PEROXISOME ASSEMBLY IN *YARROWIA LIPOLYTICA*

PEROXISOME ASSEMBLY IN *YARROWIA LIPOLYTICA*

by

ANTHONY MATTHEW BRADE

A thesis submitted to the School of Graduate Studies

in partial fulfilment of the requirements

for the degree

Master of Science

McMaster University

MASTER OF SCIENCE (1992) McMaster University
(Biochemistry) Hamilton, Ontario

TITLE: Peroxisom Assembly in *Yarrowia lipolytica*

AUTHOR: Anthony Matthew Brade, B. Sc. (McMaster University)

SUPERVISOR: Professor R.A. Rachubinski

NUMBER OF PAGES: viii, 80.

Acknowledgments

First and foremost, I would like to extend my sincere thanks to Rick for his enthusiasm and his support, especially during the tough times. I came into the lab and I learned, I don't think I could have asked for anything more. Thanks to everybody in the lab, past and present, John Aitchison, Andrea Bodnar, Gary Eitzen, John Glover, Sandra Marcus, Wayne Murray, Bill Nuttley, Gill Regoeczi, Jim Sloots, Rachel Szilard, Baowei Zhang (and Rick) for their help, their humour, their thoughts and ideas. We had a lot of good times and a lot of laughs (a few come to mind, the OFO award, the perils of Johnny G's sea turtle, and don't ask me about the cure for hiccups, I think we learn that in third year), I'll remember it all fondly.

Thanks also to everybody else who helped/commiserated/supported *etc.* over the last two years, the Ananth, Andrews, Smiley, Harley, Gerber, Freeman, and Capone labs, Ramsi, Chris, Craig and the rest of the enCounter therapy group.

My thanks also to my family and my in-laws for all of their love and support throughout my project, I could not have done it without them. Special thanks must go to my wife Tiziana for her love, strength and encouragement during my project and the write up. I am especially indebted to her for the sacrifices she has made and for everything she has brought into my life.

Technical Acknowledgments

I am very grateful to many members of the lab for their physical and intellectual help and guidance with my project.

Bill Nuttley isolated and characterized the *pay4* strain, upon which essentially the entire project is based. He also performed all of the fractionations and enzyme analyses referred to in this thesis. I would also like to thank Dr. Rick Rachubinski for coordinating the sequencing effort and handling the lion's share of the sequencing. In addition, I would like to acknowledge the contributions of John Aitchison, Gary Eitzen, Bill Nuttley and Rachel Szilard, who all helped out with the sequencing of *PAY4*. I must also thank Gill Regoeczi for her absolute consistency and reliability in the making of dozens of solutions and literally thousands of plates for this project.

I especially want to thank Gary for his assistance, when the crunch was on, in assembling the *PAY4* 'knock-out' construct, and for analyzing potential knock-out strains. John Glover and Rick have also helped out tremendously in wrapping up and extending the project with their work on Pay4p expression.

Clearly, this lab works as a team and I would like to thank Rick for making this possible.

Table of Contents

List of Figures

ABSTRACT

The primary goal of the research presented in this thesis has been to establish the yeast *Yarrowia lipolytica* as a model genetic system for the study of peroxisome biogenesis. To facilitate this, three steps were necessary. First, mutant strains of *Y. lipolytica* that manifested peroxisomal defects were generated; second, a genomic DNA library was created to rescue the mutants and, hence, clone the gene(s) involved; and third, the parameters governing high efficiency DNA transformation of *Y. lipolytica* were defined to make complementation of mutants with the genomic library feasible. This work culminated in the cloning, sequencing and characterization of a gene, dubbed *PAY 4*, that is required for peroxisomal assembly in *Y. lipolytica* .

Two of the mutant strains of *Y. lipolytica* isolated, *ole 2* and *ole 4*, were identified as peroxisome assembly mutants as they; (i) grew in acetate medium but failed to grow on oleic acid medium, and thus were peroxisomal as opposed to mitochondrial mutants; (ii) lacked recognizable peroxisomal structures when observed by immunofluorescence using a polyclonal serum that recognized a number of peroxisomal proteins and; (iii) had several peroxisomal enzyme activities localized to the cytosol, i.e. they manifested mistargeting of otherwise lumenal peroxisomal proteins.

These were complemented with a genomic library, which represented at least 19 000 independent recombinants (insert size ~5-7 kilobases) cloned into a vector (pINA445) capable of autonomous replication in *Y. lipolytica*. Estimations of the size of the *Y. lipolytica* genome indicate that probability of this library containing a given portion of the genome is .999. To utilize the library effectively, an electro-poration protocol was developed that could transform *Y. lipolytica* at high efficiency ($>5 \cdot 10^4$ transformants$\cdot \mu g^{-1}$ of DNA).

Using these resources, the gene complementing the *ole 4* mutation was cloned and sequenced. It encodes Pay4p, a protein of ~112 kDa, containing two putative ATPase modules. Preliminary studies indicate that expression of Pay4p is induced slightly by growth on oleic acid.

1. INTRODUCTION

1.0 What is a Peroxisome?

Peroxisomes are part of a family of membrane bound subcellular organelles, known as microbodies, and are found in nearly all eukaryotes, They were first observed by Rhodin in 1954, and recorded in 1956 as small (~0.5 μm) electron dense vesicles in rodent kidney and liver, each bounded by a single unit membrane (Rouiller and Bernhard, 1956). A series of seminal cell fractionation experiments by de Duve and his colleagues, in the late fifties and early sixties, demonstrated the co-localization of urate oxidase and D-amino acid oxidase, two hydrogen peroxide producing enzymes, and catalase, a peroxide degrading enzyme, with the hepatic microbody (reviewed by de Duve and Baudhuin, 1966).

In 1966, de Duve and Baudhuin defined the peroxisome based on these results, hypothesizing that they represented important sites of hydrogen peroxide metabolism, although at that time the specific function of the peroxisome remained unclear (de Duve and Baudhuin, 1966). Nevertheless, this general definition of the peroxisome; a subcellular organelle bounded by a single lipid bilayer and containing both catalase and a H_2O_2 producing oxidase; still applies to almost all types of microbodies known today.

1.0.1 Functions of Microbodies

Three types of organelles, the peroxisome, the plant glyoxysome and the trypanosomal glycosome, make up the microbody family. The enzymatic complement, and hence the function, of the microbody varies substantially between species and within the organism depending on the tissue it is found in and/or the metabolic environment

(reviewed in van den Bosch *et al.*, 1992). As alluded to above, however, microbodies share some common functional features.

All microbodies, with the exception of the trypanosomal glycosome, contain at least one H_2O_2 producing oxidase and a catalase (de Duve, 1983; Borst, 1989; Opperdoes, 1987). Some trypanosomes have glycosomes that contain catalase, but no H_2O_2 producing oxidases have been localized to this organelle in any of these species.

All microbodies studied to date have a ß-oxidation system of some kind to metabolize fatty acids (Lazarow and Fujiki, 1985; Veenhuis and Harder, 1987; van den Bosch *et al.*, 1992). In mammals, peroxisomal ß-oxidation plays an important role in the breakdown of long-chain (C_{10}-C_{30}) fatty acids, with mitochondria carrying out ß-oxidation on fatty acids of shorter chain lengths ($<C_{18}$) (Lazarow, 1978). The importance of this role is emphasized by the dramatic induction of rat liver peroxisomes in response to treatment of the mammals with hypolipidemic agents (Thorp and Waring, 1962; Azarnoff *et al.*, 1965; Oliver, 1963; Hess *et al.*, 1965; Svoboda and Azarnoff, 1966). In yeast, peroxisomes are the sole site of fatty acid ß-oxidation (Veenhuis and Harder, 1987). The ß-oxidation of lipids in peroxisomes generally results in the production of H_2O_2 as the first reaction is catalyzed by an acyl-CoA oxidase, which couples acyl chain oxidation to the reduction O_2 to H_2O_2. The exception to this rule is the trypanosomal glycosome, which likely contains an acyl-CoA dehydrogenase. This enzyme is likely analogous to the enzyme found in the mammalian mitochondrion, which couples oxidation of the acyl chain with the reduction of FAD^+ to $FADH_2$ (Borst, 1989).

Some microbodies contain enzymes involved in the glyoxyate cycle, or pathway (see Tolbert, 1981). This cycle allows the conversion of fat to carbohydrate by forming malate from acetyl-CoA and glyoxylate. The enzymes that define this pathway were originally found in the glyoxysomes of the castor bean endosperm (Kornberg and Beevers, 1957). Yeast peroxisomes also contain the glyoxylate cycle enzymes malate

synthase and isocitrate lyase (van den Bosch *et al.*, 1992). Interestingly, these two enzymatic activities have also been found in vertebrate tissues, including toad bladder peroxisomes, rat liver and adipose tissue of the hibernating bear (Goodman *et al.*, 1980; Davis *et al.*, 1989; Davis *et al.*, 1990). The usefulness of the glyoxylate cycle in the latter case is readily apparent. It should be noted, however, that this pathway is not likely to be of quantitative significance in mammals, as the requirement of C_3 compounds, notably pyruvate, for gluconeogenesis in the liver and kidney is well established.

Glucose itself is metabolized in the trypanosomal glycosome. Seven glycolytic enzymes are localized to this microbody allowing the blood stream form of *Trypanosoma brucei* an unprecedented rate of energy production from glycolysis compared to other eukaryotes (Opperdoes, 1987). The glycosome, although somewhat smaller (0.2-0.3 μm in diameter), generally resembles other members of the microbody family, bound by a single lipid bilayer and occasionally containing a paracrystalline core (Opperdoes *et al.*, 1977).

The glycosome and the mammalian peroxisome contain enzymes involved in the synthesis of ether-linked lipids (plasmalogens), notably dihydroxyacetone phosphate-alkyltransferase (DHAP-AT) and alkyl-DHAP synthase. To date, only mammalian peroxisomes and trypanosomal glycosomes have been shown to contain significant quantities of these enzymes. In addition the mammalian peroxisome has recently been identified as a site of cholesterol biosynthesis (reviewed in van den Bosch *et al.*, 1992) and has been shown to be essential for normal cholesterol synthesis in man (Hodge *et al.*, 1991).

As outlined above, not all members of the microbody family conform to the strict definition of the peroxisome (de Duve and Baudhuin, 1966). Borst has suggested that the "old term microbody seems suitable to describe a multi-purpose organelle present in virtually all eukaryotic cells, containing a ß-oxidation system and a

diversity of other (often inducible) enzyme systems, depending on the organism and tissue studied." (Borst, 1989). Similar lines of thought led Osumi and Fujiki (1990) to suggest that all microbodies be reclassified as peroxisomes. The common structural and functional features shared by the members of the microbody family have also been offered by some as evidence of a common evolutionary origin for all members of the family (Borst, 1989).

1.0.2 The Study of Peroxisomes in Yeast

"Peroxisome-like" structures were described in the yeast *Saccharomyces cerevisiae* more than a decade after the discovery of the mammalian peroxisome (Avers and Federman, 1968). The *S. cerevisiae* peroxisome, however, proved to be somewhat intractable in nature, being small , scarce, and apparently uninducible (Szabo and Avers, 1969; Hoffman *et al.*, 1970). Certain "non-conventional" yeasts, primarily of *Candida sp.*, were later found to exhibit massive peroxisomal proliferation and induction of peroxisomal ß-oxidation enzymes when grown in media containing *n*-alkanes as the sole carbon source (Osumi *et al.*, 1975; Tanaka *et al.*, 1982). Similarly, growth in media containing methanol was found to induce peroxisomes and methanol assimilation enzymes of methylotrophic yeasts, including *Hansenula polymorpha, Candida boidinii,* and *Pichia pastoris* (reviewed by Veenhuis and Harder, 1987). Studies of peroxisomal proteins and proliferation in these yeasts were initially carried out using EM and classical biochemistry, and later refined when cDNA clones of various peroxisomal proteins became available. The recent development of transformation systems for unconventional yeasts, (Gleeson and Sudbery, 1988a; this study) and the demonstration of induced ß-oxidation and peroxisome proliferation and in *S. cerevisiae* (Veenhuis *et al.*, 1987), has finally allowed classical yeast genetics and heterologous gene expression techniques to be applied to the study of peroxisomal biogenesis and peroxisomal protein targeting (see below).

1.1 Peroxisome Biogenesis

A number of distinctions can be made in contrasting the biogenesis of peroxisomes from the biogenesis of other organelles, notably chloroplasts and mitochondria. Peroxisomes lack DNA (Kamiryo *et al.*, 1982), unlike chloroplasts and mitochondria, and hence all peroxisomal proteins must be encoded by genes in the nucleus. In addition, nearly all peroxisomal proteins, both lumenal and membrane bound, are imported into peroxisomes following synthesis in the cytoplasm on free polysomes, without prior passage through the endoplasmic reticulum or any other cellular membrane system (Rachubinski *et al.*, 1984; Fujiki *et al.*, 1984; reviewed in Lazarow and Fujiki, 1985). Although at present, nothing is known about the mechanism of division and segregation of peroxisomes, new peroxisomes apparently arise from the division or budding of pre-existing peroxisomes (summarized in Lazarow and Fujiki, 1985). Consistent with this idea, proliferating mammalian peroxisomes have been shown, by reconstruction of serial EM sections, to form complex, intertwined, tubular structures, designated the 'peroxisomal reticulum', and to exhibit no contiguity with the endoplasmic reticulum (Gorgas, 1984; 1985).

Yeast studies have also supported the theory of peroxisomal autonomy. Yeast cells grown under noninducing conditions always contain at least one peroxisome per cell (Veenhuis *et al.*, 1987; Thieringer *et al.*, 1991). When yeast cells are shifted to an appropriate inducing medium, peroxisomes proliferate (Osumi *et al.*, 1975; Veenhuis *et al.*, 1979). Proliferation is characterized by the elongation of preexisting peroxisomes followed by the apparent detachment of new peroxisomes from these structures (Veenhuis *et al.*, 1978; Veenhuis *et al.*, 1987; Veenhuis and Goodman, 1990). While peroxisomes will enlarge in response to import of matrix proteins under noninducing conditions, this enlargement does not appear to drive budding or "proliferation" (Distel *et*

al., 1988). Whether peroxisomes segregate and proliferate independently remains undetermined. Studies aimed at answering these intriguing questions will likely benefit tremendously from the application of both classical yeast genetics and modern molecular and cellular biological techniques.

It is noteworthy, however, that results obtained in this lab suggest at least a partial role played by the ER in targeting of a 50 kDa peroxisomal integral membrane protein to peroxisomes in rat liver (Bodnar and Rachubinski, 1991). An antiserum directed against peroxisomal integral membrane proteins (PMPs) was used to demonstrate that the 50 kDa membrane protein was, uniquely, synthesized by membrane bound polysomes of the ER. This result was corroborated by Walton *et al.* (1992), who reported that the same antiserum appears to stain the ER of human fibroblasts by immunofluorescence. Intriguingly, when alcohol oxidase, a peroxisomal matrix protein from *Pichia pastoris*, is microinjected into these cells, it not only appears to be imported into peroxisomes, but also results in a rapid (~1-2 minutes) shift of anti-PMP immunoreactivity from the ER to vesicle-like structures. The identity of the vesicles, i.e. whether or not they are peroxisomes, is unknown at present and is the subject of current research.

1.2 How are Proteins Imported into Peroxisomes?

1.2.0 PTS-1

Since most peroxisomal proteins are synthesized completely in the cytosol and are not, by and large, posttranslationally modified, the information directing proper targeting of the protein must reside entirely within the primary amino acid sequence of the mature protein (reviewed in Fujiki and Lazarow, 1985; Borst, 1989). Groundbreaking work in the laboratory of Dr. S. Subramani demonstrated that firefly luciferase was targeted to the peroxisomes of both firefly cells and monkey CV-1 cells, implying a

conserved topological signal within the luciferase molecule was involved in peroxisomal targeting (Gould *et al.*, 1987). Further studies from the same laboratory localized the essential targeting information to a tripeptide motif, with the consensus sequence Ser/Ala/Cys-Lys/Arg/His-Leu/Met (designated PTS-1, for peroxisomal targeting sequence 1), found at the carboxytermini (C-termini) of luciferase (as SKL) and a number of soluble mammalian peroxisomal proteins (Gould *et al.*, 1988; 1989). This conserved motif has been shown to be necessary and sufficient for targeting to peroxisomes of mammalian (CV-1) cells (reviewed in Subramani, 1992).

This striking, cross-species conservation of a peroxisomal targeting motif has been further emphasized by the demonstration that luciferase is also localized to the peroxisomes of the yeasts *S. cerevisiae* and *H. polymorpha* (Gould *et al.*, 1990b). Mutagenesis of the C-terminal SKL of luciferase rendered the protein cytosolic in *S. cerevisiae* (Distel *et al.*, 1992). Endogenous yeast proteins have also been found to bear PTS-1 type signals that direct them to peroxisomes. Our laboratory defined the first PTS of a yeast protein (Aitchison *et al.*, 1991). The peroxisomal trifunctional enzyme (Hydratase-Dehydrogenase-Epimerase) of *C. tropicalis*, was cloned, sequenced and found to encode a protein bearing AKI at its C-terminus (Nuttley *et al.*, 1988). When heterologously expressed in *Candida albicans* or *S. cerevisiae*, HDE was correctly targeted to and imported into peroxisomes (Aitchison *et al.*, 1991). Deletion of the C-terminal tripeptide or the terminal Ile abrogated targeting of the protein to the peroxisomal matrix of *C. albicans*, although substitutions of Gly or Gln at positions -3 and -2, respectively, could be tolerated. Neither of these mutations allowed correct targeting of HDE to peroxisomes in *S. cerevisiae*. A survey of available sequences of yeast lumenal peroxisomal proteins revealed that approximately half have C-terminal tripeptides with significant similarity to PTS-1 (Aitchison *et al.*, in press). A recent study of one of these proteins (*H. polymorpha* catalase) demonstrated that its C-terminal SKI

tripeptide is required for correct targeting of the enzyme to peroxisomes of *H. polymorpha* (Didion and Roggenkamp, 1992). As neither AKI nor SKI appears to operate a PTS in mammalian cells (Gould *et al.*, 1989), these results emphasize the varying primary sequence requirements for functional PTS-1 directed targeting between species. Nevertheless, the similarity between PTS-1 signals of the various experimental organisms can be generalized to a small, uncharged amino acid at position -3, a basic or hydrophilic amide-containing amino acid at position -2, and a hydrophobic amino acid at position -1, although certain combinations of amino acids appear to function more efficiently than others (Swinkels *et al.*, 1992).

Foreshadowing the likely discovery of other PTS-1 containing proteins, antibodies directed against the C-terminal AKI and SKL motifs recognize a number of as yet uncharacterized proteins specific to the microbody fractions of a wide variety of species. Anti-AKI and anti-SKL antibodies react specifically with certain peroxisomal proteins of *C. tropicalis, C. albicans, S. cerevisiae* and *Y. lipolytica*, suggesting that the PTS-1 type signal may target many proteins to peroxisomes in a number of yeasts (Aitchison *et al.*, 1992). Anti-SKL antibodies have also been used to identify peroxisomal proteins from mammalian cells, glyoxysomal proteins from *P. pastoris* and *Neurospora crassa*, and glycosomal proteins from *Trypanosoma brucei* that end with an epitope related to SKL (Gould *et al.*, 1990a; Keller *et al.*, 1991). In addition, luciferase is targeted to peroxisomes when expressed in *S. cerevisiae*, *Nicotiana tabacum* (Gould *et al.*, 1990b) and *Xenopus laevis* (Holt *et al.*, 1990).

1.2.1 Thiolase

As mentioned above, PTS-1-like motifs are not found in all peroxisomal proteins. Another targeting signal, designated PTS-2, has been shown to target rat liver thiolase to peroxisomes. The signal appears to reside in the cleavable aminoterminal (N-terminal) extension of the immature form of thiolase (Swinkels *et al.*, 1991). This

sequence does not exhibit extended stretches of similarity to other known peroxisomal proteins, although there may be some limited but significant conservation between rat liver thiolase and the thiolases of *S. cerevisiae*, *Y. lipolytica* and *C. tropicalis* (Aitchison *et al.*, in press). Regardless, PTS-1 and PTS-2 do not appear to account for all types of peroxisomal targeting signals, and as such the existence of at least two additional types of PTSs (one for lumenal and one for integral membrane proteins) is anticipated (Aitchison *et al.*, in press).

The maintenance of a polypeptide in an unfolded, import competent state, mediated through binding of the polypeptide to cytosolic heat shock protein 70, has been shown to be critical for import of proteins into the mitochondria (Murikami and Mori, 1990). However, Walton *et al.* (1992) demonstrated that microinjected human serum albumin 'decorated' with SKL peptides is imported into peroxisomes of human cells. Thus, peroxisomal import may not require an unfolded polypeptide chain or even a linear unit polypeptide chain. Mislocalized peroxisomal enzymes can assemble correctly in the cytosol of peroxisome deficient human cells (see below), acquiring necessary cofactors and adopting proper oligomeric structures. Following the assembly of functional peroxisomes in these cells, the enzymatic activities can rapidly relocalize to the peroxisomal compartment in the absence of new protein synthesis (Brul *et al.*, 1988). This argues against a requirement for denaturation of peroxisomal proteins prior to import. On the other hand, aminotriazole, an irreversible inhibitor of catalase, appears to interfere with catalase import following fusion of complementing strains of peroxisome deficient cells. The authors of this study suggest that aminotriazole prevents or retards unfolding of catalase, and that this, in and of itself would inhibit import of catalase (Middelkoop *et al.*, 1991).

1.3 Peroxisomal Loss/Disfunction

1.3.0 Zellweger's Syndrome and Related Disorders

The biological importance of the peroxisome is emphasized in humans by the existence of several genetic diseases that appear to accrue directly from the loss of some or all of the known peroxisomal functions (Zellweger, 1989). The amorphous clinical manifestations of these diseases underscore the variety of processes and systems in which peroxisomes play critical roles. The first disease to be identified as a peroxisomal disorder was cerebro-hepato-renal syndrome or Zellweger's syndrome. Goldfischer *et al.* (1973) demonstrated that morphologically distinct peroxisomes were absent in liver and kidney cells derived from patients with this disease. This has been shown to result in a variety of biochemical anomalies. These include increased levels of long-chain fatty acids due to deficiencies in the activities of the ß-oxidation enzymes hydratase-dehydrogenase, fatty acyl-CoA oxidase and thiolase (Moser *et al.*, 1984; Tager *et al.*, 1985; Suzuki *et al.*, 1986); reduced levels of plasmalogens resulting from low levels of the two peroxisomal enzymes, dihydroxyacetone phosphate synthase, and dihydroxyacetone phosphate-alkyltransferase (Heymans *et al.*, 1983; Datta *et al.*, 1984; Schutgens *et al.*, 1984; Schrakamp *et al.*, 1985), that catalyze the initial steps in plasmalogen biosynthesis; low levels of bile acids and bile acid synthesis (Hanson *et al.*, 1979; Pedersen and Gustafsson, 1980); impaired degradation of pipecolic acid due to a loss of pipecolic acid oxidase (Wanders *et al.*, 1989); increased levels of phytanic acid resulting from deficiencies in phytanic acid α-oxidation (Lazarow and Moser, 1989). In addition, a number of enzymatic activities, including catalase, D-amino acid oxidase, L-α–hydroxyacid oxidase, and alanine:glyoxylate aminotransferase, are present at near normal levels but appear to be mislocalized to the cytosol (Lazarow and Moser, 1989).

