MOLECULAR CLONING AND EXPRESSION OF CYTOCHROME P-450 MONOOXYGENASES FROM *RHODOTORULA* SPP. IN *YARROWIA LIPOLYTICA*

BY

ANDREAS NDAFUDIFWA SHININGAVAMWE

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

PHILOSOPHIAE DOCTOR

IN THE

DEPARTMENT OF MICROBIAL, BIOCHEMICAL & FOOD BIOTECHNOLOGY
FACULTY OF NATURAL AND AGRICULTURAL SCIENCES
UNIVERSITY OF THE FREE STATE
BLOEMFONTEIN 9300
SOUTH AFRICA

MAY 2004

PROMOTER: PROF. M.S. SMIT
CO-PROMOTER: DR. J. ALBERTYN

“Imagination is more important than knowledge...”

Albert Einstein (1879-1955)

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards the following people:

Prof. M.S. Smit for her invaluable guidance, patience and support throughout this project especially when "tough" was really pushing its way. Thank you for the hard work you have invested in this project.

Dr. J. Albertyn for his guidance and always willing to help and share some "molecular gossips". Thank you for your time!

Dr. J-M. Nicaud for making his expression vectors available to my project.

Dr. M. E. Setati for making her recombinant strain available to my project.

The Kakia family for their moral support and cheering me up when I was feeling down.

The Nangombe family for their encouragement especially when "tough" was going.

Katrina Lugambo Ashitse for her moral support, caring, love and patience that she expressed to me during the difficult times of my project.

My family: my mother (**to whom I dedicate this thesis**), brothers and sisters for their prayers and support.

National Research Foundation (**NRF**) for the financial support of this project.

To all friends for their support and friendship that they expressed to me.

Above all, the **Heavenly Father**.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

1.1 Introduction 1
1.2 Aim of study 4
1.3 References 6

CHAPTER 2: PCR BASED METHODS FOR THE CLONING OF UNKNOWN DNA SEQUENCES FLANKING A KNOWN SEQUENCE - LITERATURE REVIEW

2.1 Introduction 11
2.2 Rapid amplification of cDNA ends (RACE) 12
2.3 The RACE PEETA (primer extension, electrophoresis, elution, tailing and amplification) technique 16
2.4 CapFinder method 18
2.5 Step out PCR 18
2.6 Rapid amplification of genomic ends (RAGE) 20
2.7 Inverse PCR (IPCR) 21
2.8 Panhandle PCR 23
2.9 Ligation-Mediated PCR 25
2.9.1 Biotin capture PCR 26
2.9.2 Vectorette PCR 27
2.9.3 Splinkerette PCR 29
2.9.4 Ligation –Anchored PCR (LA-PCR) 30
2.9.5 Reverse ligation –mediated PCR (RLM-PCR) 30
2.10 Suppression PCR 31
2.11 Chase PCR 32
2.12 Random PCR 34
2.13 Non-specific primed, Nested suppression PCR (NSPS-PCR) 34

2.14 Conclusions 36
2.15 References 39

CHAPTER 3: MOLECULAR CLONING AND CHARACTERIZATION OF A NOVEL CYTOCHROME P450 GENE OF *RHODOTORULA* SP. CBS 8446

3.1 Introduction 47
3.2 Materials and Methods 48
3.2.1 Strains, vectors and media 48
3.2.2 Isolation of total genomic DNA 49
3.2.3 PCR amplification of P450 gene fragments 49
3.2.4 Nucleotide sequence analysis 50
3.2.5 Southern hybridization 51
3.2.6 Inverse Polymerase Chain Reaction (IPCR) amplification of *Pst*I fragment 52
3.2.7 Isolation of total RNA and preparation of poly $(A)^{+}$ mRNA 53
3.2.8 RT-PCR amplification of mRNA isolated from decane and limonene grown cells 54
3.2.9 Amplification of flanking regions of cDNA using the 5'/3' RACE technique 54
3.2.10 RT- PCR amplification of 5' end of cDNA using walking primers 55
3.2.11 Isolation of full-length cDNA by RT-PCR amplification 56

3.3 Results 57
3.3.1 PCR amplification and sequence analyses of cytochromeP450 gene fragments 57
3.3.2 Southern hybridization 58
3.3.3 Inverse Polymerase Chain Reaction (IPCR) amplification of the putative CYP gene from the *Pst*I fragment from *Rhodotorula* sp. CBS 8446 59

