1992). Progress in developing methylotrophic yeasts as expression systems. Trends Biotechnol. *10,* 413-417.

Goeddel, D.V. (1990). Systems for heterologous gene expression. Methods Enzymol. *185,* 3-7.

Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R., and Seeburg, P.H. (1979). Direct expression in Escherichia coli of a DNA sequence coding for human growth hormone. Nature *281,* 544-548.

Gonzalez-Barrera, S., Cortes-Ledesma, F., Wellinger, R.E., and Aguilera, A. (2003). Equal sister chromatid exchange is a major mechanism of double-strand break repair in yeast. Mol. Cell *11,* 1661-1671.

Guarente, L. (1984). Yeast promoters: positive and negative elements. Cell *36,* 799-800.

Gwynne, P., and Page, G. (2006). Biotechnology: Personal Portraits of an Evolving Industry. Science (Supplemental) *311,* supplement.

Haber, J.E., and Debatisse, M. (2006). Gene amplification: yeast takes a turn. Cell *125,* 1237-1240.

Haber, J.E., Ira, G., Malkova, A., and Sugawara, N. (2004). Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *359,* 79-86.

Hastings, P.J. (1988). Recombination in the eukaryotic nucleus. Bioessays *9,* 61-64.

Higgins, D.R., Busser, K., Comiskey, J., Whittier, P.S., Purcell, T.J., and Hoeffler, J.P. (1998). Small vectors for expression based on dominant drug resistance with direct multicopy selection. Methods Mol. Biol. *103,* 41-53.

Higgins, D.R., and Cregg, J.M. (1998). Introduction to *Pichia pastoris*. Methods Mol. Biol. *103,* 1-15.

Hinnen, A., Hicks, J.B., and Fink, G.R. (1978). Transformation of yeast. Proc. Natl. Acad. Sci. U. S. A. *75,* 1929-1933.

Hitzeman, R.A., Leung, D.W., Perry, L.P., Kohr, W.J., Levine, H.L., and Goeddel, D.V. (1983). Secretion of Human Interferons by Yeast. Science *219,* 620-625.

Hoy, C.A., Rice, G.C., Kovacs, M., and Schimke, R.T. (1987). Over-replication of DNA in S phase Chinese hamster ovary cells after DNA synthesis inhibition. J. Biol. Chem. *262,* 11927-11934.

Ilgen, C., Cereghino, J.L., and Cregg, J.M. (2004). *Pichia pastoris* . In Production of recombinant proteins: microbial and eukaryotic expression systems, Gellissen, G. ed., (Weinheim, Germany: Wiley-VCH Verlag) pp. 143-162.

Inbar, O., and Kupiec, M. (1999). Homology search and choice of homologous partner during mitotic recombination. Mol. Cell. Biol. *19,* 4134-4142.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., and Boyer, H.W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science *198,* 1056-1063.

Jackson, D.A., Symons, R.H., and Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *69,* 2904-2909.

Kadyk, L.C., and Hartwell, L.H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. Genetics *132,* 387-402.

Kaye, J.A., Melo, J.A., Cheung, S.K., Vaze, M.B., Haber, J.E., and Toczyski, D.P. (2004). DNA breaks promote genomic instability by impeding proper chromosome segregation. Curr. Biol. *14,* 2096-2106.

Khan, S.A. (1997). Rolling-circle replication of bacterial plasmids. Microbiol. Mol. Biol. Rev. *61,* 442-455.

Kingsman, S.M., and Kingsman, A.J. (1989). Host-vector systems. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *324,* 477-485.

Kost, T.A., and Condreay, J.P. (1999). Recombinant baculoviruses as expression vectors for insect and mammalian cells. Curr. Opin. Biotechnol. *10,* 428-433.

Kramer, M.F., and Coen, D.M. (2000). The Polymerase Chain Reaction. In Current Protocols in Molecular Biolog, Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. eds., (New York: John Wiley and Sons, Inc.) pp. 15.0.3-15.1.14.

Kuntz, M. Milestones is DNA History. *2006,* 1.

Kupiec, M., and Petes, T.D. (1988). Allelic and ectopic recombination between Ty elements in yeast. Genetics *119,* 549-559.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., and Davis, R.W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U. S. A. *94,* 13057-13062.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396,* 643-649.

Lin Cereghino, G.P., Lin Cereghino, J., Sunga, A.J., Johnson, M.A., Lim, M., Gleeson, M.A., and Cregg, J.M. (2001). New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. Gene *263,* 159-169.