A critical observation was made by Lazarow and co-workers (Santos *et al.*, 1988a; 1988b), who showed that peroxisomal membranes ('peroxisomal ghosts') were present in the fibroblasts of Zellweger's syndrome patients. Using antisera directed against peroxisomal integral membrane proteins (22, 53 and 69 kDa) they were able to demonstrate the existence of these membranes by immunofluorescence. This evidence implied that the underlying defect in Zellweger's and related diseases was one of defective import of matrix proteins into the peroxisome, while also suggesting that PMPs are directed to peroxisomes in a manner different from that of lumenal proteins. Further studies indicate that these peroxisomal ghosts also retain other membrane proteins (50 and 35 kDa) and, in some cases, certain lumenal proteins, notably thiolase (Wilson, 1991; Suzuki *et al.*, 1992).

There are at least eight different gene products involved in peroxisome biogenesis in humans (Yajima *et al.*, 1992). This has been determined through cell fusion experiments with cultured fibroblasts from patients with Zellweger's syndrome and clinically related ailments. These studies have delineated eight distinct complementation groups that do not correlate with previously established clinical categories of peroxisome deficient disorders (Yajima *et al.*, 1992). In addition, Gärtner *et al.* (1992) have shown that mutations in PMP70 correlate with Zellweger's syndrome. The mutations were found in two patients, both of whom had been assigned to complementation group I. Interestingly, no mutations were found in screened cDNAs of an additional 19 group I patients, leading the authors to suggest that this complementation group may be genetically heterogeneous. Zellweger's syndrome has also been modeled in Chinese hamster ovary cells. One such mutant cell line has been shown to be complemented by a gene encoding a 35 kDa membrane protein dubbed peroxisome -assembly-factor -1 (PAF-1) (Tsukamoto *et al.*, 1991). The human homologue of this gene has been cloned

and shown to be mutated in a patient with Zellweger's syndrome (Shimozawa *et al.*, 1992).

1.3.1 PAS Mutants in Yeast

As mentioned earlier, recent advances in understanding the molecular biology and genetics of certain nonconventional yeasts, as well as the demonstration of peroxisome proliferation in *S. cerevisiae*, have made the study of peroxisome biogenesis in yeasts amenable to genetic dissection. The targeting and/or assembly paradigm might therefore be elucidated in these yeasts with the aim of understanding peroxisome biogenesis in more complex eukaryotes. To this end, common strategies have been employed recently to generate peroxisome assembly mutants in yeasts such as *S. cerevisiae*, *H. polymorpha* , and *Pichia pastoris*.

These mutants were selected by screening for their inability to grow on oleic acid (and/or methanol) as a sole carbon source, combined with the cytosolic localization of proteins normally located in the peroxisomal matrix (Erdmann *et al.*, 1989; Cregg *et al.*, 1990, Gould *et al.*, 1992). Two genes have been isolated and characterized thus far, through the complementation of the *S. cerevisiae* mutants *pas-1* and *pas-3*. The *PAS-1* gene encodes a 117 kDa soluble protein (Erdmann *et al.*, 1991). It is part of a family genes encoding proteins bearing a conserved, duplicated region, apparently involved in ATP binding and hydrolysis. The physiological roles of these proteins are not well understood. Some members of this family have been shown to play roles in such disparate cellular processes as vesicular fusion along the secretory pathway and control of the cell cycle. *PAS-3* encodes a 50 kDa peroxisomal integral membrane protein and its function is also unknown (Höhfeld *et al.*, 1991).

1.4 *Y. lipolytica*

1.4.0 Lifecycle

The yeast *Y. lipolytica*, known variously as *Candida lipolytica* or *Saccharomycopsis lipolytica*, is a dimorphic yeast. It forms true hyphae when metabolically stressed, e.g. when entering stationary phase in nutrient depleted media. During hyphal growth, only the apical cell of a given hypha divides. Under optimal growth conditions, it grows as a budding yeast, analogous to *S. cerevisiae*. Beckerich *et al.* (1985) classified it as a type 2 yeast, with the (A/B) haploid and diploid phases being of equal importance in its life cycle. This readily allows standard complementation studies with various mutants of interest. *Y. lipolytica* is an industrially important yeast, used to produce citric acid (Beckerich *et al.*, 1985). In addition, harnessing its polypeptide secretory pathway is of industrial interest as it secretes large quantities (1-2 g/L of culture medium) of an alkaline extracellular protease, under certain conditions (Davidow *et al.*, 1987). It also displays very strong growth on alkane or fatty acid containing media. This is accompanied by a dramatic induction of peroxisomes (Tanaka *et al.*, 1982).

1.4.1 Development of Genetic Manipulation

The combination of simple, defined genetics and readily inducible peroxisomes would seem to make *Y. lipolytica* an ideal organism to study peroxisome biogenesis. However, until recently, quick and efficient means of cloning complementing genes from *Y. lipolytica* had not been developed. Simple auxotrophic strains have been isolated, and their complementing genes cloned (Xuan *et al.*, 1988). With a great deal of difficulty, autonomously replicating sequences have also been isolated (Fournier *et al.*, 1991). These basic elements have been used to construct useful *Esherichia coli/Y. lipolytica* shuttle vectors. To date these vectors have not been

exploited to clone genes from *Y. lipolytica*, as none of them have been used successfully to construct a genomic library. Rather, cloning has been accomplished using integrating vector libraries followed by labour intensive isolation of genomic DNA, Southern blotting, endonuclease digestion, ligation and attempted rescue of complementing clones in *E. coli* (Kurischko *et al.*, 1992; Prodromou *et al.*, 1991). The construction of genomic libraries using these autonomously replicating vectors should make rapid complementation and cloning feasible. The short range goals of this study were to create such a library, isolate peroxisome deficient mutants of *Y. lipolytica* , clone and sequence the genes responsible. The overall goal was to make *Y. lipolytica* a model yeast system for the study of peroxisome biogenesis.

2. MATERIALS AND METHODS

2.0 Materials

Chemicals and reagents provided below form a list of critical reagents used for the research described in this thesis.

amino acids	Sigma Chemical Company (SCC)
ampicillin	SCC
bacto-agar	Difco Laboratories (DL)
bacto-peptone	DL
Bio-Rad protein assay	Bio-Rad Laboratories (Canada) Ltd.
Coomassie Brilliant Blue (R-250)	Bethesda Research Laboratories Inc. (BRL)
Cytochrome-c (horse heart)	SCC
deoxyribonucleotides (dATP, dCTP, dGTP, dTTP)	Pharmacia Canada Inc.
diethyl pyrocarbonate (DEPC)	SCC
dithiothreitol	SCC
fluorescein (FITC)-conjugated AffiniPure goat anti-rabbit (or guinea pig) IgG (H+L)	Jackson ImmunoResearch
Freund's Complete Adjuvant	SCC
GTG Agarose	FMC BioProducts
hemoglobin	SCC
hydrogen peroxide	SCC
isopropyl ß-D-thiogalactoside	Gibco/BRL Canada
molecular weight standards:	
(i) 1 kb DNA ladder (75-12, 216 bp)	Gibco/BRL Canada
(ii) Prestained SDS-PAGE Standards (Low and High Range)	Bio Rad Laboratories (Canada) Ltd.
(iii) Dalton Mark VII-L molecular weight markers for SDS-PAGE (14 000-66 000 Da)	SCC
(iv) phosphorlyase b (rabbit muscle; 97 000 Da)	SCC
(v) b-galactosidase (*E. coli*; 116 000 Da)	SCC
nitrocellulose (pore size-0.45 mm)	Schleicher and Schuell Inc.
Nonidet P-40 (NP-40)	SCC
Nystatin	SCC
oligo (dT) cellulose (type 3)	Collaborative Research Inc.
ovalbumin	Miles Laboratory (Pty) Ltd.
Sephadex G-50 (medium)	Pharmacia (Canada) Ltd.
Triton X-100	SCC
Tween 40 (polyoxyethylene sorbitan monolaurate)	SCC

x-gal (5-bromo-4-chloro-3-indolyl-b-D-galactoside)	Gibco/BRL (Canada) Ltd.
yeast extract	Merck Ltd.
yeast nitrogen base (without amino acids)	DL

2.0.1 Radiochemicals

^{125}I-protein A (>30 mCi·mg^{-1} of total protein A, 0.1 Ci·ml^{-1})	Amersham Canada Ltd.
[γ-^{32}P]ATP (3 000 Ci·mmol^{-1}, 10 mCi·ml^{-1})	Amersham; Dupont/NEN Canada Inc.
[α-^{32}P]ATP (3 000 Ci·mmol^{-1}, 10 mCi·ml^{-1})	Amersham; Dupont/NEN Canada Inc.

2.0.2 Enzymes

All enzymes were used as per the manufactures' recommendations, unless otherwise noted in the methods section.

DNA Ligase (T4)	Gibco/BRL; New England Biolabs (NEB)
DNA polymerase I Klenow fragment (*E. coli*)	NEB
DNA polymerase, modified T7 (Sequenase)	United States Biochemical Corp.
polynucleotide kinase	NEB
restriction endonucleases	Boehringer Mannheim; Gibco/BRL; Fermentas; NEB; Pharmacia
RNase I "A" (bovine pancreas)	Pharmacia (Canada) Inc.
Zymolyase T100	Seikagaku America Inc.

2.0.3 Oligodeoxyribonucleotides

M13 forward (5'-GTAAAACGACGACGGCCAGT-3') and reverse (5'-CAGGAAACAGCTATGAC-3') primers were purchased from United States Biochemical Corp. or Pharmacia (Canada) Inc. Custom synthesis of oligodeoxyribonucleotides was carried out at the Central Facility of the Institute for

Molecular Biology (MOBIX), McMaster University, Hamilton, Ont.. The sequences of the oligodeoxyribonucleotides, all of which were used for the sequencing of the overlapping regions of the inserts in plasmids pO1, pO2 and pO3, are found in the following table (2.0.3).

Table 2.0.3: Custom Synthesized Oligodeoxyribonucleotides

Number	Sequence
Pr1	5'-ACCCGTGAATCCACAA-3'
Pr2	5'-AAGTAGGCCTGCAGA-3'
Pr3	5'-TGGTTTACAGGCTCGCA-3'
Pr4	5'-CTGATCTCACAGCCATTG-3'
Pr6	5'-ATGGAGGGTGTCAAGA-3'
Pr7	5'-CCAGGGAGACTCTGGAG-3'
Pr8	5'-CTCCAGAGTCTCCCTGG-3'
Pr9	5'-AGACTGCCACATTTCT-3'
Pr10	5'-ATGGTTTCAGACTTGA-3'
Pr11	5'-AGCGGTTGCAGGGAGAG-3'
Pr12	5'-TTACTCCTCCGGGCACTA-3'
Pr13	5'-CCGAGACTTACTACTC-3'
Pr14	5'-TCTTAGTCGAGGATGTT-3'
Pr15	5'-TTTCAACACGACAGC-3'
Pr16	5'-AGGCTCTTACTCGAAAG-3'
Pr17	5'-CACTATTAATACTGT-3'

2.0.4 Antisera

Anti-SKL serum was the generous gift of Dr. S. Subramani, Dept. of Biology, University of California at San Diego, La Jolla, CA. It was generated by

injecting rabbits with a chemically synthesized peptide (NH_2-CRYHLKPLQSKL-COOH) of which the last nine amino acids correspond to the C-terminal nine amino acids of rat fatty acyl-CoA oxidase (Gould *et al.*, 1990a).

2.0.5 Yeast Strains

The *Y. lipolytica* strains E122 (*MatA, ura3-302, leu2-270, lys8-11*) and 22301-3 (*MatB, ura3-302, leu2-270, his1*) used in this study were grown at 30° C in the media described below. Minimal media were supplemented with uracil, leucine, histidine and lysine, each at 50 $\mu g \cdot ml^{-1}$, as required.

2.0.6 Bacterial Strains and Growth Media

E. coli DH5α cells were purchased from Gibco/BRL Canada and cultured in LB medium (Luria Broth) containing 1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.5. When required, ampicillin or tetracycline were added to final concentrations of 150 $\mu g \cdot ml^{-1}$ and 12.5 $\mu g \cdot ml^{-1}$, respectively (Ausubel *et al.*, 1992).

2.1 Mutagenesis of Yeast

Y. lipolytica E122 cells were grown overnight in 10 ml of YEPD (2% yeast extract, 1% bacto-peptone, 2% glucose). One ml of this culture was used to inoculate 250 ml of YEPD Cells were harvested (OD_{600} = 0.6), washed twice with 0.1 M sodium citrate buffer (pH 5.5), and suspended in 50 ml of this buffer containing 25 $mg \cdot ml^{-1}$ 1-methyl-3-nitro-1-nitrosoguanidine (NTG) (Gleeson and Sudbery, 1988b) The cell suspension was kept at room temperature for 20 min with gentle rotation. NTG-treated cells were then washed 5 times in water and resuspended in 50 ml of YEPD medium. Glycerol was added to 15% (w/v) final concentration, and 1 ml aliquots were frozen at -70° C. This treatment resulted in the death of approximately 95% of the cells. This protocol was carried out by Bill Nuttley.

2.1.0 Nystatin Enrichment

0.2 ml of frozen, mutagenized cells was used to inoculate 10 ml of YEPD medium. The cells were grown for 12h to an OD_{600} of 2.0. The cells were washed and resuspended in 2 ml of water. 0.4 ml of suspended cells was used to inoculate 10 ml of YNO (0.67% yeast nitrogen base, 0.05% (w/v) Tween 40, 0.1% (w/v) oleic acid) medium, which was then incubated for 6h. Nystatin was added to a final concentration of 10 $\mu g \cdot ml^{-1}$, and incubation was continued for 90 min (Snow, 1966). The cells were harvested by centrifugation, washed with water, resuspended in 1 ml of 15% (w/v) glycerol, and frozen at -70° C.

2.1.1 Screening for Mutants

50 µl of frozen, nystatin-enriched, mutagenized cells was diluted to 3 ml with water. 100 µl of this cell suspension was spread onto YEPD agar plates. The colonies (~300/10 cm plate) were replica plated onto selective YNA (0.67% yeast nitrogen base, 0.1% yeast extract, 2% sodium acetate) and YNO media (Erdmann *et al.*, 1989). Colonies that failed to grow on YNO but grew well on YNA were designated oleic acid-nonutilizing (*ole*$^-$) mutants.

2.2. Immunofluorescence

Ole$^-$ mutants were screened for the presence or absence of peroxisomes by indirect immunofluorescence using rabbit anti-SKL antibodies, essentially as previously described (Aitchison *et al.*, 1992).

2.2.0 Cell Preparation

Briefly, cells were grown overnight (to late-log phase) in oleate induction medium and fixed overnight in 3.7% formaldehyde at 4° C. After washing in 1.2 M sorbitol, 100 mM KP_i (pH 7.5), the fixed cells were converted to spheroplasts with Zymolyase 100T. Staining was carried out essentially as described by Pringle *et al.*

(1991). Rabbit anti-SKL antiserum was used as the primary label at a 50^{-1} dilution and was visualized with fluorescein-conjugated goat anti-rabbit IgG, used at a 20^{-1} dilution. Following staining with the secondary antibodies, cells were mounted under cover slips in 75% glycerol, 2.5% 1,4-diazabicyclo[2,2,2]octane.

Immunofluorescence microscopy was carried out with a Zeiss MC63A microscope, equipped for fluorescence, and photographed with a Chinon-7m 35 mm SLR camera on Kodak TMAX-400 film.

2.3 Organelle Isolation

2.3.0 Isolation of Peroxisomes and Mitochondria

Cells were grown on oleic acid-induction medium (0.67% yeast nitrogen base, 0.5% yeast extract, 0.5% peptone, 0.1% glucose, 0.1% (w/v) oleic acid, 0.5% (w/v) Tween 40; Erdmann *et al.*, 1989) for 16h. The cells were harvested by centrifugation, washed with water, and converted to spheroplasts with Zymolyase 100T. The spheroplasts were collected by centrifugation and homogenized in 1 M sorbitol, 5 mM MES [2-(N-morpholino)ethanesulfonic acid], pH 5.5 (YL disruption buffer, as per Aitchison *et al.*, 1992). The postnuclear supernatant fraction was isolated by centrifugation at 1000 x g for 10 minutes and recentrifuged at 20,000 X g for 20 min to obtain a pellet (20 kgP) enriched for peroxisomes and mitochondria (Aitchison *et al.*, 1992). The 20,000 X g supernatant (20 kgS, enriched for cytosol) and 20 kgP were assayed for marker enzyme activities (see below).

2.3.1 Purification of peroxisomes

When further purification of peroxisomes was required, to separate them from mitochondria for example, the 20 kgP was resuspended in YL disruption buffer and loaded onto a discontinuous sucrose gradient (4.67 ml of 25%, 7 ml of 35%, 14 ml of 42%, 7 ml of 53% (w/w) sucrose in YL disruption buffer) and centrifuged in a Beckman

VTi50 at 35 000 rpm (130 000 x g_{max}) for 1 hr at 4° C. The gradient was fractionated in 2 ml increments with a 3mm bore stainless steel tube, Tygon™ tubing (4mm internal bore) and a peristaltic pump. To assess purity, and determine organellar composition of each of the fractions, marker enzyme analyses (see below) were carried out on each fraction.

2.4 Marker Enzyme Assays

2.4.0 Catalase Assay

The assay for catalase was performed essentially as described by Baudhuin *et al.* (1964). 50 µl of the fraction to be analyzed was diluted with an equal volume of 2% (w/v) Triton X-100 and incubated on ice for ≥2 minutes. One ml of substrate (20 mM imidazole (pH 7.0), 1 mg·ml^{-1} BSA, 0.01% (v/v) H_2O_2) was then added and the sample was incubated in ice water for 5-15 minutes. 2 ml of $TiOSO_4$ (titanium (IV) oxysulphate in 2 N H_2SO_4) was then added to stop the reaction and convert residual H_2O_2 to yellow titanium peroxysulphate. The colour development was quantitated by measurement of the OD_{410}.

2.4.1 HDE Assay

ß-hydroxyacyl-Coenzyme A dehydrogenase activity was assayed essentially as described by Osumi and Hashimoto (1979), except that the substrate, ß-hydroxybutyryl-CoA, was added directly to the reaction mixture, consisting of 10 µM ß-hydroxybutyryl-CoA, 0.1 M Tris-HCl (pH 10.2), 0.1 M KCl, 0.1% (w/v) Triton X-100, 1 mM NaN_3, 0.1 M NAD^+. NADH production was used to measure enzyme activity and was quantitated by measuring the increase in OD_{340} over time. The rate of NADH production was estimated using an extinction coefficient E^{1M}_{1cm}= 6220 at 340 nm (Clarke and Switzer, 1977).

2.4.2 Cytochrome *C* Oxidase Assay

Cytochrome *c* oxidase activity was measured essentially as described by Cooperstein and Lazarow (1951). The assay buffer (2.5 ml of 0.38 $mg \cdot ml^{-1}$ cytochrome *c*, in 0.3 M ammonium acetate (pH 7.4) and .250 ml of 10% (w/v) Triton X-100) was mixed in a 3 ml plastic spectrophotometer cuvette. The cytochrome *c* was reduced by the addition of a small amount of Na hydrosulfite (generally a spatula tipful). The reduction was quantitated as an increase in OD_{550}. The cuvette was mixed vigorously, an appropriate amount of the sample was added, and the rate of cytochrome *c* oxidation was measured as the decrease in OD_{550}.

2.5 DNA Isolation From *Y. lipolytica*

2.5.0 Genomic DNA

A starter culture was grown overnight in 100 ml of YEPD at 30° C. This was pelleted and resuspended in 30 ml of sterile water. Ten ml of this solution was used to inoculate 1 L of sterile YEPD which was subsequently cultured overnight, with shaking, at 30° C. The cells were pelleted at 5000 x g for 10 minutes and washed three times with water. The pellet was transferred to a sterile mortar, prechilled in a liquid nitrogen bath. The flash frozen cells were ground to a fine powder with a sterile pestle and resuspended in 150 ml of 150 mM NaCl, 50 mM Tris·HCl (pH 7.5), 5 mM EDTA, 5% SDS.

The solution was warmed to room temperature under warm tap water, and extracted three times for 20 minutes with 150 ml of 25:25:1 phenol:chloroform:iso-amyl alcohol. After extraction the aqueous phase was adjusted to 0.2 M potassium acetate (pH 5.5) and nucleic acids were precipitated overnight at -20° C, following the addition of 2.5 volumes of cold ethanol. The nucleic acids were pelleted at 4° C, by centrifugation at 9000 x g, and washed with cold 70% ethanol. The pellet was dried under vacuum.

To isolate total RNA, the pellet was resuspended in 100 ml of 3 M LiCl and left at 4° C for two days. The RNA was pelleted as above, washed twice with 70% ethanol, dried in a desicator and stored for later use. The DNA was precipitated from the LiCl solution with 2.5 volumes of ethanol, washed twice with 70% ethanol, dried and resuspended in 12 ml of TE (pH 8.0). 12 g of $CsCl_2$ and 960 µl of 10 mg·ml^{-1} ethidium bromide were added and the resulting solution was divided amongst 6, 2 ml polycarbonate Beckman ultracentrifuge tubes. The tubes were sealed and spun at 100 000 rpm in a Beckman TLV-100 rotor in a Beckman TL-100 mini-ultracentrifuge. The banded genomic DNA was removed with a hypodermic syringe, pooled, and extracted with water saturated iso-butanol to remove the ethidium bromide. The solution was then dialyzed twice with 1 L of TE (pH 8.0) at 4° C for a minimum of 4 hours per dialysis.

2.5.1 Plasmid DNA isolation

Plasmid DNA was isolated from *Y. lipolytica* by a procedure modified from Ausubel *et al.* (1992). Briefly, 10 ml of selective (minimal) medium was inoculated with the transformant of interest and incubated at 30° C for 3-5 days (late log phase). The cells were pelleted and washed twice with sterile water, resuspended in 200 µl of breakage buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris·HCl (pH 8.0), 1 mM EDTA) and added to a microcentrifuge tube containing a roughly equal volume of glass beads. 200 µl of phenol:chloroform solution was added to this and the tube was vortexed at highest speed for two minutes. The sample was then spun at the highest speed in a microcentrifuge for 5 minutes. The aqueous phase was then removed and extracted with an equal volume of chloroform, mixed with 0.1 vol. of Na acetate (pH 5.0) and nucleic acids were precipitated with 2.5 volumes of ethanol at -20° C for at least 30 minutes. The nucleic acids were pelleted, washed once with 70% ethanol, dried under vacuum and redissolved in 100 µl of sterile water. The DNA was transformed into *E.*

coli by electroporation. Analytical amounts of adequately pure plasmid DNA could then be retrieved using the *E. coli* DH5α small scale isolation procedure (see below).

2.6 Plasmid DNA Isolation From *E. coli*

Plasmid DNA was recovered from *E. coli* as per Ausubel *et al.* (1992) as adapted from Birnboim and Doly (1979) and Birnboim (1983). Cells from a single colony were used to inoculate sterile LB containing 150 $\mu g \cdot ml^{-1}$ ampicillin (LBamp).