3.3.4 RT-PCR amplification of mRNA isolated from cells grown on limonene and decane 63
3.3.5 Application of Rapid Amplification of cDNA Ends (5'/3' RACE) technique 64
3.3.6 RT-PCR amplification of 5' end of cDNA using walking primers 66
3.3.7 Isolation of full-length cDNA 69
3.3.8 Sequence and alternative splice variants analyses 70
3.3.9 Analyses of flanking sequences 71
3.3.10 Deduced protein 74
3.4 Discussion 79
3.5 References 83

CHAPTER 4: FUNCTIONAL EXPRESSION OF CYTOCHROME P450 MONOOXYGENASES FROM *RHODOTORULA* SPP. IN *YARROWIA LIPOLYTICA*

4.1 Introduction 94
4.2 Materials and Methods 98
4.2.1 Strains, vectors and media 98
4.2.2 Isolation of total RNA from *Rhodotorula minuta* grown in chemical defined medium (CD) supplemented with L-phenylalanine 100
4.2.3 Isolation of total RNA and preparation of poly $(A)^+$ mRNA from *Rhodotorula* sp. CBS 8446 grown on decane 100
4.2.4 RT-PCR amplification of mRNA isolated from *Rhodotorula minuta* grown in chemical defined medium (CD) supplemented with L-phenylalanine 101
4.2.5 RT-PCR amplification of mRNA isolated from *Rhodotorula* sp. CBS 8446 grown on decane 102
4.2.6 Preparation of competent *Yarrowia lipolytica* cells 103
4.2.7 Transformation of *Yarrowia lipolytica* with vectors containing cDNA 104
4.2.8 Southern Hybridization 104
4.2.9 Biotransformation of benzoic acid 105

4.2.10 Biotransformation of hexylbenzene 106
4.2.11 Biotransformation of oleic acid 106
4.3 Results 107
4.3.1 Functional expression of the benzoate-*para*-hydroxylase from *Rhodotorula minuta* in *Y. lipolytica* 107
4.3.1.1 RT-PCR amplification of CYP53B1 cDNA 107
4.3.1.2 Biotransformation studies of benzoic acid 109
4.3.1.3 Confirmation of integration of *CYP53B1* into the *Yarrowia lipolytica* genome 110
4.3.1.4 Expression of a cytochrome NADPH reductase (*CPR*) gene in *Y. lipolytica* 111
4.3.2 Functional expression of the fatty acid omega-hydroxylase from *Rhodotorula* sp. CBS 8446 in *Y. lipolytica* 113
4.3.2.1 RT-PCR amplification of CYP557A1 cDNA and transformation into *Y. lipolytica* 113
4.3.2.2 Biotransformation of alkylbenzenes 114
4.3.2.3 Effect of transformation with *CYP557A1* on oleic acid consumption 116
4.4 Discussion 118
4.5 References 123

CHAPTER 5: CONCLUDING REMARKS 131

Summary 139
Opsomming 141

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

Cytochrome P450 proteins derived their name from their absorption of UV radiation at 450 nm when carbon monoxide (CO) is bound to the heme group (Omura and Sato, 1964). They are heme-containing monooxygenases that are widely distributed in nature and found in bacteria, archae, yeasts, plants and animals (Käppeli, 1986; Nelson, 1999). These monooxygenases contribute to vital processes in the cell such as carbon source assimilation, biosynthesis of hormones and detoxification of drugs, xenobiotics and carcinogens (Porter and Coon, 1991; Werck-Reichhart and Feyereisen, 2000).

In prokaryotes the P450s are soluble proteins, while in eukaryotes they are usually anchored in membranes of the endoplasmic reticulum (ER) or inner mitochondrial membranes by a hydrophobic N-terminal region (Black, 1992; Scheller *et al*., 1994; Menzel *et al*., 1996; Nelson and Strobel, 1988). In the cytochrome P450s the helix I (HI) and heme binding (HR2) domains are highly conserved (Gotoh, 1992). The heme-binding domain has a consensus sequence of Phe-X-X-Gly-X-Arg-X-Cys-X-Gly. The conserved cysteine acts as a fifth ligand to the heme iron. The helix I domain has a consensus sequence of Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser. This region corresponds to the proton transfer groove on the distal side of the heme and has been implicated in substrate recognition (Gotoh, 1992).

Cytochrome P450s use electrons from NAD(P)H to catalyze activation of molecular oxygen leading to the regiospecific and stereospecific oxidative attack of substrates (Capdevila *et al*., 1984; Sutter *et al*., 1990; Vogel *et al*., 1992; Gilewicz *et al*., 1979).