Lin Cereghino, G.P., Sunga, A.J., Lin Cereghino, J., and Cregg, J.M. (2001). Expression of foreign genes in the yeast *Pichia pastoris*. Genet. Eng. (N. Y) *23,* 157-169.

Lin, Y., Lukacsovich, T., and Waldman, A.S. (1999). Multiple pathways for repair of DNA double-strand breaks in mammalian chromosomes. Mol. Cell. Biol. *19,* 8353-8360.

Liu, H., Tan, X., Russell, K.A., Veenhuis, M., and Cregg, J.M. (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.

Liu, X., and Chang, K.P. (1994). Identification by extrachromosomal amplification and overexpression of a zeta-crystallin/NADPH-oxidoreductase homologue constitutively expressed in *Leishmania* spp. Mol. Biochem. Parasitol. *66,* 201-210.

Lopes, T.S., Hakkaart, G.J., Koerts, B.L., Raue, H.A., and Planta, R.J. (1991). Mechanism of high-copy-number integration of pMIRY-type vectors into the ribosomal DNA of *Saccharomyces cerevisiae*. Gene *105,* 83-90.

Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol. Rev. *60,* 512-538.

Mansur, M., Cabello, C., Hernandez, L., Pais, J., Varas, L., Valdes, J., Terrero, Y., Hidalgo, A., Plana, L., Besada, V. *et al*. (2005). Multiple gene copy number enhances insulin precursor secretion in the yeast *Pichia pastoris*. Biotechnol. Lett. *27,* 339-345.

Mellor, J., Dobson, M.J., Roberts, N.A., Tuite, M.F., Emtage, J.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J., and Kingsman, S.M. (1983). Efficient synthesis of enzymatically active calf chymosin in *Saccharomyces cerevisiae*. Gene *24,* 1-14.

Meselson, M.S., and Radding, C.M. (1975). A general model for genetic recombination. Proc. Natl. Acad. Sci. U. S. A. *72,* 358-361.

Miles, D.J., Busser, K., Stadler, C., and Higgins, D.R. (1998). Isolation of nucleic acids. In: Higgings, D.R., J.M. (Eds.), Methods in Molecular Biology: *Pichia* protocols. (Human Press, Totowa, NJ), pp. 73-80.

Moore, I.K., Martin, M.P., Dorsey, M.J., and Paquin, C.E. (2000). Formation of circular amplifications in *Saccharomyces cerevisiae* by a breakage-fusion-bridge mechanism. Environ. Mol. Mutagen. *36,* 113-120.

Morlino, G.B., Tizzani, L., Fleer, R., Frontali, L., and Bianchi, M.M. (1999). Inducible amplification of gene copy number and heterologous protein production in the yeast *Kluyveromyces lactis*. Appl. Environ. Microbiol. *65,* 4808-4813.

Morrow, J.F., Cohen, S.N., Chang, A.C., Boyer, H.W., Goodman, H.M., and Helling, R.B. (1974). Replication and transcription of eukaryotic DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *71,* 1743-1747.

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. *14,* 1613-1625.

Ogata, K., Nishikawa, H., and Ohsugi, M. (1969). A yeast capable of utilizing methanol. Agric. Biol. Chem. *33,* 1519-1520.

Orr-Weaver, T.L., and Szostak, J.W. (1985). Fungal recombination. Microbiol. Rev. *49,* 33-58.

Orr-Weaver, T.L., and Szostak, J.W. (1983a). Multiple, tandem plasmid integration in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *3,* 747-749.

Orr-Weaver, T.L., and Szostak, J.W. (1983b). Yeast recombination: the association between double-strand gap repair and crossing-over. Proc. Natl. Acad. Sci. U. S. A. *80,* 4417-4421.

Orr-Weaver, T.L., Szostak, J.W., and Rothstein, R.J. (1981). Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. U. S. A. *78,* 6354-6358.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. *63,* 349-404.

Paques, F., Leung, W.Y., and Haber, J.E. (1998). Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol. Cell. Biol. *18,* 2045-2054.

Paquin, C.E., Dorsey, M., Crable, S., Sprinkel, K., Sondej, M., and Williamson, V.M. (1992). A spontaneous chromosomal amplification of the ADH2 gene in *Saccharomyces cerevisiae*. Genetics *130,* 263-271.

Paulson, T.G., Almasan, A., Brody, L.L., and Wahl, G.M. (1998). Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate DNA breakage. Mol. Cell. Biol. *18,* 3089-3100.

Pipiras, E., Coquelle, A., Bieth, A., and Debatisse, M. (1998). Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. EMBO J. *17,* 325-333.