2.6.0 Small scale isolation

The protocol used is a modified version of the alkaline lysis method detailed in Ausubel *et al.* (1992). Small scale isolations were carried out with 1.5 ml of cells from a saturated LBamp culture. The cells were pelleted in a microfuge for 20 seconds and the media supernatant was removed by aspiration. The cells were then resuspended with vortexing in 100 µl of TE (pH 8.0) at room temperature. Following this, 200 µl of freshly prepared NaOH/SDS solution (0.2N NaOH, 1% SDS) was added, mixed gently, and incubated on ice water. After temperature equilibration, 150 µl of ice cold KAc solution (3 M potassium/5 M acetate, pH 4.8) was added and mixed vigorously for 3 seconds at highest speed on a vortex, and the tube was placed back on ice for 2 minutes. The mixture was then centrifuged at highest speed on a microfuge for 3 minutes and the supernatant was transferred to a fresh tube. 0.8 ml of room temperature 100% ethanol was added and mixed by inversion. After a 1 minute incubation the sample was microfuged for 1 minute and the supernatant was removed by aspiration. 1 ml of room temperature 70% ethanol was added, the sample was mixed on a vortex and microfuged for 3 minutes. The supernatant was removed by aspiration, with special care being taken to remove as much of the ethanol as possible. The pellet was subsequently dried under vacuum, and dissolved in 20 µl of TE (pH 8.0) containing 20 $\mu g \cdot ml^{-1}$ of RNase A.

2.6.1 Large scale isolation

Isolation of plasmid DNA from large volume LBamp cultures (>100 ml) was performed essentially as described in the large scale alkaline lysis method in Ausubel *et al.* (1992) except that the step using chicken egg lysozyme was omitted. The plasmid DNA was also purified twice by $CsCl_2$ density gradient centrifugation as per Ausubel *et al.* (1992) except the $CsCl_2$/EtBr plasmid solution was sealed into 2 ml polycarbonate Beckman ultracentrifuge tubes and spun at 100 000 rpm in a Beckman TLV-100 rotor in a Beckman TL-100 mini-ultracentrifuge (q.v. 2.5.0). The DNA was extracted with water saturated 1-butanol to remove the ethidium bromide. The solution was then dialyzed three times against 1 L of TE (pH 8.0) at 4° C for a minimum of 4 hours per dialysis.

2.7 DNA Manipulation

2.7.0 Quantitation

The concentrations of pure DNA or RNA samples were determined on a Beckman DU 64 spectrophotometer by standard OD_{260} measurements (Ausubel *et al.*, 1992).

2.7.1 Restriction Endonuclease Digestion

Restriction of DNA was performed essentially as outlined in Ausubel *et al.* (1992) with due consideration given to manufacturer recommendations and specifications.

2.7.2 Inactivation of DNA Modifying Enzymes

DNA modification enzymes were either heat inactivated as per Ausubel *et al.* (1992) or manufacturer recommendations, or were inactivated and removed with phenol:chloroform extraction. Briefly, the reaction mixture would be adjusted to a manageable volume (>100 µl) with water and extracted with an equal volume of 50:50:1 phenol:chloroform:iso-amyl alcohol. The microfuge tube would be spun briefly (~20

seconds), the aqueous phase transferred to a new tube and the step repeated with an equal volume of 50:1 chloroform:iso-amyl alcohol. The aqueous phase would be removed once again and precipitated as below.

2.7.3 DNA Precipitation

To precipitate aqueous DNA, the solution would generally be mixed with 0.1 volume of 3 M Na acetate (pH 5.0), precipitated with 2.5 volumes of absolute ethanol at -20° C for >15 minutes, pelleted in a microfuge at maximum speed for >10 minutes, washed with 70% ethanol, respun, dried under vacuum, and redissolved in an appropriate volume of water or TE (pH 8.0).

2.7.4 Generation of 'Blunt Ended' DNA Fragments

In the work described here, the need for 'blunt' ends only arose only for DNA fragments with 5' overhanging ends. Thus, the only enzyme used for this purpose was the Klenow fragment of *E. coli* DNA pol I. Reactions with this enzyme were carried out as per Ausubel *et al.* (1992). In general, all four dNTPs were added to the DNA solution to a final concentration of ~0.2 mM. 1 to 5 U of Klenow fragment was then added and followed by a 15 minute incubation at 30° C. The enzyme was then inactivated by phenol:chloroform extraction, heat inactivation (75° C) or by addition of 1 µl of 0.5 M EDTA

2.7.5 Dephosphorylation of 5' Termini

When dephosphorylated DNA fragments were required, calf intestinal alkaline phosphatase was used according to the recommendations of Ausubel *et al.* (1992). Generally, the enzyme was added to DNA in the desired quantity, generally 0.1 U/1-20 pmol of DNA termini, in any standard restriction buffer and incubated at 37° C for 30 minutes to remove accessible 5' phosphates. The enzyme was then inactivated, either by phenol:chloroform extraction or heating at 75° C for 10 minutes.

2.7.6 Ligation

DNA ligations performed using T4 DNA ligase in 1X KGB (200 mM K-glutamate, 50 mM Tris-acetate (pH 7.6), 10 mM Mg-acetate, 50 $\mu g \cdot ml^{-1}$ bovine serum albumin, 0.5 mM ß-mercaptoethanol) supplemented with 1 mM ATP at 18° C for cohesive end ligations, or in BRL Ligase buffer at room temperature for blunt end ligations. The concentrations of DNA used were determined empirically or based on the recommendations of Ausubel *et al.* (1992).

2.7.7 DNA Purification by Electroelution

DNA fragments were separated by agarose gel electrophoresis with 0.4-1.5% SeaKemGTG agarose gels in 1X TBE. The desired bands were then excised with a razor blade and the DNA purified with a unidirectional electroeluter (model UAE, International Biotechnologies Inc.). The apparatus was filled with 0.5X TBE and the gel slice placed in the chamber. 80 µl of 7.5 M ammonium acetate, 0.01% bromophenol blue was layered in the elution path. The DNA was then eluted from the gel slice at 100 V (~0.002 amp) for 1 hr and was trapped in the high salt. The DNA and salt were then removed from the well in a total volume of 350 µl, the DNA was precipitated following the addition of 1 ml of cold ethanol, pelleted, washed, and dried as previously described (q.v. 2.7.3)

2.7.8 Chemical Transformation of *E. coli*

The cells used primarily for chemical transformation in this work were DH5α Sub-cloning Efficiency cells (SE cells) from Gibco/BRL Canada. Transformation was carried out essentially as described by the manufacturer. Briefly, a 50 µl aliquot of SE cells was thawed on wet ice, and a volume of DNA (≤2.5 µl, 1 pg-1 µg total) was added with gentle mixing. The mixture was left on ice 15-30 minutes and then heat shocked in a 37° C water bath for 20-30 seconds. This was followed by a 2-5 minute incubation on ice. 0.95 ml of room temperature LB was then added, and the sample was

then incubated with agitation at 37° C for 30-60 minutes. Some or all of the transformants were then spread on an LBamp plate and incubated overnight at 37° C to allow colonies to develop. If 'blue-white' screening was necessary to select recombinants, 75 µl of 2% X-gal in dimethylformamide was spread on the LBamp plate prior to plating of the cells.

2.8 Electroporation

2.8.0 *E. coli*

1 L of LB was inoculated with 10 ml of a fresh overnight culture of *E. coli* DH5α cells. The cells were incubated at 37° C with vigorous shaking to an OD_{600} of 0.5-1.0. The flask was chilled on ice for 15-30 minutes and then the cells were pelleted at 4000 x g_{max} at 4° C for 15 minutes. The cells were resuspended in an equal volume of ice cold wash medium, 1 mM HEPES (pH 7.0), and pelleted as above. The cells were resuspended in 500 ml of cold wash medium and pelleted once again. The cell pellet was resuspended in ~20 ml of ice cold, 10% (v/v) glycerol and pelleted at 5000 x g_{max} for 15 minutes. The cells were then resuspended to a final volume of 2-3 ml in ice cold, 10% glycerol, frozen in a dry ice/ethanol bath and stored in aliquots at -70° C.

For transformation, a 20 µl aliquot of the cells was thawed on wet ice, mixed with 1 µl of the plasmid DNA solution and incubated on ice for 5 minutes. The aliquot was placed between the "bosses" of a BRL microelectroporation chamber (width ~0.15 cm) and subjected to a voltage pulse of 400 V boosted to ~2.5 kV using a BRL Cell-Porator in combination with a BRL Voltage Booster. The voltage booster uses a 2 µF capacitor and was adjusted to the 4 kΩ resistance setting.. The cells were transferred immediately to 1 ml of room temperature LB, incubated for one hour at 37° C with rotation, spread on an LBamp plate, and incubated overnight at 37° C to allow colonies to develop.

2.8.1 *Y. lipolytica*

A 50 ml YEPA (as YEPD except containing 2% acetate instead of glucose) culture of *Y. lipolytica* was grown at 30° C with vigorous shaking (>250 rpm) overnight to an OD_{600} of 0.5 - 1.0. 1 M DTT was added to a final concentration of 10 mM and the culture was incubated at 30° C for an additional 15 minutes shaking at ~100 rpm. Cells were collected in a clinical centrifuge and washed once with 50 ml of sterile water and respun at room temperature to prevent DTT precipitation. The supernatant was discarded and all subsequent manipulations were performed at 0 - 4° C. The cell pellet was washed once with 10 ml of sterile 1 M sorbitol, respun and resuspended in 100 ml of ice-cold, 1 M sorbitol.

A 20 µl aliquot of the cell suspension was mixed with 1 µl of the plasmid DNA solution and incubated on ice for 5 minutes. The aliquot was placed between the "bosses" of a BRL microelectroporation chamber (width ~0.15 cm) and subjected to a voltage pulse of 250 V boosted to ~1.6 kV using a BRL Cell-Porator in combination with a BRL Voltage Booster. The voltage booster uses a 2 µF capacitor and was adjusted to the 16 kΩ resistance setting.. The cells were transferred immediately to 100 µl of room temperature 1 M sorbitol and plated on selective YNA plates containing 1 M sorbitol. Transformants were replica plated onto the desired selective media as soon as colonies appeared, usually 36-48 hours following electroporation.

2.9 Sequence Analysis of DNA

'Sanger' dideoxy DNA sequencing was generally performed using the Sequenase™ Kit from United States Biochemicals (Sanger *et al.*, 1977; Tabor and Richardson, 1987). All sequencing was carried out using double-stranded templates as per Zhang *et al.* (1988). Briefly, 1-3 µg of plasmid DNA in 0.2 mM EDTA was denatured by the addition of freshly prepared NaOH to a final concentration of 0.13 M.

This solution was incubated at room temperature for 5 minutes and was then neutralized by the addition of ammonium acetate to 0.2 M. Approximately 2 volumes of ethanol at 4° C was then added and the denatured DNA was precipitated, washed and dried (q.v. DNA Precipitation). The DNA was resuspended in water and sequenced as per the Sequenase™ Manual (USB) using [α-^{32}P]dATP (3000 Ci/mmol) and 1 pmol of primer. The DNA fragments were separated by electrophoresis using either a 6% acrylamide gel (40:1 acrylamide:bis-acrylamide, 8M urea in 1X TBE) or a Long Ranger™ gel (J. T. Baker Chemicals) in 0.6X TBE at 60 watts. The gels were dried and exposed to Kodak X-OMAT K-1, or X-OMAT AR film.

2.10 Analysis of Proteins

2.10.0 Bio-Rad Protein Determination

Protein concentrations were determined using the Bio-Rad modified Bradford assay (Bradford, 1976). Protein samples were diluted to 100 µl in 0.1 M Na-P_i buffer (pH 7.0) and 1 ml of Bio-Rad protein assay reagent was added. Protein was quantitated by measuring the OD_{595} and comparing the value to those of a set of standards done at the same time. Ovalbumin was used as the standard and samples were generally made up at 0, 5, 10, 20, 40 and 80 µg.

2.10.1 Yeast Protein Isolation by Glass Bead Disruption

Preparation of crude yeast cell lysates by glass bead disruption was carried out essentially as described by Ausubel *et al.* (1992). Cells were pelleted in microfuge tubes and resuspended in an appropriate volume of disruption buffer (20 mM Tris·HCl (pH 7.9), 10 mM $MgCl_2$, 1 mM EDTA, 5% (v/v) glycerol, 1 mM DTT, 0.3 M ammonium sulfate, 1 mM PMSF) for protein isolation. An equal volume of glass beads (Sigma Chemical Corp.) was added to the tube and the mixture was vortexed at high speed for 30 seconds followed by a 30 second incubation on ice. The vortexing steps were repeated in

sequence three more times. The tubes were then spun in a microfuge at highest speed for two minutes to pack the beads and pellet cellular debris. As much of the supernatant as possible was then transferred to a second tube and stored at -20° C or used as required.

2.10.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Proteins were separated on the basis of size using the method of Laemmli (1970). The separating gels were made to concentrations of 7.5, 10, or 12.5% acrylamide (30:0.8 acrylamide:N,N'-methylene-bis-acrylamide), 0.37 M Tris-HCL (pH 8.8), 0.1% SDS, 0.5% (v/v) TEMED (N,N,N',N'-tetramethylethylene-diamine), 0.2% ammonium persulphate, and the stacking gels were 3% acrylamide (30:0.8 acrylamide:N,N'-methylene-bis-acrylamide), 0.06 M Tris-HCl (pH 6.8), 0.1% SDS, 0.1% (v/v) TEMED (N,N,N',N'-tetramethylethylene-diamine), 0.1% ammonium persulphate, as per Fujiki *et al.* (1986). A Hoefer vertical slab mini-gel apparatus was used to cast and run the gels. The protein samples, boiled 5 minutes in 0.0625 M Tris-HCl (pH 6.8), 2.0% SDS, 10% sucrose, 0.01 M DTT, 0.001% bromophenol blue, were run on the gels at 180-240 V for 20-40 minutes, generally until the dye front reached the bottom of the separating gel. The gels were run in 1X SDS-PAGE running buffer (0.4 M glycine, 50 mM Tris-HCl (pH 8.8), 0.1% SDS).

2.10.3 Purification of MBP/*PAY 4* Protein

A 1301 bp section of the PAY 4 ORF (Ecl136II - BglII) was cloned in frame, down stream of the maltose binding protein (MBP) open reading frame (ORF) in the pMAL-c2 vector (NEB). Expression of the MBP ORF is under the control of the 'tac' promoter, which is inducible by the *lacI* gene product (coded for by the vector) in the presence of IPTG (isopropyl-thio-ß-D-galactoside). The MBP fusion could be purified by passing a crude lysate from an induced culture over an amylose resin column, which specifically binds MBP.

2.10.3.0 Preparation of Crude Lysate

A crude lysate was prepared essentially as described by Ausubel *et al.* (1992). Briefly, 500 ml of LBamp was inoculated with 10 ml of a fresh overnight culture and incubated at 37° C with shaking (220 rpm) until an OD_{600} of 0.5. A 2 ml uninduced control aliquot was removed, pelleted and resuspended in 100 µl of 1X SDS-PAGE sample buffer. IPTG was added to the remaining culture, to a final concentration of 1 mM, and shaking was continued at 37° C for an additional 2 hours. The cells were harvested by centrifugation at 4000 x g_{max} for 10 minutes and resuspended in 20 ml of lysis buffer (10 mM Na P_i (pH 7.2), 30 mM NaCl, 0.25% (w/v) Tween 20, 10 mM ß-mercapto-ethanol, 10 mM EDTA, 10 mM EGTA). The sample then underwent 3 cycles of freezing at -20° C and thawing in cold water, at which point it was sonicated on ice (30 seconds on, 30 seconds off, total sonication time of 8 minutes) at 75% power with a Bronwill (Rochester, NY) Biosonik IV (4mm probe). The lysate was adjusted to 0.5 M NaCl and centrifuged at 9000 x g_{max} for 20 minutes to pellet insoluble material. The supernatant represented the crude extract.

2.10.3.1 Purification of Fusion Protein With an Amylose Resin Column

Amylose resin (NEB) was shipped preswollen in 30% ethanol. A column was prepared by packing the resin in a 30 ml syringe, plugged with siliconized glass wool, to a final bed volume of 10 ml. The column and extract were kept at 4° C for all procedures. The column was then washed with 30 ml of column buffer (CB: 10 mM Na P_i (pH 7.2), 0.5 M NaCl, 10 mM ß-mercapto-ethanol, 1 mM Na azide, 1 mM EGTA) + 0.25% Tween 20. The crude extract was diluted 1:4 with CB + 0.25% (w/v)Tween 20 and run on the column (flow rate ≈ 0.5 ml/min.). The flow through was collected and passed through the column a second time. The column was then washed with 30 ml of CB/Tween 20 followed by 50 ml of CB. The fusion protein was eluted from the column

with CB + 10 mM maltose. 20 one ml fractions were collected and assayed for protein (q.v. 2.10.0).

2.10.4 Purification of Proteins from Gel Fragments by Electroelution

To separate polypeptides prior to elution, 2.5 mg of purified fusion protein, boiled in 1X SDS-PAGE sample buffer, was loaded into a single, 10.5 cm wide well of a preparative SDS-PAGE gel (28 cm x 12 cm x 0.15 cm). The gel was run starting at 100 V on at constant power, for approximately 24 hrs.. Following electrophoresis, the gel was stained, and the desired band was excised with a razor blade and cut into ~2 cm x 0.5 cm pieces. These fragments were transferred to a 10 ml serological pipet, plugged at the tip with siliconized glass wool. The pipet was then mounted, with the tip at the positive electrode, in 1X SDS-PAGE running buffer, in an isoelectric focusing (IEF) apparatus (Bio-Rad, Tube-Cell model 155). The proteins were eluted at 100 V for two days into a dialysis bag sealed around the tip of the pipet. The dialysis bag was then removed, resealed and the protein sample dialyzed three times against 2 L of 50 mM ammonium bicarbonate at room temperature (≥4 hours between changes). The protein was frozen in liquid N_2, lyophilized, and dissolved in an appropriate volume of water.

2.11 Preparation and Characterization of Antisera

2.11.0 Preparation of Anti–PAY 4/MBP Anti-sera

Amylose resin purified PAY 4/MBP fusion protein was further purified by preparative SDS-PAGE, electroeluted and lyophilized as previously described (q.v. Purification of Proteins from Gel Fragments by Electroelution). Approximately 200 μg of protein was isolated for primary injections of two rabbits and two guinea pigs. For each rabbit, approximately 80 μg (in 400 μl) of protein was emulsified with an equal volume of Freund's complete adjuvant and injected subcutaneously in multiple sites,

essentially as described by Ausubel *et al.* (1992). Each guinea pig was injected in the same manner with 20 μg of similarly prepared protein. After five weeks, the animals were boosted with another set of subcutaneous injections (~40 μg of protein for each rabbit and 10μg for each guinea pig). Six weeks after the primary injection, 10 ml and 5 ml of blood (non-heparinized) was collected from the each of the rabbits and guinea pigs, respectively. The blood was allowed to coagulate for 2 hours at room temperature and then 2-12 hours overnight at 4° C. The serum was prepared by centrifuging the clotted blood for 10 minutes at 3000 rpm in an IEC clinical centrifuge. The serum was removed and stored at -20° C until needed.

2.11.1 Western Blot Analysis of Proteins With [^{125}I]-Protein A

Proteins were separated by SDS-PAGE and transferred to nitrocellulose at 200 mA for 16 to 20 hours in a Bio-Rad transblot chamber filled with western transfer buffer (20 mM Tris, 150 mM glycine, 20% methanol). The blot was blocked with 1% bovine hemoglobin in Tris-saline (0.9% NaCl, 0.01 M Tris·HCl, pH 7.5) for 30 minutes with slow rotation on an orbital shaker. The blot was then heat sealed in a plastic pouch (Kapak Corp., Minneapolis, MN) containing the primary antiserum diluted in Tris-saline/hemoglobin. The pouch was rocked for 90 minutes at room temperature. The blot was removed at this time and washed once in Tris-saline, twice in Tris-saline/NP-40 (0.05% w/v), and once again with Tris-saline at room temperature on the orbital shaker. The blot was then resealed in a pouch containing 0.5 μCi [^{125}I] protein A·ml^{-1} and rocked for 30 minutes at room temperature. The blot was then washed as above, allowed to air dry, and exposed to Kodak X-OMAT AR5 film for 16-20 hours with one intensifying screen (DuPont) at -70° C.

3. RESULTS AND DISCUSSION

3.1 Isolation and Analysis of Mutants

When I undertook this project, a large number of mutant strains of *Y. lipolytica* exhibiting impaired growth on minimal oleic acid media but not on minimal acetate media, had already been isolated. Over the course of this work, however, it became apparent that many of the strains had reverted or were inherently unstable, and no longer exhibited the ole$^-$/ace$^+$ phenotype. In addition, all of the remaining strains, except one, failed, upon retesting, to grow on minimal acetate medium. The solution to this particular conundrum rested in the recipe for acetate medium used for screening. The original recipe used had been derived from a paper detailing work done with *S. cerevisiae*, which does not grow well on acetate (Erdmann *et al.*, 1989). As a result, the authors of that particular study included 0.05% yeast extract in the acetate media. E122, the *Y. lipolytica* strain used for mutagenesis, was found to grow well on minimal acetate plates that contained no yeast extract. When the remaining mutant strains were streaked onto this medium, they did not grow. This contrasted with their ability to grow on acetate plates, that contained 0.05% YE, and indicated that they were likely auxotrophic mutants in other biosynthetic pathways. The one mutant, *ole4*, that still appeared normal, i.e. *ole$^-$/ace$^+$*, was eventually found to exhibit a deficiency in peroxisome assembly.

The obvious course of action was to isolate new mutant strains, bearing these observations in mind. After determining the titre of our frozen stocks of mutagenized cells, approximately 75 000 colonies were screened for the ole$^-$/ace$^+$ phenotype. A number of promising candidates were isolated, but these strains either failed to grow on acetate medium or reverted after a few rounds of restreaking to wild type growth on oleic acid medium. To increase the chances of finding mutants, an

aliquot of mutagenized cells was 'enriched' for mutants with nystatin (Snow, 1966). In this procedure, nystatin, a fungicide that kills dividing yeast cells,, is used to increase the proportion of desired mutants in a population of mutagenized cells. This is accomplished by first culturing the cells in non-selective liquid medium, and then switching to selective liquid medium for enough time to curtail the growth of mutant cells. Nystatin is then added after this interval and kills a percentage of the non-mutant cells. This was carried out (q.v. 2.1.0) and the enriched, mutagenized cells were recovered on YEPD plates (≈300 colonies/plate), replica plated onto selective YNO and YNA plates, and scored for growth. From a pool of approximately 80,000 colonies screened, 17 *ole*$^-$*ace*$^+$ mutants were found (see Figure 1 for examples of *ole*$^-$*ace*$^+$ phenotypes).