Cytochrome P450 proteins can be divided into four classes depending on how electrons are transferred from NAD(P)H to the catalytic site (Werck-Reichhart and Feyereinsen, 2000). Class I proteins require both an FAD-containing reductase and an iron sulfur redoxin. Class II proteins require only an FAD/FMN–containing reductase for transfer of electrons. Class III proteins are self-sufficient and do not require an electron donor or molecular oxygen for catalysis. Class IV proteins receive electrons directly from NAD(P)H. Class I and II are found in prokaryotes as well as in eukaryotes, while class III is only found in eukaryotes where it is responsible for synthesis of prostaglandins and jasmonate. Class IV is a unique eukaryotic soluble P450 that is only found in fungi.

Different researchers had proposed mechanisms to explain catalysis by cytochrome P450s (Porter and Coon, 1991; Sakaki and Inouye, 2000; Werck-Reichhart and Feyereinsen, 2000). The active center for catalysis is the iron-protoporphyrin IX (heme) with the thiolate of the conserved cysteine residue as a fifth ligand. The resting P450 is in the ferric form with the sixth coordination position occupied by a water molecule (Fig. 1). The first step is binding of the substrate to the P450 with displacement of water (the sixth ligand) and the first electron is transferred from NADPH via NADPH-P450 reductase to the P450 to reduce the ferric ion to the ferrous state. The next step is the binding of molecular oxygen and transfer of the second electron that leads to 'an activated oxygen' species. In the last step the O-O bond is cleaved to form water and an oxygenated or hydroxylated product. The dissociation of the product and the enzyme restores the P450 to the ferric state.

In mammals the P450 system has been observed in multiple forms that play an important role in metabolic functions such as steroid synthesis and detoxification of xenobiotics and drugs (Gonzalez, 1988; 1990, Nebert *et al.*, 1991). In plants the P450s are involved in biosynthesis or catabolism of hormones, oxidation of fatty acids for synthesis of cutins and the synthesis of flower pigments and defense chemicals (Tijet *et al.*, 1998; Pinot *et al.*, 1999). In microorganisms cytochrome P450s are involved in the metabolism of *n*-alkanes and fatty acids, synthesis of mycotoxins and detoxifications of xenobiotics (Eschenfedt *et al.*, 2003; Scheller *et al.,* 1998; Fukui and Tanaka, 1981).

The cytochrome P450s with their ability to degrade xenobiotics and to detoxify drugs plus their wide range of substrate specificities lend themselves to be applied in bioremediation and for the synthesis of fine chemicals and pharmaceuticals. Understanding of the structure-function relationships of cytochrome P450s can lead to generation of new proteins through protein engineering and mutagenesis.

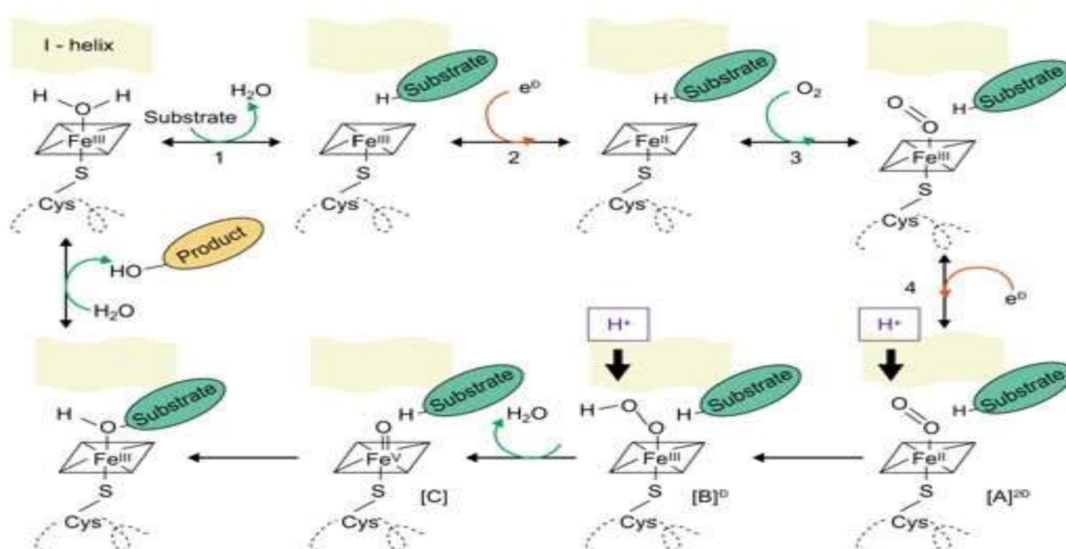


Figure 1. A schematic illustration showing the mechanism action of cytochrome P450. [A], [B] and [C] are iron-oxo intermediate species. Species [A] and [B] are nucleophiles, while [C] is electrophile (Taken from Werck-Reichhart and Feyereinsen, 2000).