Plessis, A., and Dujon, B. (1993). Multiple tandem integrations of transforming DNA sequences in yeast chromosomes suggest a mechanism for integrative transformation by homologous recombination. Gene *134,* 41-50.

Prado, F., and Aguilera, A. (1995). Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. Genetics *139,* 109-123.

Prado, F., Cortes-Ledesma, F., Huertas, P., and Aguilera, A. (2003). Mitotic recombination in *Saccharomyces cerevisiae*. Curr. Genet. *42,* 185-198.

Rai, M., and Padh, H. (2001). Expression systems for production of heterologous proteins. Current Science *80,* 1121-1128.

Rattray, A.J., Shafer, B.K., Neelam, B., and Strathern, J.N. (2005). A mechanism of palindromic gene amplification in *Saccharomyces cerevisiae*. Genes Dev. *19,* 1390-1399.

Resnick, M.A. (1976). The repair of double-strand breaks in DNA; a model involving recombination. J. Theor. Biol. *59,* 97-106.

Romanos, M. (1995). Advances in the use of *Pichia pastoris* for high-level gene expression. Current Opinion in Biotechnology *6,* 527-533.

Romanos, M., Scorer, C., Sreekrishna, K., and Clare, J. (1998). The generation of multicopy recombinant strains. Methods Mol. Biol. *103,* 55-72.

Romanos, M.A., Clare, J.J., Beesley, K.M., Rayment, F.B., Ballantine, S.P., Makoff, A.J., Dougan, G., Fairweather, N.F., and Charles, I.G. (1991). Recombinant Bordetella pertussis pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties. Vaccine *9,* 901-906.

Saffran, W.A., Smith, E.D., and Chan, S.K. (1991). Induction of multiple plasmid recombination in *Saccharomyces cerevisiae* by psoralen reaction and double strand breaks. Nucleic Acids Res. *19,* 5681-5687.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Assay for B-Galactosidase in Extracts of Mammalian Cells. In Molecular Cloning. A Laboratory Manual. Nolan, Chris ed., (Cold Springs Harbor, NY: Cold Springs Harbor Laboratory Press) pp. 16.66-16,67.

Schacherer, J., de Montigny, J., Welcker, A., Souciet, J.L., and Potier, S. (2005). Duplication processes in *Saccharomyces cerevisiae* haploid strains. Nucleic Acids Res. *33,* 6319-6326.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270,* 467-470.

Schimke, R.T. (1986). Methotrexate resistance and gene amplification. Mechanisms and implications. Cancer *57,* 1912-1917.

Schmidt, F.R. (2004). Recombinant expression systems in the pharmaceutical industry. Appl. Microbiol. Biotechnol. *65,* 363-372.

Schwacha, A., and Kleckner, N. (1995). Identification of double Holliday junctions as intermediates in meiotic recombination. Cell *83,* 783-791.

Scorer, C.A., Clare, J.J., McCombie, W.R., Romanos, M.A., and Sreekrishna, K. (1994). Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. Biotechnology (N. Y) *12,* 181-184.

Sears, I.B., O'Connor, J., Rossanese, O.W., and Glick, B.S. (1998). A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast *14,* 783-790.

Shatzman, A.R., and Rosenberg, M. (1987). Expression, identification, and characterization of recombinant gene products in *Escherichia coli*. Methods Enzymol. *152,* 661-673.

Shen, S., Sulter, G., Jeffries, T.W., and Cregg, J.M. (1998). A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. Gene *216,* 93-102.

Sibirny, A.A., Ubiyvovk, V.M., Gonchar, M.V., Titorenko, V.I., Voronovsky, A.Y., Kapultsevich, Y.G., and Bliznik, K.M. (1990). Reactions of direct formaldehyde oxidation to CO_2 are non-essential for energy supply of yeast methylotrophic growth. Arch. Microbiol *154,* 566-575.

Silberman, R., and Kupiec, M. (1994). Plasmid-mediated induction of recombination in yeast. Genetics *137,* 41-48.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu. Rev. Biochem. *72*, 449-479.

Stark, G.R., and Wahl, G.M. (1984). Gene amplification. Annu. Rev. Biochem. *53*, 447-491.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc. Natl. Acad. Sci. U. S. A. *82*, 8419-8423.

Sugawara, N., and Haber, J.E. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol. Cell. Biol. *12*, 563-575.

Sunga, A.J., and Cregg, J.M. (2004). The *Pichia pastoris* formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of *P. pastoris*. Gene *330*, 39-47.