3.1.0 Screening for PAY Mutants

In order to rapidly separate *ole* $^-$*/ace*$^+$ mutants with defects in peroxisome biogenesis from those bearing less complex phenotypes, i.e. simple defects in oleate metabolism, we took advantage of recent, related observations made in this laboratory, namely that antibodies directed against two forms of the PTS-1 sequence of peroxisomal proteins, anti-AKI and anti-SKL sera, specifically react with peroxisomal proteins of *Y. lipolytica* (Aitchison *et al.*, 1992). Anti-SKL serum was shown by western blotting to react strongly with three proteins in peroxisomal fractions (66, 62 and 33 kDa), and by immunofluorescence to delineate punctate areas within whole cells (Figure 2). Double indirect immunofluorescence studies have also demonstrated that the anti-SKL serum specifically recognizes similar punctate peroxisomal structures in *C. tropicalis* cells. We were able to exploit these observations and rapidly screen various *ole*$^-$ mutants for deficiencies in targeting of peroxisomal proteins. The primary criterion was a lack of characteristic peroxisomal structures following staining with anti-SKL serum.

3.1.0.0 Immunofluorescence

The punctate staining of parental E122 cells with anti-SKL antibodies

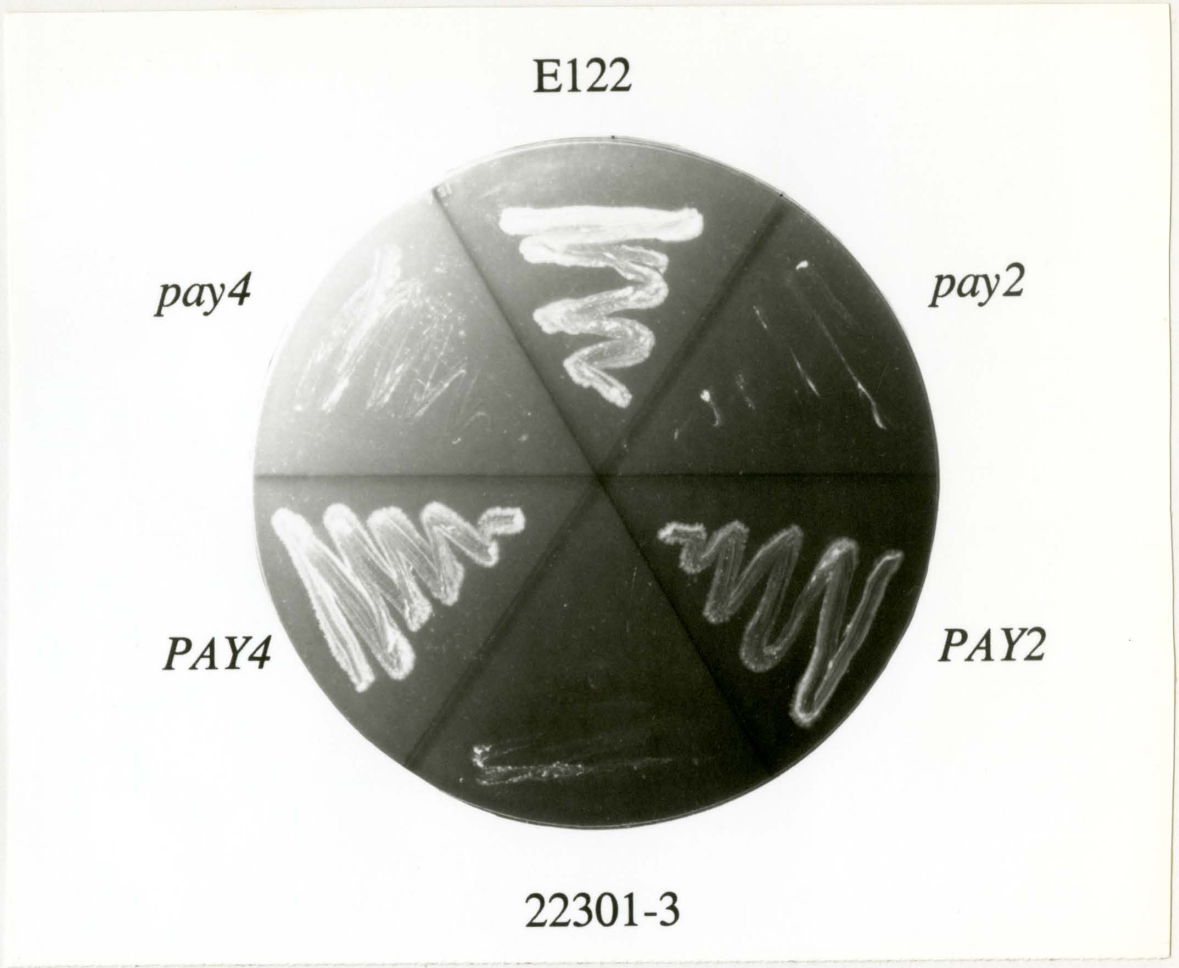


Figure 1: A YNO plate containing the supplements leucine, lysine and uracil and demonstrating oleate growth phenotypes of the wild type positive control (E122), wild type negative control (22301-3, which also requires histidine) and the mutants *pay2* and *pay4* (initially designated *ole2* and *ole4*). Also shown are the complemented mutant strains, designated *PAY2* and *PAY4*. All of the strains shown exhibited essentially similar growth phenotypes when cultured on complete (YEPD) or minimal acetate (YNA, containing required supplements) media (data not shown).

(Figure 2) was typical of the labeling pattern observed for most of the *ole*$^-$ mutants examined (data not shown). These mutants were not characterized further. The remaining *ole*$^-$ mutants lacked definable peroxisomal structures following immunolabelling. Of these mutants, most simply failed to grow at all in the peroxisome inducing medium. The dead cells presented a distinct morphology under phase contrast, lacking any definable nuclear or organellar structures. The lack of growth in induction medium rendered meaningful biochemical analysis of these mutants impossible. Two of the mutants examined, however, grew reasonably well in oleic acid-induction medium and lacked the punctate pattern of peroxisomes after immunolabelling, instead exhibiting diffuse background staining. These two potential *pay* (Peroxisome Assembly in *Y. lipolytica*) mutants, tentatively dubbed *pay2* and *pay4*, were therefore chosen for further biochemical study.

3.1.0.1 Fractionation and Marker Enzyme Analysis

Pay2 and *pay4* were confirmed to be *bona fide pay* mutants by subcellular fractionation and marker enzyme studies. E122, *pay2* and *pay4* were grown in oleic acid-induction medium to allow induction of the ß-oxidation enzymes and catalase. The cells were then homogenized and fractionated into the 20 kgS (cytosol) and 20 kgP (enriched in peroxisomes and mitochondria) fractions. The fractions were then assayed for the activities of catalase and ß-hydroxyacyl-CoA dehydrogenase, two peroxisomal marker enzymes, and the activity of cytochrome *c* oxidase, a mitochondrial marker enzyme.

In wild type E122, the 20 kgP fraction contained a mean of 66% of the dehydrogenase activity and 32% of catalase the activity, reflecting the peroxisomal location of these enzymes (Figure 3). The activities of these enzymes recovered in the 20 kgS were due, at least in part, to leakage from peroxisomes broken during isolation (Aitchison *et al.*, 1991; van den Bosch *et al.*, 1992). Whether catalase leaks

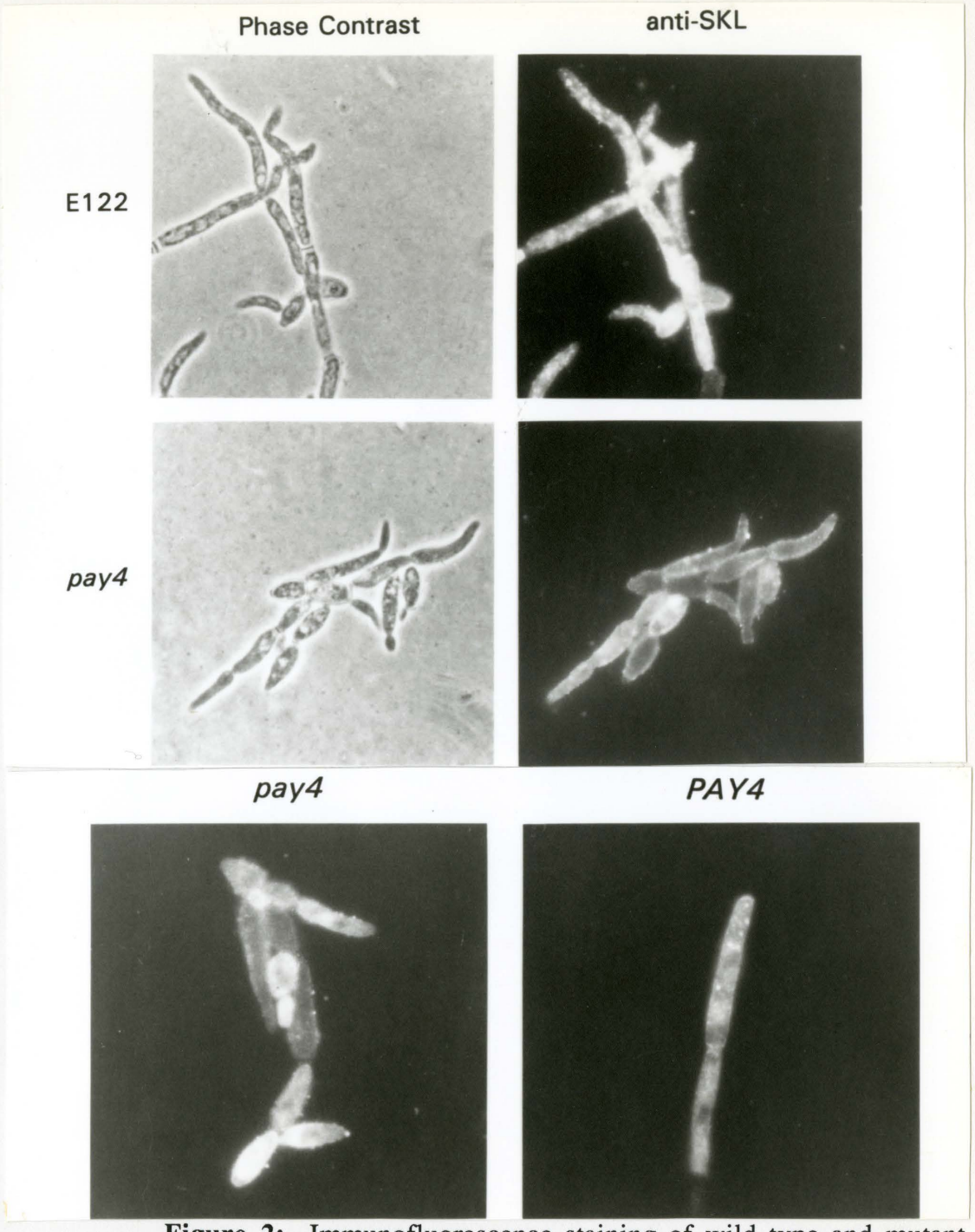


Figure 2: Immunofluorescence staining of wild type and mutant *Y. lipolytica* with anti-SKL antiserum. The top right panel demonstrates the punctate staining pattern typical of peroxisomes in wild type *Y. lipolytica* cells stained with anti-SKL sera. This contrasts with the diffuse background staining observed with the *pay4* mutant (middle right and bottom left). The bottom right panel shows a mutant cell bearing a complementing plasmid and exhibiting an essentially wild type staining pattern.

preferentially into the 20 kgS fraction has not yet been determined. The existence of a cytosolic catalase, as in *S. cerevisiae* (Hartig and Ruis, 1986; Cohen *et al.*, 1988), cannot be excluded. The 20 kgP of the *pay2* and *pay4* mutants consistently contained less than 5% of the dehydrogenase and catalase activities (Figure 3). This demonstrates the mislocalization of catalase and ß-hydroxyacyl-CoA dehydrogenase in these two mutants, and establishes them as *bona fide pay* mutants. The pelletability of the mitochondrial marker enzyme cytochrome *c* oxidase was essentially equal in the wild type E122, *pay2* and *pay4* strains, with approximately half of the activity recovered in the 20 kgP (Figure 3). These results, combined with the phenotypes and immunofluorescence profiles of these mutants, indicate that the genetic lesions present in both strains likely affect peroxisome biogenesis through pathways governing the targeting of peroxisomal proteins and/or the assembly of peroxisomes themselves.

3.2 Construction of the Genomic Library

In order to isolate and sequence the *PAY2* and *PAY4* genes, a library of *Y. lipolytica* genomic DNA was cloned into a vector, pINA445, capable of autonomous replication in *Y. lipolytica*. The library was constructed by ligating 5 to 7 kbp genomic DNA fragments, cut with Sau3A I, to the compatible ends of the unique Bam HI site of pINA445 (Figure 4). Transformation of *E. coli* DH5α cells with a total of 2 µg of ligation products yielded approximately 23 000 transformants, of which an estimated 19 000 (83%) were recombinant. The probability that this library encompasses more than a complete genome of *Y. lipolytica* is 0.999 (Clarke and Carbon, 1976; Xuan *et al.*, 1988). 30 µg of DNA was partially digested with one unit of Sau 3AI and electrophoresed on a 0.4% SeaKem GTG agarose gel. A slice of the gel containing 5-7 kbp fragments was excised and the DNA was electroeluted, yielding approximately 250 ng of DNA. The vector, pINA445, was linearized with Bam HI, which cuts the vector once in its

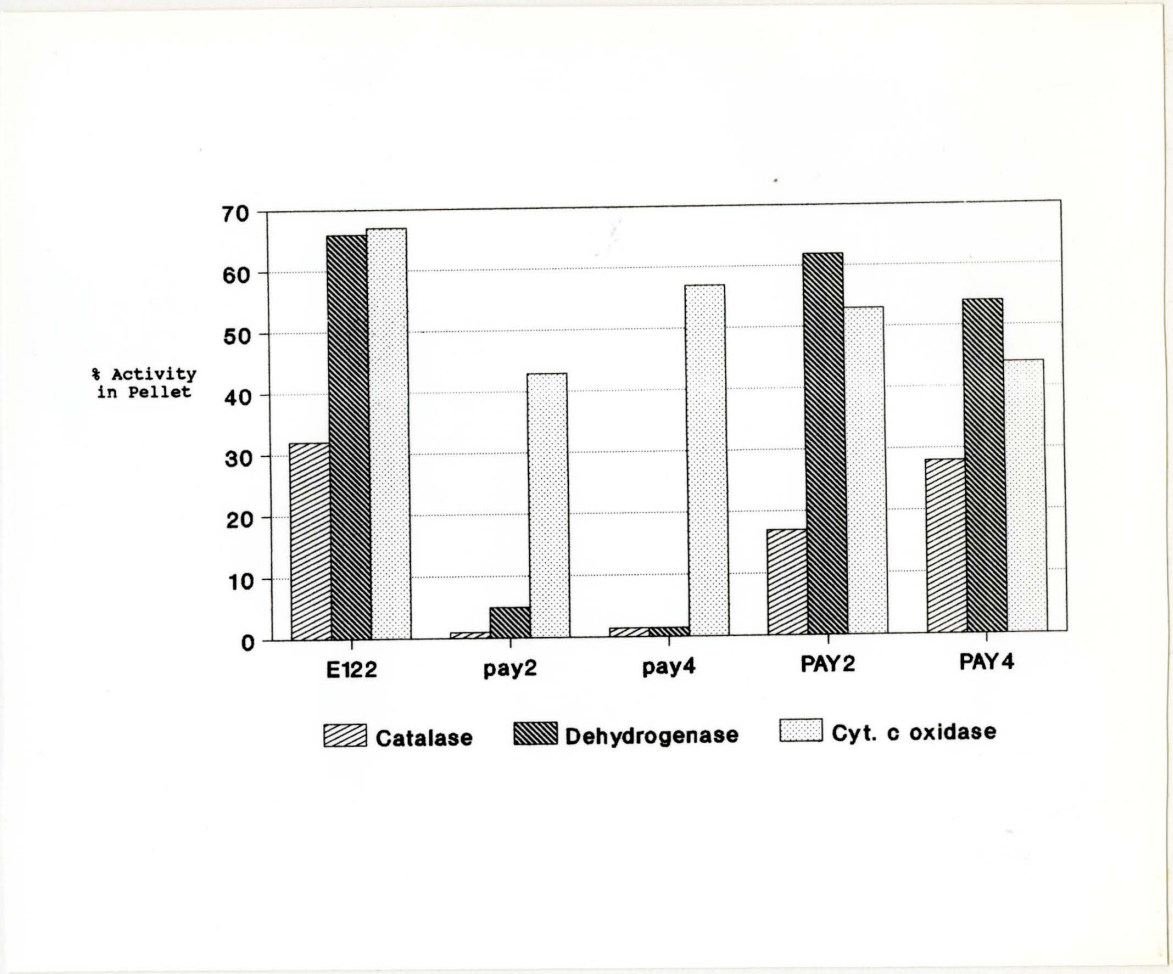


Figure 3: Distribution of peroxisomal and mitochondrial marker enzymes in the 20 kgPs of complemented and uncomplemented *pay* mutants and wild type *Y. lipolytica* . The bars of the chart indicate the average percentage of marker enzyme activity found in the 20 kgP (pellet) fraction of a given cellular homogenate, with the remainder of activity being found in the 20 kgS (cytosolic) fraction. Enzymatic activities in the nuclear fraction, containing primarily nuclei, unbroken cells and cellular debris, were not measured. Strains used are E122, wild type *Y. lipolytica*; *pay2* and *pay4* mutant strains; and *PAY2* and *PAY4*, complemented mutants.

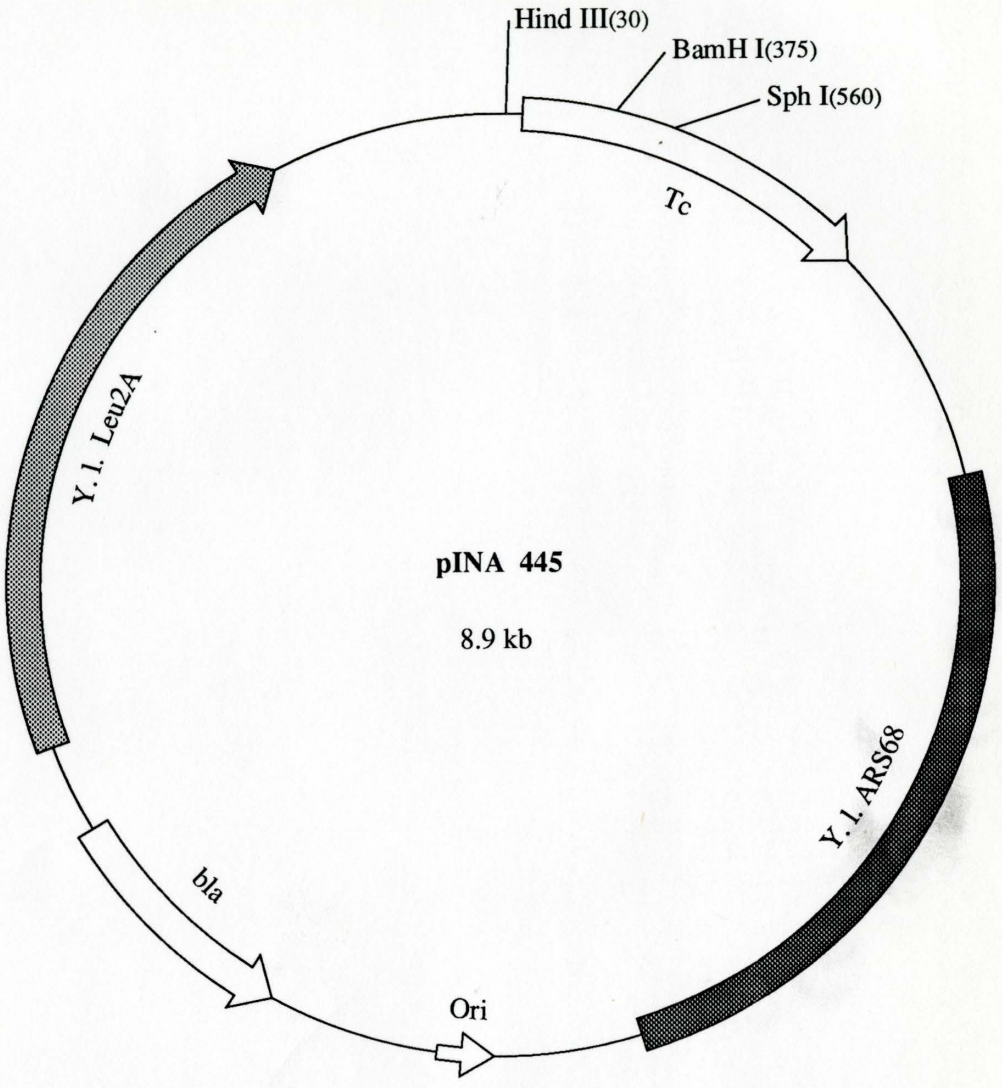


Figure 4: pINA445 *E. coli/Y. lipolytica* shuttle vector. pINA was constructed in the laboratory of Dr. C. Gaillardin by inserting the *Y. lipolytica LEU2a* gene and *Y. lipolytica ARS68* sequences into pBR322. A *Y. lipolytica* genomic library was created by cloning 5-7 kbp Sau3AI fragments into the unique Bam HI site in the pBR322 Tet^r gene. Note the flanking Sph I and Hind III sites as they figured into attempts to determine the minimal region required to complement the *pay4* mutation.

tetracycline resistance gene, and treated with calf intestinal alkaline phosphatase (CIAP). A series of seven ligations, each containing ~20 ng of insert DNA and 300 ng of CIAP treated vector in 10 µl, were diluted five fold and chemically transformed into approximately 15 ml of BRL SE competent cells, at a ratio of ~2 µl ligation mixture: 100 µl SE cells.

Each 100 µl transformation mixture was incubated on ice for 30 minutes, heat shocked for 35 seconds at 37° C, followed by another incubation on ice for at least two minutes. Each transformation was then brought up to 1 ml volume with LB and incubated for 60 minutes at 37° C with rotation. These were combined into pools of 6 ml, of which 100 µl was spread onto LBamp plates, to determine the titre of transformants in that particular pool. These pools of transformants were combined into several banks, pelleted and used to inoculate 50 ml LBamp cultures. Each bank was grown to mid-log phase, pelleted, resuspended in 5 ml of freezing medium, frozen in one ml aliquots in a dry ice-ethanol bath and stored at -70° C.

To determine the percentage of recombinants present in a given pool, all transformant colonies were also screened for growth on LBtet plates, and approximately one fifth of these were screened by plasmid isolation followed by diagnostic digestion with Pst I. All tet resistant clones screened with diagnostic DNA digestion were non-recombinants, as expected. Of the few tet sensitive clones screened in this manner that were not recombinants, most exhibited some sort of gross rearrangement of the plasmid DNA isolated. Of a total of 265 colonies screened, 220 were tet sensitive, yielding an estimated efficiency of 83%. Extrapolated to all of the pools of transformants, an estimated 23 100 individual transformants were generated, 19 200 of which were recombinants.

This represents a total length of ~115 mbp of cloned genomic DNA. Using the method of Clarke and Carbon (1976), the probability that this library

encompasses greater than one full *Y. lipolytica* genome is 0.999. This is based on a total genomic size of 15 mbp, as estimated by Xuan *et al.* (1988), who created a library containing ~65 mbp of *Y. lipolytica* genomic DNA in an integrative shuttle vector.