1.2 Aim of study

One type of hydroxylation reactions, which has been extensively studied in alkane utilizing microorganisms is the hydroxylation of alkanes and fatty acids. Two different, but very similar, enzymatic steps are important in alkane utilization by microorganisms: the mono-terminal hydroxylation of alkanes that is regarded as a first and rate–limiting step in alkane degradation as well as ω-hydroxylation of fatty acids (Eschenfedt *et al*., 2003; Craft *et al*., 2003; Scheller *et al.,* 1998; Fukui and Tanaka, 1981; Schunck *et al*., 1987, Blasig, 1988; Zimmer *et al*., 1996). The *n*-alkanes are hydroxylated to fatty alcohols that are subsequently oxidized to fatty aldehydes and fatty acids. The hydroxylation of fatty acids at the omega position and subsequent oxidation can lead to dioic acids, which are useful chemical products (Eschenfedt *et al*., 2003; Craft *et al*., 2003; Picataggio *et al*., 1992; Scheller *et al.,* 1998).

The alkane hydroxylating P450s from alkane utilizing ascomycetous yeasts are genetically well characterized (Iida *et al.,* 1998; 2000; Lottermoser *et al*., 1996; Schunck *et al.,* 1989; Ohkuma *et al*., 1991a, 1991b; Craft *et al*., 2003; Yadav and Loper, 1999). They have been classified in the CYP52 family and multiple genes are present in most alkane utilizing ascomycetous yeasts. However, no *n*-alkane or fatty acid hydroxylase encoding genes have yet been isolated from basidiomycetous yeasts such as *Rhodotorula* spp. The only cytochrome P450 encoding gene isolated from a *Rhodotorula* sp. is the *benzoate-para-hydroxylase* gene (*CYP53B1*) that was cloned from *Rhodotorula minuta* (Fujii *et al.,* 1997). Studies performed in our laboratory have shown that some *Rhodotorula* spp. can utilize not only *n*-alkanes and benzoic acid but also monoterpenes as carbon sources (Moleleki, 1998). A PCR fragment of approximately 600 bp was isolated from *Rhodotorula* sp. CBS 8446 using degenerate primers based on the conserved helix I and heme binding domains of 15 different cytochrome P450 proteins from the CYP52 family (Moleleki, 1998). The sequence of this fragment showed homology to known P450s and was used as a starting point for the cloning of this P450 gene.

The goals of the present study were:

1) To compile a literature review on PCR based methods that can be used for cloning the unknown DNA sequences flanking a known sequence.
2) Isolation and sequencing of *n*-alkane or fatty acid hydroxylase encoding gene(s) from the basidiocycetous yeast *Rhodotorula* sp. CBS 8446.
3) Heterologous expression of *Rhodotorula* P450 monooxygenases in *Yarrowia lipolytica*:
   a) Expression of the *Rhodotorula minuta* benzoate -*para*-hydroxylase (*CYP53B1*) gene.
   b) Expression of the *Rhodotorula* sp. CBS 8446 newly isolated fatty acid omega hydroxylase (*CYP557A1*) gene.

1.3 References

Black, SD. (1992). Membrane topology of the mammalian P450 cytochromes**.** *FASEB J*. **6**: 680-685.

Blasig, R., Mauersberger, S., Riege, P., Schunck, WH., Jockisch, W. Franke, P. and Müller, HG. (1988). Degradation of long- chain n-alkanes by yeast *Candida maltosa*. *Appl. Microbiol Biotechnol* **28**: 589-597.

Capdevila, J., Saeki, Y. and Falck, JR. (1984). The mechanistic plurality of cytochrome P-450 and its biological ramifications. *Xenobiotica*. **14**: 105-118.

Craft, DL., Madduri, KM., Eshoo, M. and Wilson, CR. (2003). Identification and characterization of the CYP52 family of *Candida tropicalis* ATCC 20336, important for the conversion of fatty acids and alkanes to alpha, omega-dicarboxylic acids. *Appl Environ Microbiol*. **69**: 5983-5991.

Eschenfeldt, WH., Zhang, Y., Samaha, H., Stols, L., Eirich, LD., Wilson, CR. and Donnelly, MI. (2003). Transformation of fatty acids catalyzed by cytochrome P450 monooxygenase enzymes of *Candida tropicalis*. *Appl Environ Microbiol*. **69**: 5992-5999.

Fujii, T., Nakamura, K., Shibuya, K., Tanase, S., Gotoh, O., Ogawa, T. and Fukuda, H. (1997). Structural characterization of the gene and corresponding cDNA for the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. *Mol Gen Genet*. **256**: 115-120.

Fukui, S. and Tanaka, A. (1981). Metabolism of alkanes by yeasts. *Adv. Biochem. Eng*. **19**: 217-237.