Sunga, A.J., Tolstorukov, I., and Cregg, J.M. (2008). Posttransformational vector amplification in the yeast *Pichia pastoris*. FEMS Yeast Res.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25-35.

Tabor, J.M. (1988). Genetic Engineering Technology in Industrial Pharmacy: Principles and Applications (New York: Marcel Dekker).

Talmadge, K., Kaufman, J., and Gilbert, W. (1980). Bacteria mature preproinsulin to proinsulin. Proc. Natl. Acad. Sci. U. S. A. *77*, 3988-3992.

Thill, G.P., Davis, G.R., Stillman, C., Holtz, G., Brierly, R., Engel, M., Buckholtz, R., Kenney, J., Provow, S., Vedvick, T., and Seigel, R.S. (1990). Positive and negative effects of multicopy integrated expression vectors on protein expression in *Pichia pastoris.* . In Proceedings of the Sixth International Symposium on the Genetics of Microorganisms, Heslot, H., Davies, J., Florent, J., Bobichon, L., Durand, G. and Penasse, L. eds., (Paris: Societe Francaise de Microbiologie) pp. 477-490.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., and Gingeras, T.R. (1987). Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Res. *15*, 3859-3876.

Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G., and Hall, B.D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature *298*, 347-350.

Vassileva, A., Chugh, D.A., Swaminathan, S., and Khanna, N. (2001). Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. Protein Expr. Purif. *21*, 71-80.

Veenhuis, M., Van Dijken, J.P., and Harder, W. (1983). The significance of peroxisomes in the metabolism of one-carbon compounds in yeasts. Adv. Microb. Physiol. *24*, 1-82.

Voelkel-Meiman, K., and Roeder, G.S. (1990). Gene conversion tracts stimulated by HOT1-promoted transcription are long and continuous. Genetics *126*, 851-867.

Wang, X., Ira, G., Tercero, J.A., Holmes, A.M., Diffley, J.F., and Haber, J.E. (2004). Role of DNA replication proteins in double-strand break-induced recombination in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *24*, 6891-6899.

Watanabe, T., and Horiuchi, T. (2005). A novel gene amplification system in yeast based on double rolling-circle replication. EMBO J. *24*, 190-198.

Waterham, H.R., Digan, M.E., Koutz, P.J., Lair, S.V., and Cregg, J.M. (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene *186*, 37-44.

Waterham, H.R., Keizer-Gunnink, I., Goodman, J.M., Harder, W., and Veenhuis, M. (1992). Development of multipurpose peroxisomes in *Candida boidinii* grown in oleic acid-methanol limited continuous cultures. J. Bacteriol. *174*, 4057-4063.

Watson, J.D., Gilman, M., Witkowski, J., and Zoller, M. (1992). Recombinant DNA.626.

Williamson, D.H. (1985). The yeast ARS element, six years on: a progress report. Yeast *1*, 1-14.

Windle, B.E., and Wahl, G.M. (1992). Molecular dissection of mammalian gene amplification: new mechanistic insights revealed by analyses of very early events. Mutat. Res. *276*, 199-224.

Wright, S. (2001). Legitimating Genetic Engineering. Perspectives in Biology and Medicine *44*, 235-247.

Wright, S. (1986). Recombinant DNA Technology and Its Social Transformation, 1972-1982. OSIRIS *2*, 303-360.

Yandeau-Nelson, M.D., Xia, Y., Li, J., Neuffer, M.G., and Schnable, P.S. (2006). Unequal sister chromatid and homolog recombination at a tandem duplication of the A1 locus in maize. Genetics *173*, 2211-2226.

Yoshimori, R., Roulland-Dussoix, D., and Boyer, H.W. (1972). R factor-controlled restriction and modification of deoxyribonucleic acid: restriction mutants. J. Bacteriol. *112*, 1275-1279.

Zgaga, Z., Chanet, R., Radman, M., and Fabre, F. (1991). Mismatch-stimulated plasmid integration in yeast. Curr. Genet. *19*, 329-332.

Zhu, H., Shi, J., Cregg, J.M., and Woldegiorgis, G. (1997). Reconstitution of highly expressed human heart muscle carnitine palmitoyltransferase I. Biochem. Biophys. Res. Commun. *239,* 498-502.

Zwart, K., Veenhuis, M., van Dijken, J.P., and Harder, W. (1980). Development of amine oxidase-containing peroxisomes in yeasts during growth on glucose in the presence of methylamine as the sole source of nitrogen. Arch. Microbiol. *126,* 117-126.