A large scale preparation of the library was made following the addition of representative numbers of transformants from each bank to 1 L of LBamp. This yielded 1.1 mg of DNA and contained ~74% recombinants (102:34 amp^r/tet^s:amp^r/tet^r). This apparent decrease in the proportion of recombinants, combined with the observation that slower growing (smaller) colonies were more likely to be recombinants (31:5 amp^r/tet^s:amp^r/tet^r), suggests that some of the recombinant plasmids were under slight negative selection, relative to wild type pINA445, during bacterial growth. Thus, some segments of the *Y. lipolytica* genome may be under-represented in the library used for this study. This is not unusual for DNA libraries cultured in *E. coli* , however (Ausubel *et al.*, 1992).

3.3 Transformation of *Y. lipolytica*

In order to complement mutants, a reliable *Y. lipolytica* transformation protocol was required. Initial attempts were made using a lithium acetate protocol (Gaillardin *et al.*, 1985). While a variety of permutations of this protocol were attempted, none ever yielded better than 100 colonies·μg^{-1} of DNA. Electroporation was also experimented with and found to be somewhat more successful. A basic electroporation experiment, performed essentially as described by Becker and Guarante (1991) except that cells were plated on standard selective plates (YNBD) as opposed to plates containing 1 M sorbitol, produced ~1000 colonies·μg^{-1} after the magnitude and duration of the voltage pulse was optimized. Pretreating the cells with 10 mM DTT (as per Meilhoc *et al.*, 1990) and plating the transformants on media containing 1 M sorbitol (YNBD/Sorb), increased the transformation efficiency ~5 fold.

3.3.0 High Efficiency Transformation

After observing the strong growth of E122 on minimal acetate compared to growth on minimal glucose media, an experiment was performed preculturing the yeast in YEPA, rich acetate liquid medium, instead of YEPD, and plating the transformants on selective acetate medium containing 1 M sorbitol (YNA/Sorb) instead of YNBD/Sorb. This dramatically boosted transformation efficiency to between $5 \cdot 10^4$ and $2.5 \cdot 10^5$ transformants/μg. This enabled rapid complementation of mutants , yielding between 5 and 25 complementing clones per transformation (~500 ng of library) for all mutants tested thus far. These efficiencies were approximately two orders of magnitude greater than those obtained using lithium acetate protocols and related ARS containing vectors (Fournier *et al.*, 1991).

3.4 Cloning and Characterization of *PAY 4*

3.4.0 Complementation of *ole 4*

As intimated above, once a high efficiency transformation system was established, complementing mutants became a relatively straightforward task. A single electroporation of *pay* 4 with 500 ng of the library effectively produced a lawn of transformants following 2-3 days incubation at 30° C on selective YNA/Sorb. This was then replica plated onto minimal oleate plates (YNO) and incubated at 30° C for 5-7 days. Colonies bearing complementing plasmids were clearly visible at this time as there was essentially no background growth. Similarly, *pay4* transformants bearing only pINA445 produced no colonies after being replicaplated onto selective YNO plates. In total, two plates of *pay* 4 transformed with the library yielded 42 individual clones.

3.4.1 Isolation and Mapping of Plasmids pO1, pO2 and pO3

Eight of these colonies were randomly chosen, subcultured into selective liquid medium and harvested for plasmid rescue into *E. coli* (q.v. 2.5.1). Three distinct

plasmids, designated pO1 (Figure 5), pO2 (Figure 6) and pO3 (Figure 7), were isolated from the eight clones and were shown by restriction mapping to share a common region of ~4.5 kbp.

3.4.1.0 Attempts to isolate smaller complementing fragments

In an attempt to reduce the common complementing region, and thus the amount of DNA to be sequenced, a number of convenient restriction fragments from the three plasmids were subcloned back into pINA445. The first two constructs made contained the 3.0 and 3.2 kbp Hind III fragments from pO1, cloned into the Hind III site of pINA445 (designated pH3.0 and pH3.2, see figure 8). Neither of these fragments complemented *pay* 4 and this suggested that the central Hind III site might be in or close to the sequence of the complementing gene, since the two fragments together comprised essentially the entire pO1 genomic insert, and contained, between them, the 4.5 kbp common region in its entirety. The next construct made, pS3.5, was created by subcloning the 3.5 kbp Sph I fragment common to pO2 and pO3 into the Sph I site of pINA445 (Figure 9). This fragment included ~1 kbp of the common region upstream of the Hind III site and all of the downstream region. This fragment failed to complement as well. The final construct, pP3.0, was created by cloning the 3 kbp Pvu II fragment into Bam HI cut, Klenow blunted pINA445 (Figure 9). This clone, which also failed to complement the *pay4* mutation, essentially encompassed all of the common region contained in the 3.0 kbp Hind III fragment and extended an additional ~0.5 kbp downstream of the common Hind III site. These results suggested that the sequence required to complement the *pay4* strain had to encompass a region beginning upstream of the Sph I site and extending downstream of the Pvu II site as delineated in figure 10.

3.4.2 Sequencing of *PAY 4* and Analysis of Potential Coding Regions

These observations directed initial sequencing efforts and assisted in delineating the complementing open reading frame (ORF). An ORF was defined for

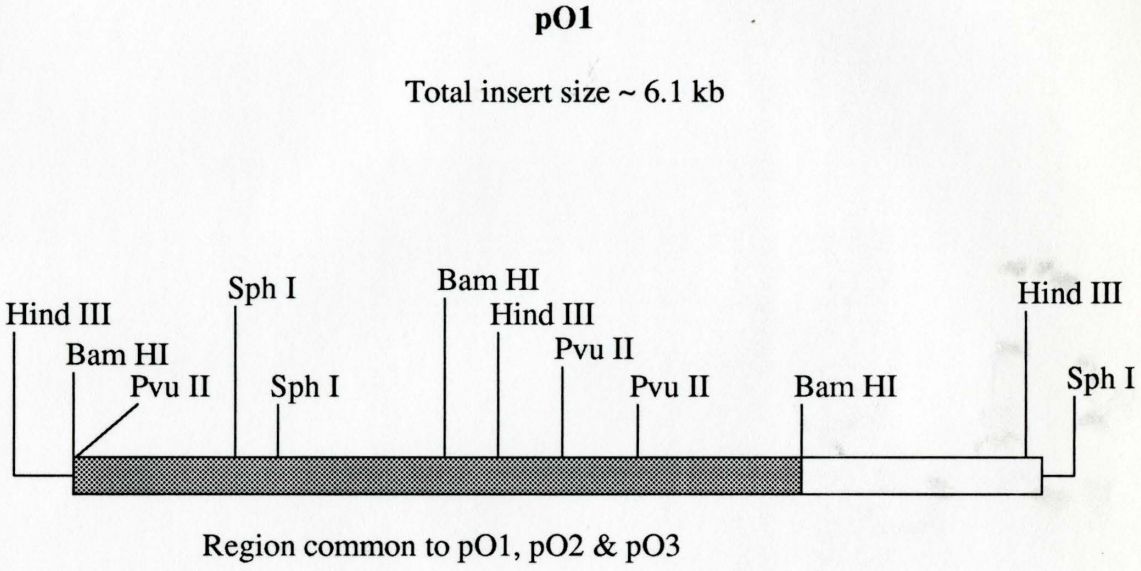


Figure 5: Restriction map of the pO1 plasmid. pO1 was the first plasmid rescued from complemented *pay4* strains. The shaded region depicts the portion of the map (~4.5 kbp) that was determined to overlap with the inserts of other distinct complementing plasmids. As indicated, the insert was estimated to span approximately 6.1 kbp.

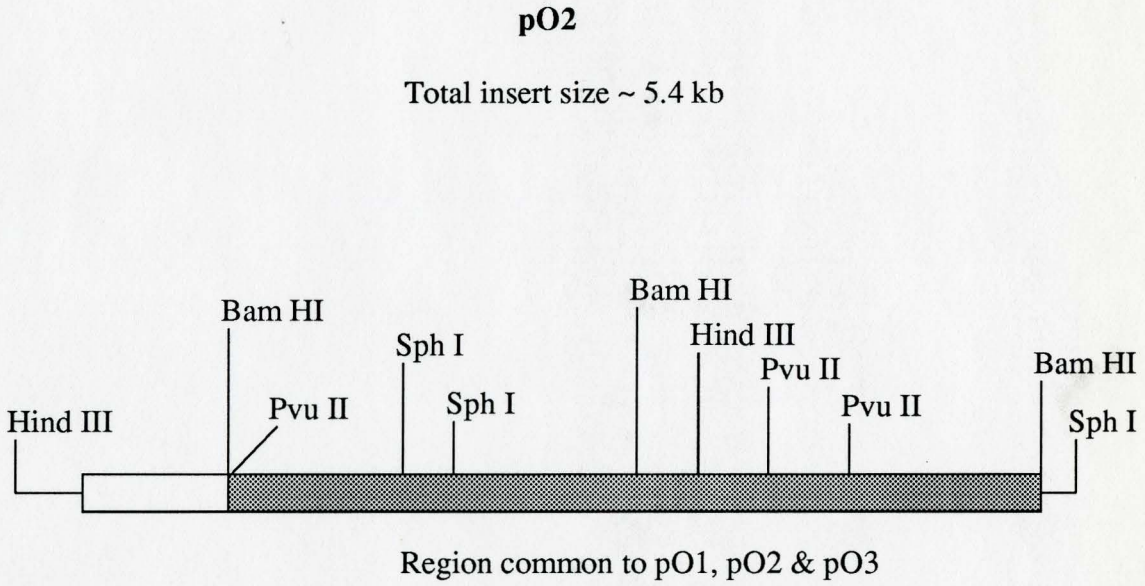


Figure 6: Restriction map of the pO2 plasmid. pO2 was the second plasmid rescued from complemented *pay4* strains. The shaded region depicts the portion of the map (~4.5 kbp) that was determined to overlap with the inserts of other distinct complementing plasmids. As indicated, the insert was estimated to span approximately 5.4 kbp.

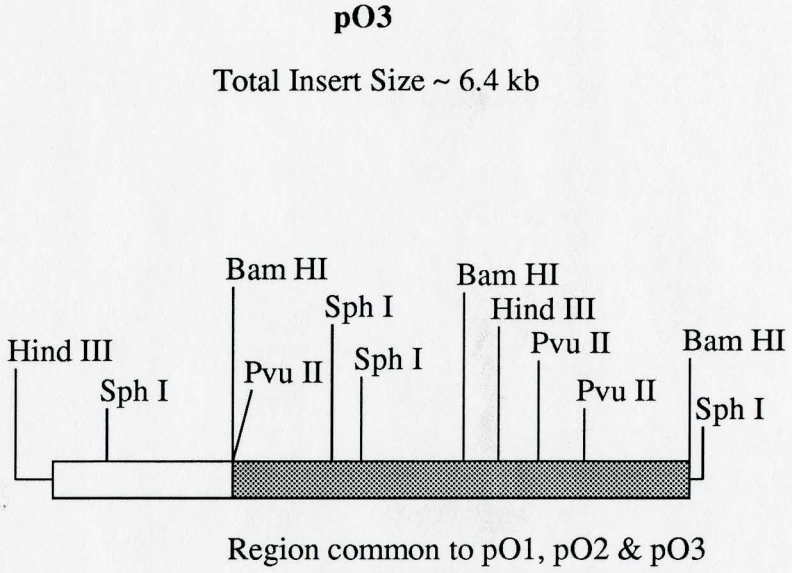


Figure 7: Restriction map of the pO3 plasmid. pO3 was the third plasmid rescued from complemented *pay4* strains. The shaded region depicts the portion of the map (~4.5 kbp) that was determined to overlap with the inserts of other distinct complementing plasmids. As indicated, the insert was estimated to span approximately 6.4 kbp.

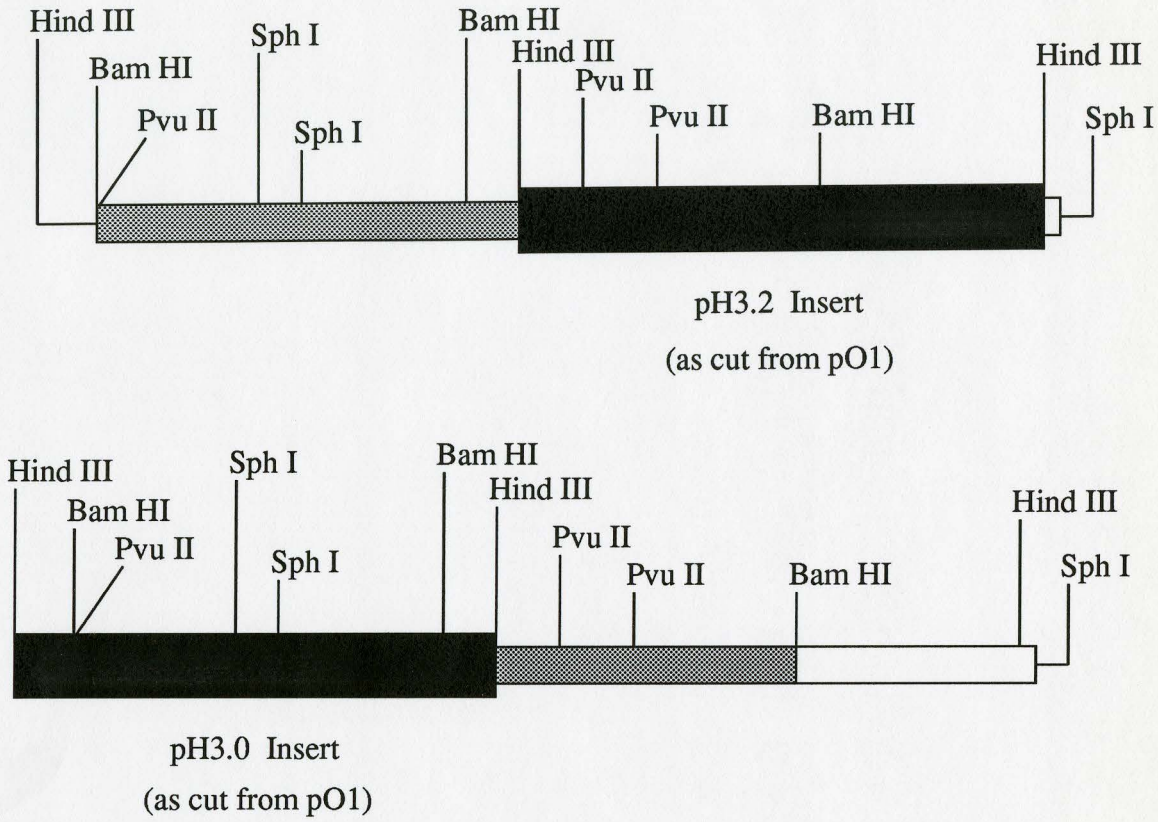


Figure 8: pO1 fragments tested for complementation of *pay4* mutation. The portions of the pO1 restriction map shaded black were subcloned back into pINA445 in an effort to minimize the amount of sequence needed to complement the *pay4* mutation. Since neither fragment complemented the mutation, this suggested that (i) they shared the required sequence between them, and (ii) the Hind III bisecting them was in or close to the putative complementing ORF.

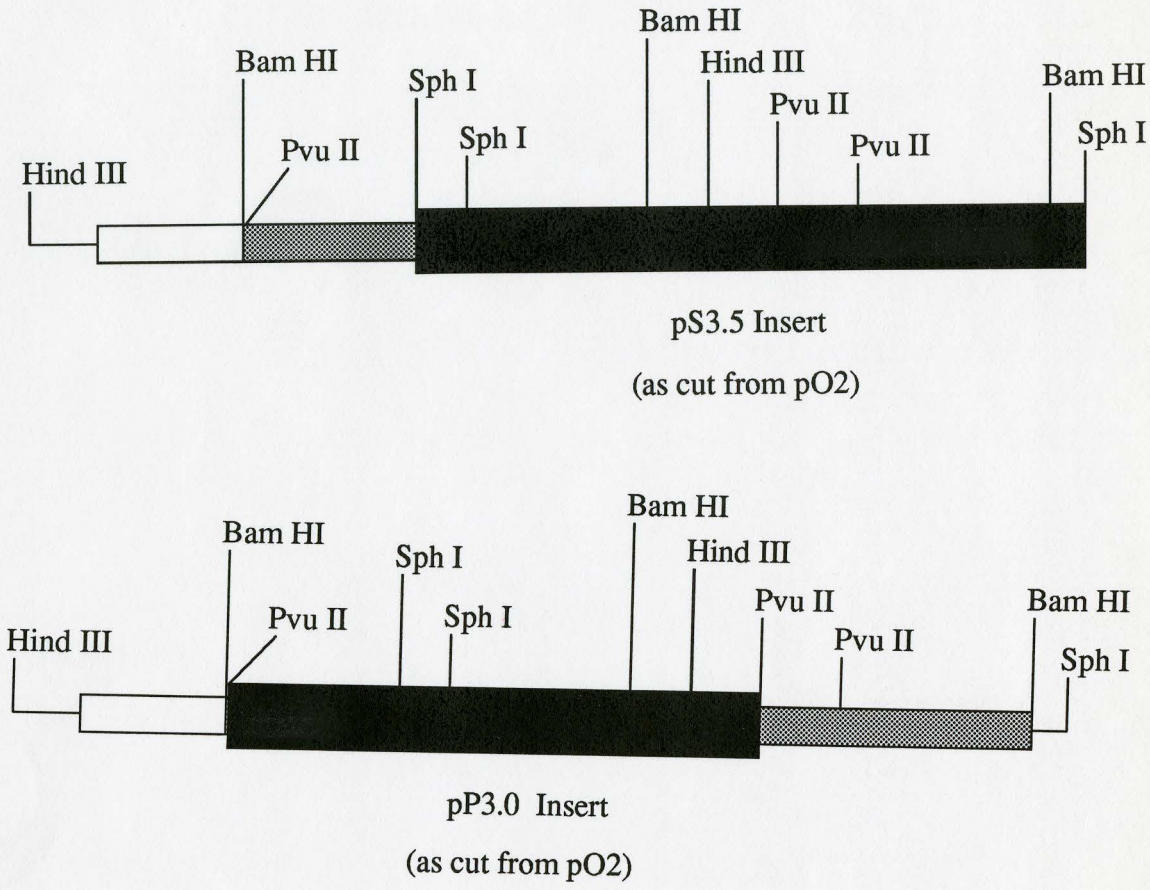


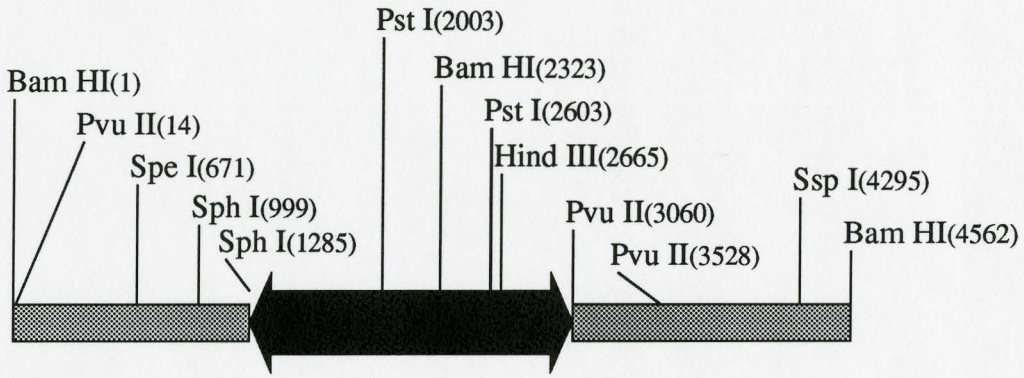
Figure 9: pO2 fragments tested for complementation of *pay4* mutation. The portions of the pO2 restriction map shaded black were subcloned back into pINA445 in an effort to minimize the amount of sequence needed to complement the *pay4* mutation. Neither fragment complemented the mutation.

these purposes as a stretch of in frame codons beginning with an ATG and ending with one of the three 'stop' codons. The total sequencing effort is summarized in figure 11 and the entire sequence is presented in appendix A.

A single ORF (designated ORF 1) was found (Figure 11) that encompassed the essential region established by pH3.0, pH3.2, pS3.5 and pP3.0 (Figure 10), it spans 3075 bp of the 4567 bp region common to the three complementing plasmids, and is flanked by canonical transcription initiation and termination elements (see below). Two other large ORFs were found on the opposite strand, one of 1563 bp and one of 654 bp, but both of these were contained entirely within pP3.0, which indicated that they were not sufficient to complement the *pay4* mutation. In addition, neither of these ORFs were flanked by obvious transcription initiation or termination elements.

3.4.2.0 Potential transcription initiation elements flanking ORF1

The 'TATA' box is one of the central motifs governing transcriptional initiation in yeasts and other eukaryotes (Heslot and Gaillardin, 1992). The position of the TATA box relative to the ATG initiation codon varies from -30 to -300 bp in different *S. cerevisiae* genes (Xuan *et al.*, 1990; Heslot and Gaillardin, 1992). For comparison, three of the four *Y. lipolytica* genes that have been sequenced and published to date have clear TATA boxes ranging from 125 to 65 nucleotides upstream of their respective ATG initiation codons (Davidow *et al.*, 1985; Davidow *et al.*, 1987; Xuan *et al.*, 1990; Köttig *et al.*, 1991). In contrast, putative TATA elements were found to occur at positions -404 and -461 relative to the initiation ATG of ORF 1. It seems doubtful, therefore, that these TATA elements are directly involved in the transcriptional control of ORF 1 expression. It is possible, however, that ORF 1 is controlled by a so called 'TATA-less' promoter (Heslot and Gaillardin, 1992), i.e. one that contains a cryptic TATA element. This is not unprecedented in *Y. lipolytica.* The *Y. lipolytica LYS5*



The shaded area denotes a region of sequence essential to pay4 complementation, as established by the plasmids pH3.0, pH3.2, pS3.5 and pP3.0.

Figure 10: Portion of the common region absolutely required for *pay4* complementation. The region shaded black represents the overlapping region between pS3.5 and pP3.0. Neither of these plasmids complemented the *pay4* mutation and this indicated that the sequence shared between them was essential for this activity. The arrow heads are used to indicate that, while this region was required for complementation, it was not sufficient. Therefore, the minimal coding region must also include sequences flanking the shaded area.