Gilewicz, M., Zacek, M., Bertrand, J-C. and Azoul ay E. (1979). Hydroxylase regulation in *Candida tropicalis* grown on alkanes. *Can. J. Microbiol*. **25**: 201-206.

Gonzalez, FJ. (1988). The molecular biology of cytochrome P450s. *Pharmacol Rev*. **40**: 243-288.

Gonzalez, FJ. (1990). Molecular genetics of the P-450 superfamily. *Pharmacol Ther*. **45**: 1-38.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem*. **267**: 83-90.

Iida, T, Ohta, A. and Takagi, M. (1998). Cloning and characterization of an *n*-alkane-inducible cytochrome P450 gene essential for *n*-decane assimilation by *Yarrowia lipolytica*. *Yeast*. **14**: 1387-1397.

Iida, T., Sumita, T., Ohta, A. and Takagi, M. (2000). The cytochrome P450ALK multigene family of an *n*-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning and characterization of genes coding for new CYP52 family members. *Yeast*. **16**: 1077-1087.

Käppeli, O. (1986). Cytochromes P-450 of yeasts. *Microbiol Rev*. **50**: 244-258.

Lottermoser, K., Schunck, WH. and Asperger, O. (1996). Cytochrome P450 of the sophorose lipid-producing yeast *Candida apicola*: heterogeneity and polymerase chain reaction - mediated cloning of two genes. *Yeast*. **12**: 565-575.

Menzel, R., Kargel, E., Vogel, F., Bottcher, C. and Schunck, WH. (1996). Topogenesis of a microsomal cytochrome P450 and induction of endoplasmic reticulum membrane proliferation in *Saccharomyces cerevisiae*. *Arch Biochem Biophys*. **330**: 97-109.

Moleleki, N. (1998). The involvement of cytochrome P-450 monooxygenases in the hydroxylation of monoterpenes by yeasts (M.Sc. Thesis).

Nebert, DW., Nelson, DR., Coon, MJ., Estabrook, RW., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, FJ., Guengerich, FP., Gunsalus, IC., Johnson, EF., Loper, JC., Sato, R., Waterman, MR. and Waxman, DJ. (1991). The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. DNA *Cell Biol*. **10**: 1-14.

Nelson, DR. (1999). Cytochrome P450 and the individuality of species. *Arch Biochem Biophys*. **369**: 1-10.

Nelson, DR. and Strobel, HW. (1988). On the membrane topology of vertebrate cytochrome P-450 proteins. *J Biol Chem.* **263**: 6038-6050.

Ohkuma, M., Hikiji, T., Tanimoto, T., Schunck, W-H., Müller, H-G., Yano, K. and Takagi, M. (1991a). Evidence that more than one gene encodes *n*-alkane-inducible cytochrome P-450s in *Candida maltosa*, found by two-step gene disruption. *Agric. Biol. Chem*. **55**: 1757-1764.

Ohkuma, M., Tanimoto, T., Yano, K. and Takagi, M. (1991b). *CYP*52 (cytochrome P450alk) multigene family in *Candida maltosa*: molecular cloning and nucleotide sequence of the two tandemly arranged genes. *DNA Cell Biol*. **10**: 271-282.

Omura, T. and Sato, R. (1964). The carbon monooxide-binding pigment of liver microsomes. *J. Biol. Chem*. **239**: 2379-2385.

Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J. and Eirich, LD. (1992). Metabolic engineering of *Candida tropicalis* for the production of long-chain dicarboxylic acids. *Biotechnology* (N Y). **10**: 894-898.

Pinot, F., Benveniste, I., Salaun, JP., Loreau, O., Noel, JP., Schreiber, L. and Durst, F. (1999). Production *in vitro* by the cytochrome P450 CYP94A1 of major C18 cutin monomers and potential messengers in plant-pathogen interactions: enantioselectivity studies. *Biochem J*. **342**: 27-32.

Porter, TD. and Coon, MJ. (1991). Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem*. **266**: 13469-13472.

Sakaki, T. and Inouye, K. (2000). Practical application of mammalian cytochrome P450. *J. Biosci. Bioeng*. **90**: 583-590.

Scheller, U., Kraft, R., Schroder, KL., Schunck, WH. (1994). Generation of the soluble and functional cytosolic domain of microsomal cytochrome P450 52A3. *J Biol Chem*. **269**: 12779-12783.

Scheller, U., Zimmer, T., Becher, D., Schauer, F. and Schunck, WH. (1998). Oxygenation cascade in conversion of *n*-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A3. *J Biol Chem*. **273**: 32528-32534.

Schunck, WH., Kargel, E., Gross, B., Wiedmann, B., Mauersberger, S., Kopke, K., Kiessling, U., Strauss, M., Gaestel, M. and Muller, HG. (1989). Molecular cloning and characterization of the primary structure of the alkane hydroxylating cytochrome P-450 from the yeast *Candida maltosa*. *Biochem Biophys Res Commun*. **161**: 843-850.

Schunck, WH., Mauersberger, S., Huth, J., Riege, P., and Müller, HG. (1987). Function and regulation of cytochrome P-450 in alkane- assimilating yeast. *Arch Microbiol*. **147**: 240-244.

Sutter, TR., Sanglard, D., Loper, JC. and Sangard, D. (1990). Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from *Candida tropicalis*. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins. *J Biol Chem*. **265**: 16428-16436.

Tijet, N., Helvig, C., Pinot, F., Le Bouquin, R., Lesot, A., Durst, F., Salaun, JP. and Benveniste, I. (1998). Functional expression in yeast and characterization of a clofibrate-inducible plant cytochrome P-450 (*CYP94A1*) involved in cutin monomers synthesis. *Biochem J*. **332**: 583-589.

Vogel, F., Gengnagel, C., Kärgel, E., Müller, H. and Schunck, W-H. (1992). Immunocytochemical localization of alkane - inducible cytochrome P-450 and its NADPH - dependent reductase in the yeast *Candida maltosa*. *Eur. J. Cell Biol*. **57**: 285-291.

Werck-Reichhart, D., Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biology*. **1**: 3003.1-3003.9

Yadav, JS. and Loper, JC. (1999). Multiple P450alk (cytochrome P450 alkane hydroxylase) genes from the halotolerant yeast *Debaryomyces hansenii*. *Gene*. **226**: 139-146.

Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. and Schunck, WH. (1996). The CYP52 multigene family of *Candida maltosa* encodes functionally diverse *n*-alkane-inducible cytochromes P450. *Biochem Biophys Res Commun*. **224**: 784-789.

CHAPTER 2 –LITERATURE REVIEW

PCR BASED METHODS FOR THE CLONING OF UNKNOWN DNA SEQUENCES FLANKING A KNOWN SEQUENCE

2.1 Introduction

The polymerase chain reaction (PCR) is a powerful tool for amplifying a specific DNA sequence starting with a very minute amount of DNA (Saiki *et al.,* 1988). PCR amplification is performed with two oligonucleotides that flank the known DNA fragment. It involves repeated cycles of denaturation of DNA, annealing of primers to their complementary sequences and extension of the annealed primers by *Taq* polymerase, a thermostable enzyme. The primers hybridise to the opposite strands of the amplified DNA fragment doubling the amount of DNA in each cycle. In some instance unknown fragments can be amplified if it is flanked by conserved known sequences to which primers can bind. Basic PCR cannot be used to amplify a region that has never been characterized or that is not flanked by known sequences. This limits the usefulness of PCR in biotechnological research. To overcome this limitation, several PCR based methods such as the RACE technique (Frohman *et al.,* 1988), inverse PCR (Ochman *et al.,* 1988), ligation-mediated PCR methods (Mueller and Wold, 1989) and random PCR (Dominguez and Lopez-Larrea, 1994) have been developed to amplify the unknown DNA sequences that flank a known sequence.

The RACE (rapid amplification of cDNA ends) method involves the amplification of the unknown 5' and 3' ends flanking the known sequence of cDNA. Several methods such as the Capfinder method (Schmidt and Mueller, 1999; Schramm *et al*., 2000), the PEETA (primer extension, electrophoresis, elution, tailing and amplification) method (Flouriot *et al*., 1999) and step out PCR (Matz *et al*., 1999) method have been devised to improve the

RACE method. Inverse PCR requires the inverse circularisation of DNA and a pair of primers facing in "outward" orientations. In ligation-mediated methods adaptors or linkers are ligated to the digested DNA to act as annealing sites for primers in order to amplify the unknown fragments. Random PCR methods use a gene specific primer in combination with a non-specific primer. Even though above mentioned methods appear simple and straightforward, each has its own problems. Therefore the aim of this review is to outline the principles of these PCR based methods, their applications in molecular biology and their limitations for cloning the unknown sequences flanking a known sequence.