PAY4 ORF Sequencing Summary

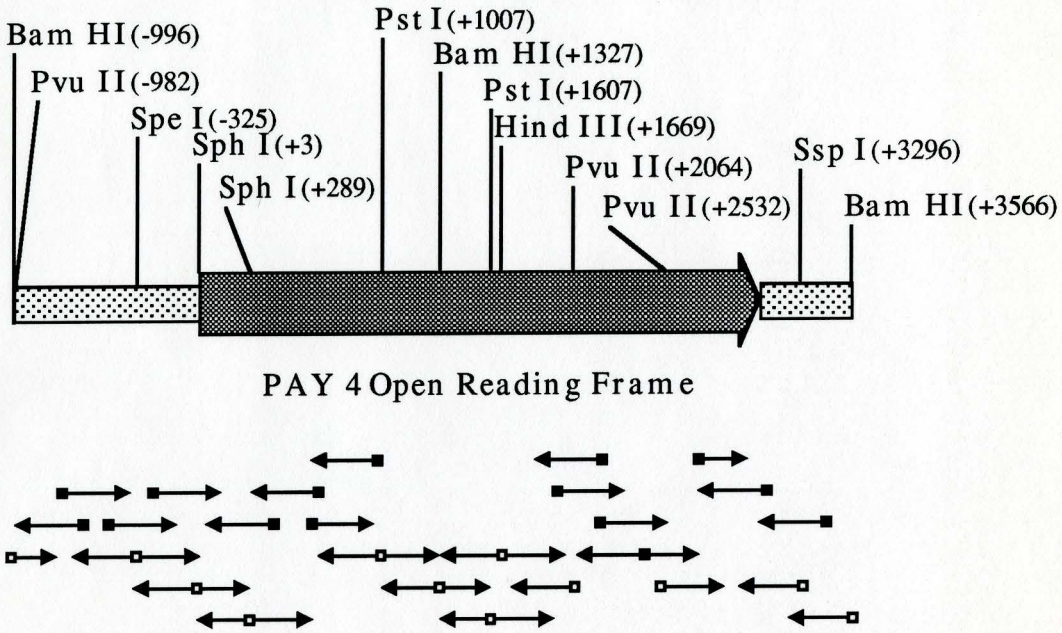


Figure 11: Sequencing summary of the 4.5 kbp common region and *PAY4* ORF. The darker shaded region denotes the extent of the *PAY4* open reading frame, with the arrow delineating the direction of transcription. All sequencing reactions were carried out using pGEM subcloned fragments of inserts from pO1 pO2 or pO3 as templates. The small arrows beneath the sequence schematic represent individual sequencing reactions, with the direction depicting which strand was sequenced. Solid boxes indicate a reaction carried out with a sequence specific primer, while open boxes are indicative of reactions carried out using the M13 forward or reverse primers.

gene, for example, does not display an obvious TATA box (Xuan *et al.*, 1990). Transcription of this gene was shown by primer extension to initiate in a 13 bp CA rich region (position -50 relative to the ATG) immediately downstream of an 11 bp CT rich. This CT rich region was preceded by another CT rich region (11 bp) approximately 70 bp upstream (position -130). A strikingly similar situation exists upstream of the ORF 1 ATG. A 14 bp CA rich region begins at position -20 and is immediately preceded by a 12 bp CT rich stretch. Another CT rich region (11 bp) occurs at position -70 (See appendix A).

The sequence surrounding the ORF 1 initiator was also compared to the consensus sequence for a translation initiator in yeast. 75% of AUG start codons in *S. cerevisiae* are preceded by an A at the -3 position and 50% are followed by a C at the +5 position (reviewed by Cigan and Donahue, 1987). The -3 A is proposed to be important for translation initiation on yeast ribosomes. The ORF 1 ATG conforms to this 'rule' (AGC ATG CC). Two other ATG triplets can be found within 500 bases upstream of the putative ORF 1 ATG. Neither of these has this consensus sequence. In addition, both of these ATG triplets are followed closely (< 120 bp) by in frame stop codons.

3.4.2.1 Transcription termination elements flanking ORF 1

Several sequence motifs have been identified with transcription termination and polyadenylation of yeast mRNA transcripts. TTTTTTATA (Henikoff and Cohen, 1984) and TAG...TGAT...TTT (Sutton and Broach, 1985, Zaret and Sherman, 1982) sequences on the sense strand of a transcribed gene represent two of the most common elements associated with yeast transcriptional termination and polyadenylation. In *Y. lipolytica* , the TAG...TGAT...TTT motif appears to be most common, as it is present at the 3' end of the four *Y. lipolytica* genes whose sequences have been published (Davidow *et al.*, 1987; Xuan *et al.*, 1990; Davidow *et al.*, 1985; Köttig *et al.*, 1991). A TAG...TGAT...TTT termination motif was found beginning at nucleotide +3152, 73 nt

downstream of the ORF 1 stop codon. A TTTTTTATA termination motif was also found, beginning at nucleotide +3264. The region 3' of ORF1, therefore, appears to represent a potentially strong mediator of transcriptional termination.

This sequence analysis, combined with restriction mapping data and the attempts to find a smaller complementing fragment to complement *pay4*, strongly suggest that ORF 1 encodes a gene that complements the *pay4* mutation. By extension this gene has tentatively been designated *PAY4*.

3.4.3 Computer Analysis of the *PAY4* Protein

The PC Gene application was used for all computer analysis of the *PAY4* nucleotide and putative peptide sequences. Several pertinent observations were made using this program. Two consensus ATP/GTP binding sites were identified in the *Pay4* gene using the PROSITE program and the RAOARGOS program highlighted two hydrophobic regions of sufficient length to theoretically span a lipid bilayer.

3.4.3.0 The ATP/GTP binding domain

A significant portion of ATP/GTP binding proteins have been shown by comparative studies using sequence and crystallographic data to share several reasonably conserved motifs. The sequence motif (A or G)-X-X-X-X-G-K-(S or T), known as the 'P-loop' or 'A' motif, represents the most conserved of these regions and is generally believed to form a flexible phosphate binding loop between a ß-strand and an α-helix (Walker *et al.*, 1982; Saraste *et al.*, 1990). This sequence was found twice in Pay4p at positions 477-484 and 760-767.

Interestingly, this arrangement of two duplicated ATP binding domains separated by 200-300 amino acids is characteristic of a family of presumed Mg^{2+} dependent ATPases that include the peroxisomal assembly protein Pas1p cloned from *S. cerevisiae* (Erdmann *et al.*, 1991). Other members of this family include a *S. cerevisiae* protein, Sec18p, and its mammalian homologue NSF (*N*-ethylmaleimide sensitive factor,

cloned from Chinese hamster) both involved in vesicular fusion along the secretory pathway (Eakle *et al.*, 1988; Wilson *et al.*, 1989), *S. cerevisiae* CDC48p, a cell cycle protein (Fröhlich *et al.*, 1991) and two of its vertebrate homologues VCP, porcine "vasolin" containing protein (Koller and Brown, 1987), and p97 from *Xenopus laevis* (Peters *et al.*, 1990). All of these proteins share considerable sequence similarity over ~400 amino acids beginning with the first P-loop and extending ~200 amino acids past the second such motif.

p97 has been shown to exhibit strong ATPase activity, comparable to other highly active ATPases such as yeast elongation factor 3 and mammalian kinesin (Peters *et al.*, 1992; Uritani and Miyazaki, 1988; Kunznetsov and Gelfland, 1986). This activity was demonstrated *in vitro* to be dependent upon a homo-hexameric form of p97 containing two trimeric rings stacked on top of one another (Peters *et al.*, 1992). Furthermore, ATPase activity has been shown to be exhibited by NSF, which also apparently forms oligomers (Peters *et al.*, 1992; Block *et al.*, 1988).

Searches of the CDPROT20 database with the FASTSCAN program in PC/Gene to find proteins in the database with short regions of homology to Pay4p, demonstrated that the ATP binding modules (encompassing the P-loop and flanking region) of the proteins discussed above were also highly similar to the putative P-loops and flanking regions of Pay4p. More detailed comparisons with the proteins were carried out with the program PALIGN using the 'Structure-Genetic-Matrix' and open gap and unit gap costs of 5 and 3, respectively.

In addition, the primary sequence of a peroxisome assembly protein from the yeast *P. pastoris* (Pas5p, S. Subramani personal communication) was found to exhibit a very high level of homology to Pay4p with 59% identity and 15.6% similarity over their ATPase modules and 46% identity/15.3% similarity over their entire length. Much less striking conservation was found between Pay4p and *S. cerevisiae* Pas1p with 28.9%

identity/14% similarity over the conserved ATPase modules and 21.5% identity/14.9% similarity over their entire length . This relative lack of conservation between the primary sequences of Pay4p and Pas1p is emphasized by the similarity between Pay4p and porcine p97 (VCP) and its *S. cerevisiae* homologue CDC48p, with 38.1% identity/18.5% similarity and 34.9% identity/18.1% similarity, respectively over their conserved ATPase modules.

Interestingly, the full length sequences of *P. pastoris* Pas5p and *S. cerevisiae* Pas1p were also found to be tenuously related (24.1% identity/15.1% similarity). The scant similarity that is observed between full length Pay4p/Pas5p and Pas1p seems, in fact, to derive entirely from the (by definition) conserved ATPase modules. While Pay4p and Pas5p appear to be closely related at the primary sequence level, clearly the same cannot be said for their relatedness with Pas1p. More study will be necessary, therefore, to determine whether these proteins collectively represent a common factor in the peroxisomal assembly machinery of yeast. This question could be addressed through heterologous expression studies. Ideally, one would attempt to rescue each mutant strain with the corresponding genes from the other species, i.e. ascertaining whether the *PAY4* and/or *PAS5* genes would complement the *pas1 S. cerevisiae* strain and *vice versa*.

3.4.3.1 Potential membrane spanning helices

Three programs in PCGene, RAOARGOS (Rao and Argos, 1986; Argos *et al.*, 1982), HELIXMEM (Eisenberg *et al.*, 1984) and SOAP (Kyte and Dolittle, 1982; Klein *et al.*, 1985) can be used to predict regions in polypeptides that may span a lipid bilayer. Using RAOARGOS to scan the putative Pay4p sequence, (Parameters: 16 amino acids minimum length, buried helix parameter ≥ 1.13) two putative membrane spanning regions were identified (amino acids 489-508 and 767-782). Each of these sequences contains, centrally, one charged and one polar amino acid (Q/E and K/N, respectively)

within them, casting some doubt on the veracity of the prediction (Reithmeier and Deber, 1991). In addition, neither of these sequences surpassed the detection threshold for putative membrane spanning helices in the SOAP or HELIXMEM programs in PCGene. It should be noted that all of the homologous proteins discussed above have been shown to be soluble proteins localized to the cytosol and/or nucleus, with the exception of Pas5p, the subcellular distribution of which is yet to be investigated. The method of Rao and Argos has been shown to erroneously identify membrane spanning regions in some soluble proteins (Rao and Argos, 1986). Clearly, subcellular fractionation combined with western blotting using anti-Pay4p antibodies will be required to determine the localization of Pay4p unequivocally.

3.4.4 Knock-out of the *PAY 4* Gene

An attempt was made to knock out the *PAY* 4 gene. This was done by cloning the *Y. lipolytica LEU*2 gene into a fragment of the *PAY* 4 ORF and transforming the fragment back into wild type *Y. lipolytica.* The internal 2.1 kbp Sst I fragment of the *PAY* 4 ORF (from nucleotide +71 to nucleotide +2133) was subcloned into the Sst I site of pGEM 3Z f(+). This construct was then digested with Stu I and Bgl II and gel purified. This removed a fragment of the *PAY4* ORF from +1009 to +1817. The *Y. lipolytica LEU*2 gene was isolated on an Eco47 III/Bgl II fragment from pINA445. This fragment was gel purified and ligated into the Stu I/Bgl II digested *PAY* 4 /pGEM 3Z construct. The resulting hybrid was digested with Sph I and Sca I, each of which cleaved once within the *PAY4* ORF, at positions +289 and +2113 respectively. This step was chosen, rather than cleaving the construct with Sst I as the *Y. lipolytica LEU*2 gene contains an Sst I site. The resulting linear fragment contained the *Y. lipolytica LEU*2 flanked 5' by nucleotides +289 to +1009 of the *PAY4* ORF and 3' by nucleotides +1817 to +2113. This fragment was gel purified and electroporated into E122 and 22301-3 haploid

strains. These strains were then selected for leucine prototrophy and clones were screened for their abilities to grow on oleic acid.

Preliminary results indicate that some E122 *Leu*$^+$ transformants isolated were *ole*$^-$, suggesting that the gene we have designated *PAY4* is, in fact the gene mutated in the *pay4* strain. No *Leu*$^+$ 22301-3 transformants isolated to date have exhibited an *ole*$^-$ phenotype, the reasons for this are unknown at present. Southern blotting experiments carried out by Gary Eitzen have determined that one of the *Leu*$^+$/*ole*$^-$ E122 transformants bears a single copy of the *PAY4* gene that has been disrupted by the knock-out construct. A demonstration that this clone lacks recognizable peroxisomes and contains mislocalized peroxisomal enzymes will prove that the gene encoded by ORF 1, dubbed *PAY4*, is indeed the gene mutated in the *pay4* strain.

3.4.5 Construction of *Pay 4* /Maltose Binding Protein Fusion

A portion of the *PAY4* ORF was cloned in frame with the *E. coli* maltose binding protein gene in the vector pMAL-c2 (NEB). The fusion protein was constructed in order to raise an anti-serum against Pay4p to facilitate expression and localization studies. It should be noted that the fragment chosen coded for a region on the C-terminal side of the conserved, putative ATPase modules. Thus, the resulting anti-sera were expected to be specific for Pay4p and not to cross-react appreciably with any *Y. lipolytica* homologues of CDC48p or Sec18p.

pMAL-c2 was digested with Xmn I and Bam HI and gel purified. A 1299 bp Ecl136 II (+69)/Bgl II (+1370) fragment of the *PAY* 4 ORF was digested from pO2 and also gel purified. The vector and insert were ligated together, transformed into DH5α *E. coli* , and 12 colonies picked for small scale plasmid isolation and restriction endonuclease screening. Of these clones, 8 recombinants were found and used for further study. Small scale (5 ml) inductions (1 mM IPTG) were carried out on the eight recombinants, and one non-recombinant control, to determine the amount and size of the

fusion protein they expressed. Based on the predicted size of the polypeptide encoded by the ORF 1 fragment (~43 kDa) and the size of MBP (~42 kDa), the in frame fusion was predicted to encode a protein of approximately 85 kDa. Five of the eight recombinants produced a protein of this size, strongly induced with 1 mM IPTG, and the remaining three produced a much smaller protein (~50 kDa). The ~50 kDa species is consistent with the predicted fusion protein that would be produced if a single base deletion occurred during the ligation of the Xmn I/Ecl136 II blunt ends. This type of ligation event would have fused an open reading frame of 228 bp (76 a.a.) onto the end of the MBP gene. The control non-recombinant produced an induced polypeptide of ~42 kDa, as expected.

The 85 kDa fusion protein was purified away from contaminating *E. coli* proteins by passing the induced cell lysate over an amylose resin column (q.v. 2.10.3), which binds the MBP portion of the fusion protein. The fusion protein was subsequently eluted by washing the column with binding buffer containing free maltose. The proteins in the peak eluate fractions were then further separated by preparative SDS-PAGE followed by electroelution of the band containing the 85 kDa fusion protein. The eluate from the electroelution was concentrated with a lyophilizer and injected as described (q.v. 2.11.0) into two guinea pigs and two rabbits to generate polyclonal sera against the fusion protein.

3.4.6 Characterization of *PAY 4* Expression

Studies on the expression of Pas1p in *S. cerevisiae* demonstrated that it is oleate inducible but very difficult to detect by western blotting. This problem was solved by isolating protein from cells bearing a multi-copy vector carrying a functional *PAS1* gene (Erdmann *et al.*, 1991). A similar tack was therefore chosen for the preliminary studies on expression of Pay4p, i.e. protein was isolated from E122 cells carrying the pO1 complementing plasmid. It should be noted that, while most *S. cerevisiae* multi-copy

vectors are present in the cell at an average copy number of 10/cell, *ARS68* based vectors, like pINA445 in this case, are present on average in lower quantities (~3/cell) in *Y. lipolytica* (C. Gaillardin, personal communication). This preliminary study (data not shown) demonstrated that, the anti-Pay4p/MBP serum generated in one of the rabbits, specifically recognized a protein of ~110-120 kDa (the size estimation was inexact as only prestained molecular weight markers were loaded on the gel). Figure 12 depicts a western blot done later by John Glover using that α-Pay4p/MBP serum that clearly demonstrates induction of Pay4p following growth in oleate medium. In addition, the empirical molecular weight can be seen to match the predicted molecular weight. This signal was stronger with lysates made from cells bearing the pO1 vector versus lysates made with cells bearing pINA445 or no vector at all. The induction with oleate, increased expression in cells bearing the complementing plasmid, and close correlation between predicted and observed molecular weight strongly suggest that the anti-serum specifically recognizes the *PAY4* gene product.

As implied above, Pay4p was also visualized by western blotting in extracts derived from induced and uninduced wild type E122 cells not carrying the pO1 plasmid. This contrasts with the expression of Pas1p in *S. cerevisiae,* which is not detectable unless the *PAS1* gene is present in the cells on a multicopy vector (Erdmann *et al.*, 1991). Interestingly, this difference in expression correlates with the respective abilities of *Y. lipolytica* and *S. cerevisiae* to grow on oleic acid. *Y. lipolytica* also exhibits significant basal levels of expression of Pay4p when grown on media containing glucose or acetate. When carrying *PAS1* on a multi-copy vector, *S. cerevisiae* also demonstrates a measurable basal level of expression in 0.3% glucose. This expression was shown to be repressed by growth of *S. cerevisiae* cells in high glucose (5%) media (Erdmann *et al.*, 1991). Glucose repression of *PAY4* expression has not yet been tested.

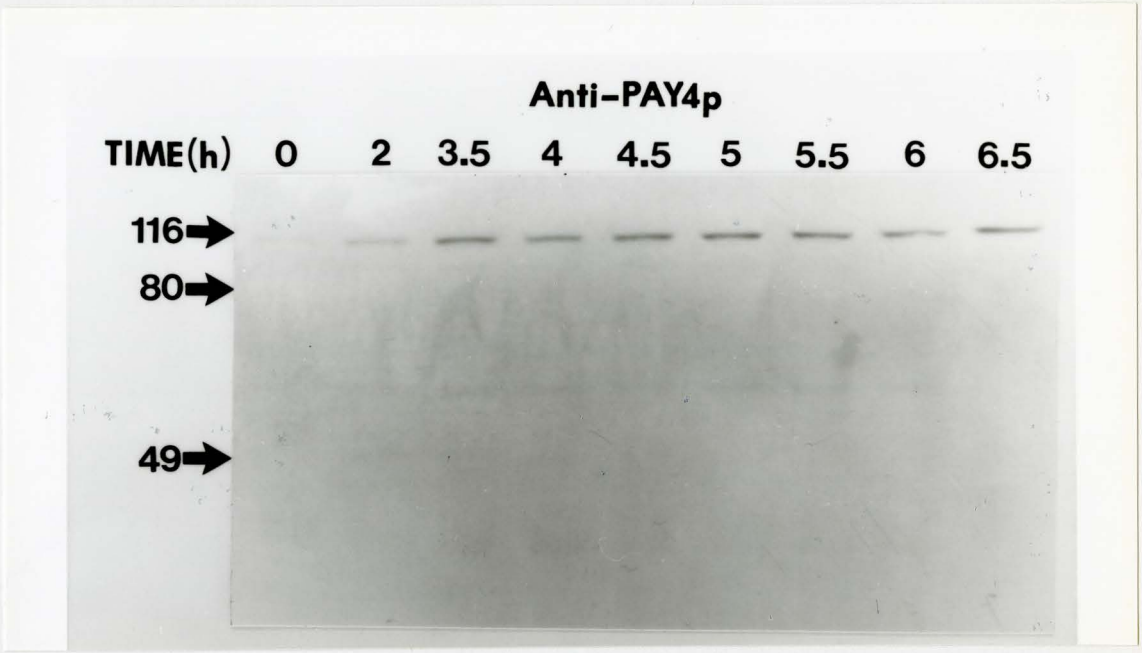


Figure 12: Western blot using rabbit αPay4p/MBP serum to assess induction of Pay4p expression following transfer of 3 g (wet weight) of *E122* cells from glucose containing medium (500 ml of YEPD) to oleate containing medium (500 ml YPBO: 0.3% yeast extract, 0.3% bacto-peptone, 0.5% K_2HPO_4, 0.5% KH_2PO_4, 1.0% Brij 35 (SCC), 1.0% oleate, Kamiryo *et al.*, 1982). Cells were removed at the time points noted in the figure and processed using the glass bead disruption (2.10.1). 300 µg of each time point was loaded onto a 7.5 % SDS-PAGE gel for separation.

3.5 Future Directions

The work presented above leads in a number of obvious directions. Clearly, there are some 'loose ends' to be tied up regarding Pay4p expression. In addition, the sub-cellular location of Pay4p is a critical piece of missing information. Preliminary results using immunofluorescence suggest that Pay4p is not compartmentalized in the cell following growth in induction medium. This question can be answered best using western analysis, and such studies are presently under way in our laboratory. From this point, work will hopefully lead to the eventual elucidation of the biochemical function of Pay4p.

NOTE ADDED IN PROOF: The sequence of a third *S. cerevisiae PAS* gene (*PAS2*) was published as this thesis was completed (Wiebel and Kunau, 1992). The findings of this paper do not significantly affect the interpretation of the results presented here. Pas2p is apparently a novel ubiquitin conjugating enzyme.

4. REFERENCES

Aitchison, J.D., Murray, W.W. and Rachubinski, R.A. (1991) The carboxy-terminal tripeptide Ala-Lys-Ile is essential for targeting *Candida tropicalis* trifunctional enzyme to yeast peroxisomes. *J. Biol. Chem.* **266**, 23197-23203.

Aitchison, J.D., Nuttley, W.M., Szilard, R.K., Brade, A.M., Glover, J.R. and Rachubinski, R.A. (1992) Peroxisome biogenesis in yeast. *Molec. Microbiol.* **6**, (in press).

Aitchison, J.D., Szilard, R.K., Nuttley, W.M. and Rachubinski, R.A. (1992) Antibodies directed against a yeast carboxyl-terminal peroxisomal targeting signal specifically recognize peroxisomal proteins from various yeasts. *Yeast* **8**, 721-734.

Argos, P., Rao, M.J.K. and Hargrave, P.A. (1982) Structural prediction of membrane proteins. *Eur. J. Biochem.* **128**, 565-575.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J. and Struhl, K. (eds.) (1992) *Current Protocols in Molecular Biology*. Greene Publishing and Wiley-Interscience, New York.

Azarnoff, D., Tucker, D. and Barr, G. (1965) Studies with ethyl chlorophenoxyisobutyrate (clofibrate). *Metab. Clin. Exp.* **14**, 959-965.

Baudhuin, P., Beaufay, H., Rahman-Li, Y., Sellinger, O.Z., Wattiaux, R., Jacques, P. and de Duve, C. (1964) Tissue fractionation studies. Intracellular distribution of monoamine oxidase, aspartate aminotransferase, alanine aminotransferase, D-amino acid oxidase and catalase in rat-liver tissue. *Biochem. J.* **92**, 179-184.

Becker, D.M. and Guarente, L. (1991) High-efficiency transformation of yeast by electroporation. *Meth. Enzymol.* **194**, 182-187.

Beckerich, J., Fournier, P., Gaillardin C., Heslot, H., Rochet, M. and Trenton, B. (1985) Yeasts. In Ball, C. (ed.) *Genetics and Breeding of Industrial Microorganisms*, pp. 115-158, CRC Press Inc., Boca Raton, FL, USA.

Birnboim, H.C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. *Meth. Enzymol.* **100**, 243-255.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.

Bodnar, A.G. and Rachubinski, R.A. (1991) Characterization of the integral membrane polypeptides of rat liver peroxisomes isolated from untreated and clofibrate-treated rats. *Biochem. Cell Biol.* **69**, 499-508.