2.2 Rapid amplification of cDNA ends (RACE)

To overcome the limitation of ordinary PCR, which can only amplify a fragment between two known sequences, Frohman *et al*, (1988) developed the 5'/3' RACE (rapid amplification of cDNA ends) technique in which cDNA is used as a template. This method is used to achieve amplification and cloning of the region between a single short sequence in the cDNA molecule and its unknown 3' or 5' end. Different researchers use different names for the 5'/3' RACE technique, some referred to it as one-single sided PCR (Ohara *et al*., 1989), or anchored PCR (A-PCR)(Loh *et al*., 1989). Even though there are different names for these methods they apply the same principles. In this discussion the term "RACE technique" will be used. This method is comprised of two parts: 3' end and 5' end amplifications.

a) 3' end amplification of cDNA

The synthesis of cDNA from RNA is performed using a oligo(dT_{15}) -anchored primer. The dT residue part of the primer contains 15 dT residues that are complementary to the polyA tail of eukaryotic mRNA with an anchor sequence flanking the end of the primer (Fig. 1). PCR is subsequently performed using the synthesised cDNA as a template with a gene specific primer (GSP) and an anchor specific primer (AP). The anchor specific primer binds to the flanking part of the oligo(dT)-anchor primer. The specificity of amplification depends on the base-

pairing of the GSP only to the molecules representing the cDNA of interest. The anchor specific primer also improves the specificity of amplification, because it has been observed that the long stretch of oligo(dT) can create a number of mismatches when it is used in PCR (Frohman *et al.*, 1988). When one-single sided PCR (Ohara *et al.*, 1989) is used to amplify the 3’ end, cDNA synthesis is performed using oligo(dT_{20}) instead of oligo(dT_{15}) that is used in 3’ RACE. In addition no oligo(dT) anchor primer is used.

b) 5’ end amplification of cDNA

cDNA is first synthesised from RNA using a gene specific primer (GSP1) and the synthesised cDNA is purified to remove excess primers (Fig. 1). A homopolymer A-tail is added at the 3’ end of the cDNA by terminal deoxynucleotidyl-transferase (TdT). The A-tailed cDNA is amplified in a PCR reaction using a second gene specific primer (GSP2) and an oligo(dT)-anchor primer. Finally, PCR is performed using the nested GSP3 primer and the anchor-specific primer (Fig. 1). The nested primer increases the specificity and efficiency of the amplification because it only binds to the DNA of interest. So called anchored PCR (Loh *et al.*, 1989) introduced a polyG tail instead of a polyA tail.

The RACE technique has been used for structural and expression studies of RNA molecules as well as amplification and cloning of rare mRNAs. It has for example been used to clone *int*-2, a gene that is present in multiple transcripts expressed at a very low abundance (Frohman *et al.,* 1988), to clone the 3’ region of the human parvalbumin gene (Berchtold *et al.*, 1989), to amplify the cDNA sequences for the skeletal muscle α-tropomyosins of a frog and zebrafish (Ohara *et al.*, 1989) and to characterize 5’ mRNA of human T cell receptor (TCR) δ chains (Loh *et al.,* 1989).

Even though this method has been widely applied, it can be difficult to obtain the desired results (Frohman *et al.*, 1988; Loh *et al.,* 1989; Ohara *et al.*, 1989; Kriangkum *et al.*, 1992; Schaefer, 1995). These difficulties can be attributed to the enzymatic steps that are involved. The high background of non-specific or truncated products can also be a

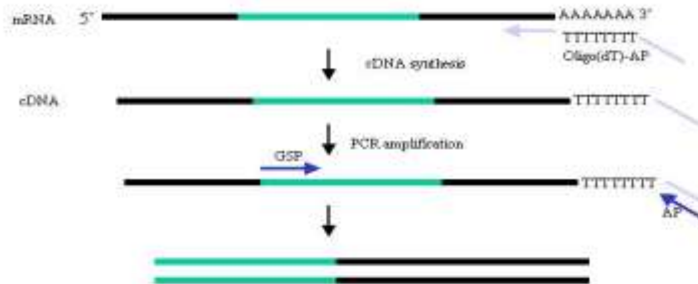
problem. Thus, there is a need for modifications in some steps. There are two crucial steps in the RACE technique: the reverse transcription step and addition of the homopolymeric tail to the 3’ end of the cDNA.

The reverse transcription step is crucial in the synthesis of full-length cDNA. During the cDNA synthesis truncated cDNAs are also generated. These short truncated sequences can reduce the generation of specific full-length cDNA. Truncation occurs when the reverse transcriptase is unable to continue extension in the region of mRNA that has formed stable structures. The transcripts with higher GC/AU ratios usually contain regions of these stable structures. The truncation of cDNA can also arise if the oligo(dT)-anchor primer binds in A-rich regions in the coding sequence upstream of the polyA tail.