Borst, P. (1989) Peroxisome biogenesis revisited. *Biochim. Biophys. Acta* **1008**, 1-13.

Bradford, M.M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-284.

Brul, S., Wiemer, E.A.C., Westervald, A., Strijland, A., Wanders, R.J, Schram, A.W., Heymans, H., Schutgens R.B.H., van den Bosch, H.S.A. and Tager, J.M. (1988) Kinetics of the assembly of peroxisomes after fusion of complementary cell lines from patients with the cerebro-hepato-renal (Zellweger) syndrome and related disorders. *Biochem. Biophys. Res. Commun.* **152**, 1083-1089.

Cigan, A.M. and Donahue, T.F. (1987) Sequence and structural features associated with translational initiator regions in yeast--a review. *Gene* **59**, 1-18.

Clarke, J. and Switzer, R. (1977) *Experimental Biochemistry* (2nd Ed.), pp. 256 W. H. Freeman and Co., San Francisco.

Clarke, L. and Carbon, J. (1976) A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9**, 91-99.

Cohen, G., Rapatz, W. and Ruis, H. (1988) Sequence of the *Saccharomyces cerevisiae CTA1* gene and amino acid sequence of catalase A derived from it. *Eur. J. Biochem.* **176**, 159-163.

Cooper , T.G. and Beevers, H. (1969) ß-oxidation in glyoxysomes from the castor bean endosperm. *J. Biol. Chem.* **244**, 3514-3520.

Cooperstein, S.J. and Lazarow, A. (1951) A microspectrophotometric method for the determination of cytochrome oxidase. *J. Biol. Chem.* **189**, 665-670.

Cregg, J.M., van der Klei, I.J., Sulter, G.J., Veenhuis, M. and Harder, W. (1990) Peroxisome-deficient mutants of *Hansenula polymorpha. Yeast* **6**, 87-97.

Datta, N.S., Wilson, G.N. and Hajra, A.K. (1984) Deficiency of enzymes catalyzing the biosynthesis of glycero-ether lipids in Zellweger syndrome. *N. Engl. J. Med.* **311**, 1080-1083.

Davidow, L.S., Apostolakos, D., O'Donnell, M.M., Proctor, A.R., Ogrydziak, D.M., Wing, R.A., Stasko, I., De Zeeuw, J.R. (1985), Integrative transformation of the yeast *Yarrowia lipolytica. Curr. Genet.* **10**, 39-48.

Davidow, L.S., O'Donnell, M.M.,Kaczmarek, F.S., Pereira, D.A., De Zeeuw, J.R. and Franke, A.E. (1987), Cloning and sequencing of the alkaline extracellular protease gene of *Yarrowia lipolytica. J. Bacteriol.* **169**, 4621-4629.

Davis, W.L., Goodman, D.B., Crawford, L.A. Cooper, O.J. and Matthews, J.L. (1990) Hibernation activates glyoxylate cycle and gluconeogenesis in black bear adipose tissue. *Biochim. Biophys. Acta* **1051**, 276-278.

Davis, W.L., Matthews, J.L. and Goodman, D.B. (1989) Glyoxylate cycle in the rat liver: effect of vitamin D_3 treatment. *FASEB J.* **3**, 1651-1655.

de Duve, C. (1983) Microbodies in the living cell. *Sci. Am.* **248**, 74-84.

de Duve, C. and Baudhuin, P. (1966), Peroxisomes (microbodies and related particles). *Physiol. Rev.* **46**, 323-357.

Didion, T. and Roggenkamp, R. (1992) Targeting signal of the peroxisomal catalase in the methylotrophic yeast *Hansenula polymorpha. FEBS Lett.* **303**, 113-116.

Distel, B., Gould, S.J. ,Voorn-Brouwer, T. ,van der Berg, M., Tabak, H.F. and Subramani, S.(1992) The carboxyl-terminal tripeptide serine-lysine-leucine of firefly luciferase is necessary but not sufficient for peroxisomal targeting in yeast. *New Biol.* **4**, 157-165.

Distel, B., van der Klei, I., Veenhuis, M. and Tabak, H.F. (1988) Alcohol oxidase expressed under nonmethylotrophic conditions is imported, assembled, and enzymatically active in peroxisomes of *Hansenula polymorpha. J. Cell. Biol.* **107**, 1669-1675.

Eakle, K.A., Bernstein, M. and Emr, S.D. (1988) Characterization of a component of the yeast secretion machinery: Identification of the *SEC18* gene product. *Mol. Cell. Biol.* **8**, 4098-4109.

Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci.* USA **81**, 140-144.

Erdmann, R., Veenhuis, M., Mertens, D. and Kunau, W.-H. (1989) Isolation of peroxisome-deficient mutants of *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci.* USA **86**, 5419-5423.

Erdmann, R., Wiebel, F.F., Flessau, A. ,Rytka, J., Beyer, A., Fröhlich, K.-U. and Kunau, W.-H. (1991) *PAS 1*, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. *Cell* **64**, 499-510.

Fournier, Ph., Guyaneux, L., Chasles, M. and Gaillardin, C. (1991) Scarcity of *ARS* sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica. Yeast* **7**, 25-36.

Fröhlich, K.-U., Fries, H.W., Rudiger, M., Erdmann, R., Botstein, D. and Mecke, D. (1991) A yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. *J. Cell. Biol.* **114**, 443-453.

Fujiki, Y., Rachubinski, R.A. and Lazarow, P.B. (1984) Synthesis of a major integral membrane polypeptide of rat liver peroxisomes on free polysomes. *Proc. Natl. Acad. Sci. USA* **81**, 7127-7131.

Fujiki, Y., Rachubinski, R.A., Zentella-Dehesa, A. and Lazarow, P.B. (1986) Induction, identification, and cell-free translation of mRNAs coding for peroxisomal proteins in *Candida tropicalis*. *J. Biol. Chem.* **261**, 15787-15793.

Gaillardin, C., Ribet, A.M. and Heslot, H. (1985) Integrative transformation of the yeast *Yarrowia lipolytica*. *Curr. Genet.* **10**, 40-58.

Gärtner, J., Moser, H.W. and Valle, D. (1992) Mutations in the 70K peroxisomal membrane protein gene in Zellweger syndrome. *Nature Genetics* **1**, 16-23.

Gleeson, M.A. and Sudbery, P.E. (1988a) The methylotrophic yeasts. *Yeast* **4**, 1-15.

Gleeson, M.A. and Sudbery, P.E. (1988b) Genetic analysis in the methylotrophic yeast *Hansenula polymorpha*. *Yeast* **4**, 293-303.

Goldfischer, S., Moore, C.L., Johnson, A.B., Spiro, A.J., Valsamis, M.P., Wisniewski, H.K. Ritch, R.H., Norton, W.T., Rapin, I. and Gartner, L.M. (1973) Peroxisomal and mitochondrial defects in the cerebro-hepato-renal syndrome. *Science* **182**, 62-64.

Goodman, D.B., Davis, W.L. and Jones, R.G. (1980) Glyoxylate cycle in toad bladder: possible stimulation by aldosterone. *Proc. Natl. Acad. Sci. USA* **77**, 1521-1525.

Gorgas, K. (1984) Peroxisomes in sebaceous glands. V. Complex peroxisomes in the mouse preputial gland: serial sectioning and three-dimensional reconstruction studies. *Anat. Embryol.* **169** 261-270.

Gorgas, K. (1985) Serial section analysis of mouse hepatic peroxisomes. *Anat. Embryol.* **172**, 21-32.

Gould, S.J., Keller, G.-A. and Subramani S. (1987) Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J. Cell. Biol.* **105**, 2923-2931.

Gould, S.J., Keller, G-.A. and Subramani, S. (1988) Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. *J. Cell Biol.* **107**, 897-905.

Gould, S.J., Keller, G-.A., Hosken, M., Wilkinson, J. and Subramani, S. (1989) A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **108**, 1657-1664.

Gould, S.J., Krisans, S., Keller, G.-A. and Subramani, S. (1990a) Antibodies directed against the peroxisomal targeting signal of firefly luciferase recognize multiple mammalian proteins. *J. Cell. Biol.* **110**, 27-34.

Gould, S.J., Keller, G.-A., Schneider, M., Howell S.H., Garrard, L.J., Goodman, J.M., Distel, B., Tabak, H. and Subramani, S. (1990b) Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J.* **9**, 85-90.

Gould, S.J., McCollum, D., Spong, A.P., Heyman, J.A. and Subramani, S. (1992) Development of the yeast *Pichia pastoris* as a model organism for a genetic and molecular analysis of peroxisome assembly. *Yeast* **8**, 613-628.

Hanson, R.F., Szczopanik-van Leeuwen, P., Williams, G.C., Grabowski, G. and Sharp, H.L. (1979) Defects of bile acid synthesis in Zellweger's syndrome. *Science* **203**, 1107-1108.

Hartig, A. and Ruis, H. (1986) Nucleotide sequence of the *Saccharomyces cerevisiae CTT1* gene and deduced amino-acid sequence of yeast catalase T. *Eur. J. Biochem.* **160**, 487-490.

Hellman, L., Zumoff, B., Kara, E., Rubin, I.L. and Rosenfeld, R.S. (1963) Reduction of cholesterol and lipids in man by *p*-chlorophenoxyisobutyrate. *Ann. Int. Med.* **59**, 477-494.

Henikoff, S. and Cohen, E.H. (1984) Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 1515-1520.

Heslot, H. and Gaillardin, C. (1992) *Molecular Biology and Genetic Engineering of Yeasts*, pp. 118-125, CRC Press Inc., Boca Raton FL, USA.

Hess, R., Stäubli, W., Reiss, W. (1965) Nature of the hepato-megalic effect produced by ethyl- chlorophenoxy-isobutyrate in the rat. *Nature* **208**, 856-858.

Heymans, H.S., Schutgens, R.B., Tan, R., van den Bosch, H. and Borst P. (1983) Severe plasmalogen deficiency in tissues of infants without peroxisomes (Zellweger syndrome). *Nature* **306**, 69-70

Hodge, V.J., Gould, S.J., Subramani, S., Moser, H.W. and Krisans, S. (1991) Normal cholesterol synthesis in human cells requires functional peroxisomes. *Biochem. Biophys. Res. Commun.* **181**, 537-541.

Höhfeld, J., Veenhuis, M. and Kunau, W.-H. (1991) *PAS3*, a *Saccharomyces cerevisiae* gene encoding a peroxisomal integral membrane protein essential for peroxisome biogenesis. *J. Cell. Biol.* **114**, 1167-1178.

Holt, C.E., Garlick, N. and Cornel, E. (1990) Lipofection of cDNAs in the embryonic central nervous system. *Neuron* **4**, 203-214.

Hyman L.E., Seiler, S.H., Whorisky, J. and Moore, C.L. (1991) Point mutations upstream of the yeast ADH2 poly(A) site significantly reduce the efficiency of 3'-end formation. *Mol. Cell. Biol.* **11**, 2004-2012.

Kamiryo, T., Abe M., Okazaki, S. and Shimamoto, N. (1982) Absence of DNA in peroxisomes of *Candida tropicalis*. *J. Bacteriol.* **152**, 269-274.

Kawamoto, S., Nozaki, C., Tanaka, A. and Fukui, S. (1978) Fatty acid ß-oxidation system in microbodies of *n*-alkane grown *Candida tropicalis*. *Eur. J. Biochem.* **83**, 609-613.

Keller, G.-A. ,Gould, S.J. ,Deluca, M. and Subramani, S. (1989) Firefly luciferase is targeted to peroxisomes in mammalian cells. *Proc. Natl. Acad. Sci. USA.* **84**, 3264-3268.

Keller, G.-A. Krisans, S., Gould, S.J., Sommer, J.M., Wang, C.C., Schiebls, W., Kunau, W.-H., Brody, S. and Subramani, S (1991) Evolutionary conservation of a microbody targeting signal that targets proteins to peroxisomes, glyoxysomes, and glycosomes. *J. Cell Biol.* **114**, 893-904.

Klein, P., Kanehisa, M. and DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* **815**, 468-476.

Koller, K.J. and Brownstein, M.J. (1987) Use of a cDNA clone to identify a supposed precursor protein containing vasolin. *Nature* **325**, 542-545.

Kornberg, H.L. and Beevers, H. (1957) A mechanism of conversion of fat to carbohydrates in castor beans. *Nature* **180**, 35-36.

Köttig, H., Rottner, G., Beck, K.-F., Schweizer, M. and Schweizer, E. (1991) The pentafunctional *FAS1* genes of *Saccharomyces cerevisiae* and *Yarrowia lipolytica* are co-linear and considerably longer than previously estimated. *Mol. Gen. Genet.* **226**, 310-314.

Kurischko, C., Fournier, P., Chasles, M., Weber, H. and Gaillardin, C. (1992) Cloning of the mating type gene *MATA* of the yeast *Yarrowia lipolytica*. *Mol. Gen. Genet.* **232**, 423-426.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lazarow, P.B. (1978) Rat liver peroxisomes catalyze the ß-oxidation of fatty acids. *J. Biol. Chem.* **253**, 1522-1528.

Lazarow, P.B. and Fujiki, Y. (1985) Biogenesis of peroxisomes. *Ann. Rev. Cell Biol.*, **1**, 489-530.

Lazarow, P.B. and Moser, H.W., (1989) Disorders of peroxisome biogenesis. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle (eds.),*The Metabolic Basis of Inherited Disease* (6th ed.), pp. 1479-1509, McGraw-Hill, New York.

Lazarow, P.B. and de Duve, C. (1976) A fatty acyl CoA oxidizing system in rat liver peroxisomes: enhancement by clofibrate, a hypolipidemic drug. *Proc. Natl Acad. Sci. USA.* **73**, 2043-2046.

Meilhoc, E., Masson, J.M. and Teissie, J. (1990) High efficiency transformation of yeast cells by electric field pulses. *Bio/Technology* **8**, 223-227.

Middelkoop, E., Strijland A.and Tager, J.M. (1991) Does aminotriazole inhibit import of catalase into peroxisomes by retarding unfolding? *FEBS Lett.* **279**, 79-82.

Moser, A.E., Singh, I., Brown, F.R., Solish, G.I., Keeley, R.I., Benke, P.J. and Moser, H.W. (1984) The cerebrohepatorenal (Zellweger) syndrome. Increased levels and impaired degradation of very-long-chain fatty acids and their use in prenatal diagnosis. *N. Engl. J. Med.* **310**, 1141-1146.

Murikami, K. and Mori, M. (1990) Purified presequence binding factor (PbF) forms an import competent complex with a purified mitochondrial precursor protein. *EMBO J.* **9**, 3201-3208.

Novikoff, A.B. and Shin, W.-Y. (1964) The endoplasmic reticulum in the Golgi zone and its relation to microbodies, Golgi apparatus and autophagic vesicles in rat liver cells. *J. Micros.* **3**, 187-206.

Novikoff, P.M., Novikoff, A.B., Quintana, N. and Davis, C. (1964) Studies on microperoxisomes. III. Observations on human and rat hepatocytes. *J. Histochem. Cytochem.*, **21**, 540-558.

Nuttley, W.M., Aitchison, J.D. and Rachubinski, R.A. (1988) cDNA cloning and primary structure determination of the peroxisomal trifunctional enzyme hydratase-dehydrogenase-epimerase from the yeast *Candida tropicalis* pK233. *Gene* **69**, 171-180.

Oliver, M.F. (1963) Further observations on the effects of 'Atromid' and of ethyl chlorophenoxyisobutyrate on serum lipid levels. *J. Atheroschlerosis Res.* **3**, 427-444.

Opperdoes, F.R. (1987) Compartmentation of carbohydrate metabolism in trypanosomes. *Ann. Rev. Microbiol.* **41**, 127-151.

Opperdoes, F.R., Borst, P., Bakker, S and Keene, W. (1977) Localization of glycerol-3-phosphate oxidase in the mitochondrion and particulate NAD^+-linked glycerol-3-phosphate dehydrogenase in the microbodies of the bloodstream form of *Trypanosoma brucei*. *EMBO J.* **4**, 3811-3817.

Osumi, M., Fukuzumi, F. ,Teranishi, Y., Tanaka, A. and Fukui, F. (1975) Development of microbodies in *Candida tropicalis* during incubation in a *n*-alkane containing medium. *Arch. Microbiol.* **103**, 1-11.

Osumi, T. and Fujiki, Y. (1990) Topogenesis of peroxisomal proteins. *Bioessays* **12**, 217-222.

Osumi, T. and Hashimoto, T. (1979) Peroxisomal ß-oxidation system of rat liver. Copurification of enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase. *Biochem. Biophys. Res. Commun.* **89**, 580-584.

Pederson, J.I. and Gustafsson, J. (1980) Converison of 3α, 7α, 12α-trihydroxy-5ß-cholestanoic acid into cholic acid by rat liver peroxisomes. *FEBS Lett.* **121**, 345-348.

Peters, J.-M., Harris, J.R., Lustig, A., Müller, S., Engel, A., Volker, S. and Franke, W.W. (1992) Ubiquitous soluble Mg^{2+}-ATPase complex: A structural study. *J. Mol. Biol.* **223**, 557-571.

Peters, J.-M., Walsh, M.J. and Franke, W.W. (1990) An abundant and ubiquitous homo-oligomeric ring shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. *EMBO J.* **9**, 1757-1767.

Pringle, J.R., Adams, A.E., Drubin, D.G., and Haarer, B.K. (1991) Immunofluorescence methods for yeast. *Methods Enzymol.* **194**, 565-602.

Prodromou, C., Wright, I.P., Evans, I.H. and Bevan, E.A. (1991) Cloning of the *HIS3* gene of *Yarrowia lipolytica. Antonie Van Leeuwwenhoek* **60**, 95-99.

Rachubinski, R.A., Fujiki. Y., Mortensen R.M. and Lazarow. P.B. (1984) Acyl-CoA oxidase and hydratase-dehydrogenase, two enzymes of the peroxisomal ß-oxidation system, are synthesized on free polysomes of clofibrate-treated rat liver. *J. Cell. Biol.* **99**, 2241-2246.

Rao, M.J.K. and Argos, P. (1986) A conformational preference parameter to predict helices in integral membrane proteins. *Biochim. Biophy. Acta* **869**, 197-214.

Reithmeier, R. and Deber, C., (1992). In Yeagle, P. (ed.) *The Structure of Biological Membranes*. CRC Press Inc., Boca Raton, FL.

Rouiller, C. and Bernhard, W. (1956) "Microbodies" and the problem of mitochondrial regeneration in liver cells. *J. Biophys. Biochem. Cytol.* Suppl. **2**, 355-359.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.* USA **74**, 5463-5467.

Santos, M.J., Imanaka, T., Shio, H., Small, G.M. and Lazarow, P.B. (1988a) Peroxisomal membrane ghosts in Zellweger syndrome-aberrant organelle assembly. *Science* **239**, 1536-1538.

Santos, M.J., Imanaka, T., Shio, H. and Lazarow, P.B. (1988b) Peroxisomal integral membrane proteins in control and Zellweger fibroblasts. *J. Biol. Chem.* **263**, 10502-10509.

Saraste, M., Sibbald, P.R. and Wittinghofer, A. (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430-434.

Schrakamp, G., Roosenboom, C.F., Schutgens, R.B., Wanders, R.J., Heymans, H.S.A., Tager, J.M. and van den Bosch, H. (1985) Alkyl dihydroxyacetone phosphate synthase in human fibroblasts and its deficiency in Zellweger syndrome. *J. Lipid Res.* **26**, 867-873.

Schutgens, R.B. Romeyn, G.J., Wanders, R.J., van den Bosch, H., Schrakamp, G. and Heymans, H.S.A. (1984) Deficiency of acyl-CoA: dihydroxyacetone phosphate acyl transferase in patients with Zellweger (cerebro-hepato-renal) syndrome. *Biochem. Biophys. Res. Commun.* **12**, 179-184.

Shimozawa, N., Tsukamoto, T., Suzuki, Y., Orii, T., Shirayoshi, Y., Mori, T. and Fujiki, Y. (1992) A human gene responsible for Zellweger syndrome that affects peroxisome assembly. *Science* **255** , 1132-1134.

Snow, R. (1966) An enrichment method for auxotrophic yeast mutants using the antibiotic 'Nystatin'. *Nature* **211**, 206-207.

Subramani, S. (1992) Targeting of proteins into the peroxisomal matrix. *J. Memb. Biol.* **125**, 99-106.

Sutton, A., and Broach, J.R. (1985) Signals for transcription initiation and termination in the *Saccharomyces cerevisiae* plasmid 2 μm circle. *Mol. Cell. Biol.* **5**, 2770-2780.

Suzuki, Y., Orii, T., Mori, M., Tatibana, M. and Hashimoto, T. (1986) Deficient activities and proteins of peroxisomal ß-oxidation enzymes in infants with Zellweger's syndrome. *Clin. Chim. Acta* **156**, 191-196.

Suzuki, Y., Shimozawa, N., Yajima, S., Orii, T., Yokota, S., Tashiro, Y., Osumi, T. and Hashimoto, T. (1992) Different intracellular localization of peroxisomal proteins in fibroblasts from different patients with aberrant peroxisome assembly. *Cell. Struct. Funct.* **17**, 1-8.

Svoboda, D.J. and Azarnoff, D.L. (1966) Response of hepatic microbodies to a hypolipidemic agent, ethyl chlorophenoxyisobutyrate (CPIB). *J. Cell Biol.* **30**, 442-450.

Swinkels B.W., Gould, S.J. and Subramani, S. (1992) Targeting efficiencies of various permutations of the consensus C-terminal tripeptide peroxisomal targeting signal. *FEBS Lett.* **305**, 133-136.

Swinkels, B.W., Gould, S.J., Bodnar, A.G., Rachubinski, R.A. and Subramani, S. (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of rat 3-ketoacyl-CoA thiolase. *EMBO J.* **10**, 3255-3262.

Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci.* USA **84**, 4767-4771.

Tager, J.M., van der Beek, W.A., Wanders, R.J., Hashimoto, T., Heymans, H.S.A., van den Bosch, H., Schutgens, R.B. and Schram, A.W. (1985) Peroxisomal ß-oxidation enzyme proteins in the Zellweger syndrome. *Biochem. Biophys. Res. Cummun.* **126**, 1269-1275.

Tanaka, A., Osumi, M. and Fukui, S. (1982) Peroxisomes of alkane-grown yeast: fundamental and practical aspects. *Ann. N. Y. Acad. Sci.* **386**, 183-199.

Thieringer, R., Shio, H., Han, Y., Cohen, G. and Lazarow, P.B. (1991) Peroxisomes in *Saccharomyces cerevisiae*: Immunofluorescence analysis and import of catalase A into isolated peroxisomes. *Mol. Cell. Biol.* **11**, 510-522.

Thorp, J.M. and Waring, W.S. (1962) Modification of metabolism and distribution of lipids by ethyl chlorophenoxyisobutyrate. *Nature* **194**, 948-949.

Tolbert, N. (1981) Metabolic pathways in peroxisomes and glyoxysomes. *Ann. Rev. Biochem.* **50**, 133-157.

Tsukamoto, T., Miura, S. and Fujiki, Y. (1991) Restoration by a 35K membrane protein of peroxisome assembly in a peroxisome-deficient mammalian cell mutant. *Nature* **350**, 77-81.

Uritani, M. and Miyazaki, M. (1988) Characterization of the ATPase and GTPase activities of elongation factor 3 (EF-3) purified from yeasts. *J. Biochem. (Tokyo)* **103**, 522-530.

van den Bosch, H., Schutgens, R., Wanders, R. and Tager, J. (1992) Biochemistry of peroxisomes. *Ann. Rev. Biochem.* **61**, 157-197.

Veenhuis, M., Keizer, I. and Harder, W. (1979) Characterization of peroxisomes in glucose-grown *Hansenula polymorpha* and their development after the transfer of cells into methanol-containing media. *Arch. Microbiol.* **120**, 167-175.

Veenhuis, M. and Goodman, J.M. (1990) Peroxisomal assembly: Membrane proliferation precedes the induction of the abundant matrix proteins in the methylotrophic yeast *Candida boidinii*. *J. Cell Sci.* **96**, 583-590.

Veenhuis, M. and Harder, W. (1987) Metabolic significance and biogenesis of microbodies in yeast. In: Fahimi, H. and Sies, H. (eds.) *Peroxisomes in Biology and Medicine*, pp. 436-458, Springer-Verlag, Berlin.

Veenhuis, M., Mateblowski, M., Kunau, W.-H. and Harder, W. (1987) Proliferation of microbodies in *Saccharomyces cerevisiae*. *Yeast* **3**, 77-84.

Veenhuis, M., van Dijken, J.P.., Pilon, S.A. and Harder, W. (1978) Development of crystalline peroxisomes in methanol-grown cells of the yeast *Hansenula polymorpha* and its relation to environmental conditions. *Arch. Microbiol.* **117**, 153-163.

Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.

Walton, P.A., Gould, S.J., Feramisco, J.R. and Subramani, S. (1992) Transport of microinjected proteins into peroxisomes of mammalian cells: inability of Zellweger cell lines to import proteins with the SKL tripeptide peroxisomal targeting signal. *Mol. Cell. Biol.* **12**, 531-541.

Wanders, R.J., Romeyn, G.J., Schutgens, R.B. and Tager, J.M. (1989) L-pipecolate oxidase: a distinct peroxisomal enzyme in man. *Biochem. Biophys. Res. Commun.* **164**, 550-555.

Wiebel, F.F. and Kunau, W.-H. (1992) The Pas2 protein is essential for peroxisome biogenesis and is related to ubiquitin-conjugating enzymes. *Nature* **359**, 73-76.

Wilson, D.W., Wilcox, C.A., Flynn, G.C., Chen, E., Kuang, W.-J., Henzel, W.J., Block, M.R., Ullrich, A. and Rothman, J.E. (1989) A fusion protein required for vesicle mediated transport in both mammalian cells and yeast. *Nature* **339**, 355-359.

Wilson, G. (1991) Structure function relationships in the peroxisome: implications for human disease. *Biochem. Med. Metab. Biol.* **46**, 288-298.

Xuan, J.-W., Fournier, P. and Gaillardin, C. (1988) Cloning of the *LYS5* gene encoding saccharopine dehydrogenase from the yeast *Yarrowia lipolytica* by targeted integration. *Curr. Genet.* **14**, 15-21.

Xuan, J.-W., Fournier, P., De Clerck , N., Chasles, M. and Gaillardin, C. (1990) Overlapping reading frames at the *LYS5* locus in the yeast *Yarrowia lipolytica*. *Mol. Cell. Biol.* **10**, 4795-4806.

Yajima, S., Suzuki, Y., Shimozawa, S., Orii, T., Fujiki, Y., Osumi, T., Hashimoto, T. and Moser, H. (1992) Complementation study of peroxisome-deficient disorders by immunofluorescence staining and characterization of fused cells. *Hum. Genet.* **88**, 491-499.

Zaret, K.S. and Sherman, F. (1982) DNA sequence required for efficient transcription termination in yeast. *Cell* **9**, 563-573.

Zellweger, H. (1989) Peroxisomopathies: new developments. *Dev. Med. Child. Neurol.* **31**, 689.

Zhang, H., Scholl, R., Browse, J. and Somerville, C. (1988) Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Res.* **16**, 1220.

Appendix A

The following pages contain the sequence of the 4.5 kbp common region in its entirety. The sense strand sequence (with respect to the *PAY4* ORF) is listed with the first nucleotide of the *PAY4* ORF designated +1. Note the underlined nucleotides in the 5' and 3' flanking regions as these denote possible regulatory elements as discussed in sections 3.4.2.0 and 3.4.2.1.

-996 GGATCC
-990 TCTTCCAGCTGCCCTTCTTTGTTTATGGAGCCTACAAGCTGCTAACCAAGACCGCCACTCCCACGACATACCTGTGGATGCTGGTGTACGGTGTCAATG
-891 CCAGCCTAACGACTCTGGCATGTCTCGCGGAGGTGTGGGCTCGAGCCGGTCTTACCGATGCCCTTCGATACAACCTGCTCGCTGTCTACGCCCCCTTTT
-792 TCTTCATTTCTGGCTACATTGTCATTGACGTATTCCAGCGGTTGCAGGGAGAGCTGAGAAAGGTCAAGACTGATTAAGCAACGAATTACTATAACCAAA
-693 CCGCCAAACAAGTGATGCCACACGACTCACGCAGGAATGTACCGCATACTATGTACTGTGTACTGTTCTGGCAATACCCAATGGTTCAGCTTTCCAACA
-594 TCCTCGACTAAGAAGGAACTACTGTATTGGAACTGTCACTGGTGCTGACGACTTGTGTGTTTATTGTCAATGAGATTGGACTCGGTTACTTACTCCTCC
-495 GGGCACTACAAGTACCGTACCTCAAGTCCTGTGTGTATATATTATCCGTATTATGATTTCAAAACCATTACATTGTAAAGTCTCTCTCAATTGACTGAT
-396 TCATATAGAACTACCAGTAGAAGCATCTCCGAAGTGTTGTTTTTTCTAGTACCGTATGTACAGTATTCATACTAGTTAGCTACAGTACATAACGCGCGC
-297 CCAGGCTAACATCATCACGGAACAACTTGTACTCATCATTCAGAGCCGAAATTTGAACAGAAACAAAAAGAATTGAATTTCCCATTATTCCAGTCACTA
-198 CTGGGAATCTTCACCAACATAAAAAACACCCACCTTGACGTCTATCACGTGCTTCATTCATACCGTTATTCTGGCTCTTCCCCACACCCCACTTTTCAC
-99 ATTAACCTTGATGTCCGCCAAACTCTCTTTTCTGCACGCTCCCTTAACCTCGGCCGACAGATCATAGTCATTCTTCTCTACAACACCACCACATCGAGC

+1 ATG CCG TCG ATC AGC CAC AAA CCC ATC ACA GCC AAG CTG GTG GCT GCC CCG GAT GCC ACC AAG CTC GAG CTC TCG
Met Pro Ser Ile Ser His Lys Pro Ile Thr Ala Lys Leu Val Ala Ala Pro Asp Ala Thr Lys Leu Glu Leu Ser 25

+76 TCC TAC CTG TAC CAG CAG CTC TTT TCC GAC AAG CCA GCC GAG CCA TAT GTT GCG TTC GAA GCT CCT GGA ATC AAG
Ser Tyr Leu Tyr Gln Gln Leu Phe Ser Asp Lys Pro Ala Glu Pro Tyr Val Ala Phe Glu Ala Pro Gly Ile Lys 50

+151 TGG GCT CTG TAT CCT GCT TCT GAG GAC CGA TCT CTG CCG CAG TAC ACT TGC AAG GCC GAC ATT CGC CAC GTG GCA
Trp Ala Leu Tyr Pro Ala Ser Glu Asp Arg Ser Leu Pro Gln Tyr Thr Cys Lys Ala Asp Ile Arg His Val Ala 75

+226 GGC AGT CTC AAA AAG TTC ATG CCC GTG GTG CTC AAA CGG GTC AAC CCT GTC ACG ATT GAG CAT GCC ATC GTG ACC
Gly Ser Leu Lys Lys Phe Met Pro Val Val Leu Lys Arg Val Asn Pro Val Thr Ile Glu His Ala Ile Val Thr 100

+301 GTG CCT GCC AGC CAG TAC GAA ACG CTT AAC ACA CCC GAA CAA GTG CTC AAG GCT CTC GAG CCG CAA CTC GAC AAG
Val Pro Ala Ser Gln Tyr Glu Thr Leu Asn Thr Pro Glu Gln Val Leu Lys Ala Leu Glu Pro Gln Leu Asp Lys 125

+376 GAC CGA CCG GTA ATC CGA CAG GGC GAC GTG CTG CTC AAC GGA TGC AGA GTG CGT CTG TGC GAG CCT GTA AAC CAG
Asp Arg Pro Val Ile Arg Gln Gly Asp Val Leu Leu Asn Gly Cys Arg Val Arg Leu Cys Glu Pro Val Asn Gln 150

+451 GGC AAG GTG GTC AAG GGA ACC ACC AAG CTG ACG GTC GCC AAA GAG CAG GAA ACC ATC CAA CCG GCC GAC GAA GCT
Gly Lys Val Val Lys Gly Thr Thr Lys Leu Thr Val Ala Lys Glu Gln Glu Thr Ile Gln Pro Ala Asp Glu Ala 175

+526 GCC GAC GTG GCC TTT GAC ATT GCA GAG TTT CTC GAC TTT GAC ACA TCT GTT GCC AAG ACC CGT GAA TCC ACA AAC
Ala Asp Val Ala Phe Asp Ile Ala Glu Phe Leu Asp Phe Asp Thr Ser Val Ala Lys Thr Arg Glu Ser Thr Asn 200

+601 CTC CAA GTG GCA CCT CTT GAA GGA GCC ATC CCT ACT CCT CTC TCG GAC CGG TTT GAC GAC TGC GAA AGC CGA GGC
Leu Gln Val Ala Pro Leu Glu Gly Ala Ile Pro Thr Pro Leu Ser Asp Arg Phe Asp Asp Cys Glu Ser Arg Gly 225

+676 TTC GTC AAG TCT GAA ACC ATG TCG AAA CTC GGA GTC TTT TCC GGC GAC ATT GTG TCC ATC AAG ACC AAA AAC GGA
Phe Val Lys Ser Glu Thr Met Ser Lys Leu Gly Val Phe Ser Gly Asp Ile Val Ser Ile Lys Thr Lys Asn Gly 250

+751 GCC GAA CGG GTG CTC CGA CTG TTT GCA TAC CCC GAA CCA AAC ACA GTC AAG TAC GAC GTG GTC TAC GTC TCG CCC
Ala Glu Arg Val Leu Arg Leu Phe Ala Tyr Pro Glu Pro Asn Thr Val Lys Tyr Asp Val Val Tyr Val Ser Pro 275

+826 ATT TTG TAC CAC AAC ATT GGC GAC AAG GAG ATT GAG GTG ACT CCC AAC GGT GAG ACA CAT AAG TCT GTA GGA GAG
Ile Leu Tyr His Asn Ile Gly Asp Lys Glu Ile Glu Val Thr Pro Asn Gly Glu Thr His Lys Ser Val Gly Glu 300

+901 GCG CTG GAT TCC GTG CTT GAA GCT GCT GAG GAA GTC AAG CTG GCA AGA GTG CTT GGT CCT ACT ACC ACA GAC AGA
Ala Leu Asp Ser Val Leu Glu Ala Ala Glu Glu Val Lys Leu Ala Arg Val Leu Gly Pro Thr Thr Thr Asp Arg 325

+976 ACC TTC CAA ACA GCC TAC CAC GCA GGT CTG CAG GCC TAC TTT AAG CCC GTA AAG AGA GCC GTT CGA GTA GGT GAT
Thr Phe Gln Thr Ala Tyr His Ala Gly Leu Gln Ala Tyr Phe Lys Pro Val Lys Arg Ala Val Arg Val Gly Asp 350

+1051 CTG ATC CCC ATT CCC TTT GAC TCT ATT CTC GCT CGA ACT ATT GGC GAA GAT CCT GAA ATG AGC CAC ATT CCT CTG
Leu Ile Pro Ile Pro Phe Asp Ser Ile Leu Ala Arg Thr Ile Gly Glu Asp Pro Glu Met Ser His Ile Pro Leu 375

+1126 GAA GCT CTG GCA GTC AAG CCC GAC TCC GTG GCG TGG TTT CAG GTC ACT TCT CTC AAC GGA AGT GAA GAC CCT GCC
Glu Ala Leu Ala Val Lys Pro Asp Ser Val Ala Trp Phe Gln Val Thr Ser Leu Asn Gly Ser Glu Asp Pro Ala 400

+1201 TCC AAA CAG TAC CTT GTG GAC TCC TCT CAG ACA AAG TTG ATC GAG GGA GGA ACT ACT TCT TCT GCT GTA ATC CCC
Ser Lys Gln Tyr Leu Val Asp Ser Ser Gln Thr Lys Leu Ile Glu Gly Gly Thr Thr Ser Ser Ala Val Ile Pro 425

+1276 ACC AGT GTT CCA TGG CGC GAA TAT CTC GGT CTG GAT ACC TTG CCC AAG TTT GGA TCC GAG TTT GCA TAC GCC GAC
Thr Ser Val Pro Trp Arg Glu Tyr Leu Gly Leu Asp Thr Leu Pro Lys Phe Gly Ser Glu Phe Ala Tyr Ala Asp 450

+1351 AAG ATC CGC AAC CTG GTT CAG ATC TCG ACC TCG GCT CTC TCG CAC GCC AAA CTC AAC ACC TCG GTA CTT TTG CAT
Lys Ile Arg Asn Leu Val Gln Ile Ser Thr Ser Ala Leu Ser His Ala Lys Leu Asn Thr Ser Val Leu Leu His 475

+1426 TCC GCC AAG CGA GGC GTC GGA AAG TCA ACT GTT CTT CGT TCC GTG GCT GCC CAG TGC GGA ATT TCT GTG TTC GAA
Ser Ala Lys Arg Gly Val Gly Lys Ser Thr Val Leu Arg Ser Val Ala Ala Gln Cys Gly Ile Ser Val Phe Glu 500

+1501 ATC TCG TGT TTT GGA CTT ATT GGA GAC AAT GAG GCC CAG ACA CTG GGA ACT CTG CGA GCC AAG CTA GAC CGG GCC
Ile Ser Cys Phe Gly Leu Ile Gly Asp Asn Glu Ala Gln Thr Leu Gly Thr Leu Arg Ala Lys Leu Asp Arg Ala 525

+1576 TAT GGC TGC TCT CCT TGC GTG GTC GTT CTG CAG CAT CTC GAG TCG ATT GCA AAG AAG AGC GAC CAG GAC GGA AAG

```
      Tyr Gly Cys Ser Pro Cys Val Val Val Leu Gln His Leu Glu Ser Ile Ala Lys Lys Ser Asp Gln Asp Gly Lys   550

+1651 GAT GAG GGA ATA GTT TCA AAG CTT GTG GAC GTT CTT GCG GAC TAC TCC GGA CAC GGA GTG CTG CTG GCA GCT ACG
      Asp Glu Gly Ile Val Ser Lys Leu Val Asp Val Leu Ala Asp Tyr Ser Gly His Gly Val Leu Leu Ala Ala Thr   575

+1726 TCT AAC GAT CCC GAC AAG ATT TCC GAG GCC ATT CGA TCT CGA TTT CAG TTT GAA ATT GAG ATC GGA GTG CCC TCT
      Ser Asn Asp Pro Asp Lys Ile Ser Glu Ala Ile Arg Ser Arg Phe Gln Phe Glu Ile Glu Ile Gly Val Pro Ser   600

+1801 GAG CCC CAA AGA CGT CAG ATC TTT TCT CAT CTG ACG AAA TCC GGT CCT GGG GGC GAT TCA ATC AGA AAT GCA CCC
      Glu Pro Gln Arg Arg Gln Ile Phe Ser His Leu Thr Lys Ser Gly Pro Gly Gly Asp Ser Ile Arg Asn Ala Pro   625

+1876 ATT TCT CTG CGT TCT GAT GTC TCT GTT GAG AAC CTT GCG CTC CAA TCT GCC GGT CTG ACC CCT CCT GAT CTC ACA
      Ile Ser Leu Arg Ser Asp Val Ser Val Glu Asn Leu Ala Leu Gln Ser Ala Gly Leu Thr Pro Pro Asp Leu Thr   650

+1951 GCC ATT GTC CAG ACT ACC CGT CTG AGA GCC ATT GAT CGA CTT AAC AAG CTC ACC AAG GAA CGT CCG ATA GCC ACT
      Ala Ile Val Gln Thr Thr Arg Leu Arg Ala Ile Asp Arg Leu Asn Lys Leu Thr Lys Glu Arg Pro Ile Ala Thr   675

+2026 CTA GAC GAT CTC CTG ACC CTT TCG CAT GGT ACT CTC CAG CTG ACT CCC TCC GAC TTT GAT GAC GCT ATT GCT GAT
      Leu Asp Asp Leu Leu Thr Leu Ser His Gly Thr Leu Gln Leu Thr Pro Ser Asp Phe Asp Asp Ala Ile Ala Asp   700

+2101 GCT CGA CAA AAG TAC TCC GAC TCC ATT GGA GCT CCC CGA ATC CCC AAT GTC GGA TGG GAC GAT GTT GGA GGC ATG
      Ala Arg Gln Lys Tyr Ser Asp Ser Ile Gly Ala Pro Arg Ile Pro Asn Val Gly Trp Asp Asp Val Gly Gly Met   725

+2176 GAG GGT GTC AAG AAG GAT ATT CTG GAC ACT ATC GAG ACT CCT CTA AAG TAC CCC CAC TGG TTC TCT GAC GGT GTA
      Glu Gly Val Lys Lys Asp Ile Leu Asp Thr Ile Glu Thr Pro Leu Lys Tyr Pro His Trp Phe Ser Asp Gly Val   750

+2251 AAG AAG CGA TCG GGT ATT CTG TTT TAC GGT CCT CCC GGT ACT GGT AAG ACT CTT CTC GCG AAA GCC ATT GCC ACG
      Lys Lys Arg Ser Gly Ile Leu Phe Tyr Gly Pro Pro Gly Thr Gly Lys Thr Leu Leu Ala Lys Ala Ile Ala Thr   775

+2326 ACT TTC TCG CTC AAC TTC TTC TCC GTC AAG GGT CCC GAG CTG CTC AAC ATG TAC ATT GGT GAA TCC GAG GCC AAC
      Thr Phe Ser Leu Asn Phe Phe Ser Val Lys Gly Pro Glu Leu Leu Asn Met Tyr Ile Gly Glu Ser Glu Ala Asn   800

+2401 GTG CGA CGA GTA TTC CAA AAG GCC CGA GAT GCC AAG CCC TGT GTT GTC TTC TTT GAC GAG TTG GAT TCC GTG GCG
      Val Arg Arg Val Phe Gln Lys Ala Arg Asp Ala Lys Pro Cys Val Val Phe Phe Asp Glu Leu Asp Ser Val Ala   825

+2476 CCT CAG AGA CGG AAC CAG GGA GAC TCT GGA GGA GTC ATG GAC CGA ATT GTG TCG CAG CTG CTA GCT GAG CTC GAT
      Pro Gln Arg Arg Asn Gln Gly Asp Ser Gly Gly Val Met Asp Arg Ile Val Ser Gln Leu Leu Ala Glu Leu Asp   850

+2551 GGA ATG TCT ACT GCG GGA GGA GAG GGT GTT TTC GTT GTG GGA GCT ACT AAC CGT CCT GAT CTG TTG GAC GAG GCT
      Gly Met Ser Thr Ala Gly Gly Glu Gly Val Phe Val Val Gly Ala Thr Asn Arg Pro Asp Leu Leu Asp Glu Ala   875

+2626 CTT CTG CGA CCT GGA CGA TTC GAT AAG ATG CTG TAC TTG GGT ATC TCT GAT ACC CAC GAG AAG CAG CAG ACT ATT
      Leu Leu Arg Pro Gly Arg Phe Asp Lys Met Leu Tyr Leu Gly Ile Ser Asp Thr His Glu Lys Gln Gln Thr Ile   900

+2701 ATG GAG GCT CTT ACT CGA AAG TTC CGA CTT GCT GCC GAT GTG TCT CTT GAG GCC ATC TCC AAA CGA TGT CCG TTT
      Met Glu Ala Leu Thr Arg Lys Phe Arg Leu Ala Ala Asp Val Ser Leu Glu Ala Ile Ser Lys Arg Cys Pro Phe   925

+2776 ACT TTC ACC GGC GCC GAT TTC TAC GCT CTG TGT TCA GAT GCC ATG CTA AAT GCC ATG ACT CGA ACT GCC AAC GAG
      Thr Phe Thr Gly Ala Asp Phe Tyr Ala Leu Cys Ser Asp Ala Met Leu Asn Ala Met Thr Arg Thr Ala Asn Glu   950

+2851 GTT GAT GCC AAG ATC AAG CTG CTC AAC AAG AAC AGG GAG GAG GCC GGA GAG GAG CCC GTC TCC ATC AGA TGG TGG
      Val Asp Ala Lys Ile Lys Leu Leu Asn Lys Asn Arg Glu Glu Ala Gly Glu Glu Pro Val Ser Ile Arg Trp Trp   975

+2926 TTT GAC CAC GAG GCT ACC AAG AGT GAC ATT GAG GTT GAG GTT GCT CAG CAG GAT TTT GAG AAG GCC AAG GAC GAG
      Phe Asp His Glu Ala Thr Lys Ser Asp Ile Glu Val Glu Val Ala Gln Gln Asp Phe Glu Lys Ala Lys Asp Glu  1000

+3001 CTG AGT CCA TCT GTT TCG GCG GAG GAG CTG CAA CAC TAT CTC AAG CTG AGA CAG CAA TTT GAG GGA GGC AAG AAA
      Leu Ser Pro Ser Val Ser Ala Glu Glu Leu Gln His Tyr Leu Lys Leu Arg Gln Gln Phe Glu Gly Gly Lys Lys  1025

+3076 TGA ACAGACGAGGGATGATGGGTGGATTTGTATAAATATATTACGTGATAATTATGACATTGCTCAAATACTGGAGTAGGAAGACCACCAAAACTTGA
      ***

+3174 TTATTTACACCGATAAGAGAAATTATATTCTCACAGTACAATCTGCTTGCGTCACGCAACTACTCGTACTGGTTTTTATGTTTTTTTATGTTTTTTATA
+3273 TTAATTTTTTTTTAACTTTTTATAAATATTTGGATTATAGGAAAAATGTCTCGTTCCTTAGTGCATCTGAGATGCCCCAAAATTACAGTAATCGGACTG
+3372 CAAAAGTGGGGACGACTGCATGACGGAGATAGGAGTGATGATATTGAGCAGTGAACTGTACTCGTGCTTGTGCTGTCGTGTTGAAATTGGGGATGAAGA
+3471 AAAAACCACCGGACAAAATCTTTCCAATCACGACCAAAATATAATCGACACCTTTCTACCGACTTTTCTATCAACGGTCTCCATCTACAAGTAGTGGAT
+3570 CC
```