The addition of the homopolymer to the 3’ end of the synthesized cDNA is the second critical step. The homopolymer serves as a substrate for the oligo(dT)-anchor primer in subsequent PCR. Sometimes tailing may fail for various reasons (Schaefer, 1995). During tailing full-length cDNAs, as well as truncated cDNAs, are tailed causing non-specific amplification. If the primers used for reverse transcription are not eliminated from the reaction they can also be tailed together with the cDNA. The homopolymer–tailed primers inhibit the amplification of tailed cDNA in PCR. The mechanism for inhibition is still unknown, however it is speculated that the homopolymer specific primer is depleted quickly as a result of side reactions (Kriangkum *et al*., 1992).

There are a few other problems that are experienced with the RACE technique. When different 5’ ends are isolated for instance in the case of genes that encode for different members of a multigene family, the cDNA synthesis must be performed for each transcript under investigation. This could be a time consuming procedure. This problem has been overcome by Harvey and Darlison, (1991) by using random hexanucleotides primers in combination with 5’ RACE. Another drawback of the RACE technique is the introduction of errors in the DNA as a result of multiple rounds of nested PCR amplifications.

3’ RACE



5’ RACE

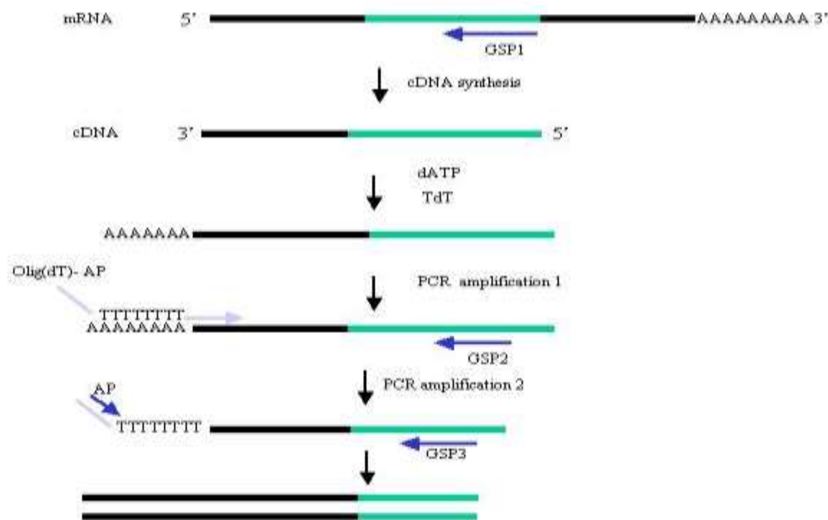


Figure 1. Schematic illustration of the RACE technique. cDNA synthesis is performed with oligo(dT)-AP (oligod(T_{15}) anchored primer or GSP), and it is tailed with TdT (terminal deoxynucleotidyl- transferase). PCR amplification is performed with AP (anchor specific primer) and GSP (gene specific primer). The figure is adapted from 5’/3’ RACE kit (Roche).

2.3 The RACE PEETA (primer extension, electrophoresis, elution, tailing and amplification) technique

Usually the 5' ends mapped by the RACE technique do not correspond to the actual transcription termination sites, because the reverse transcriptase prematurely terminates the reaction, thus generating products with different sizes. The shortest and most abundant products are preferentially amplified. This makes it difficult to study changes in 5' ends, which result from alternative splicing and promoter usage. It is also difficult to isolate specific extension products after reverse transcription and process them individually in the subsequent steps. To overcome these problems the PEETA (primer extension, electrophoresis, elution, tailing and amplification) technique was developed as an improvement of the 5' RACE method (Flouriot *et al*., 1999).

The technique involves the use of a biotinylated radioactively labelled long primer (>226 bp). Two specific primers are designed to amplify a known gene fragment that will act as the long primer in cDNA synthesis. One of the primers is biotinylated at the 5' end (Fig. 2). After purification using magnetic streptavidin beads the biotinylated long primer is radioactively labelled with [α ^{32}P]dCTP. The labelled primer is hybridized to RNA and reverse transcription is performed. The synthesised cDNA is separated on a denaturing polyacrylamide/urea gel. The band of interest is cut out of the gel and purified and a poly(C) tail is added to the purified product. PCR is subsequently performed using an oligo(dG)-anchor primer and a gene specific primer (Fig. 2).

The long labelled primer increases the sensitivity and specificity of the primer extension step. It has been observed that PEETA obtains more of the 5' ends of mRNA than the RACE technique. Thus showing that some parts of the 5' ends of cDNA are either incomplete or missing when RACE is used (Flouriot *et al*., 1999).

The PEETA method was used to identify the differentially expressed 5' end mRNA of the human isoforms of estrogen receptor-α (hER- α) (Flouriot *et al*., 1999). Thus, this method can be applied to study genes that are involved in alternative splicings.

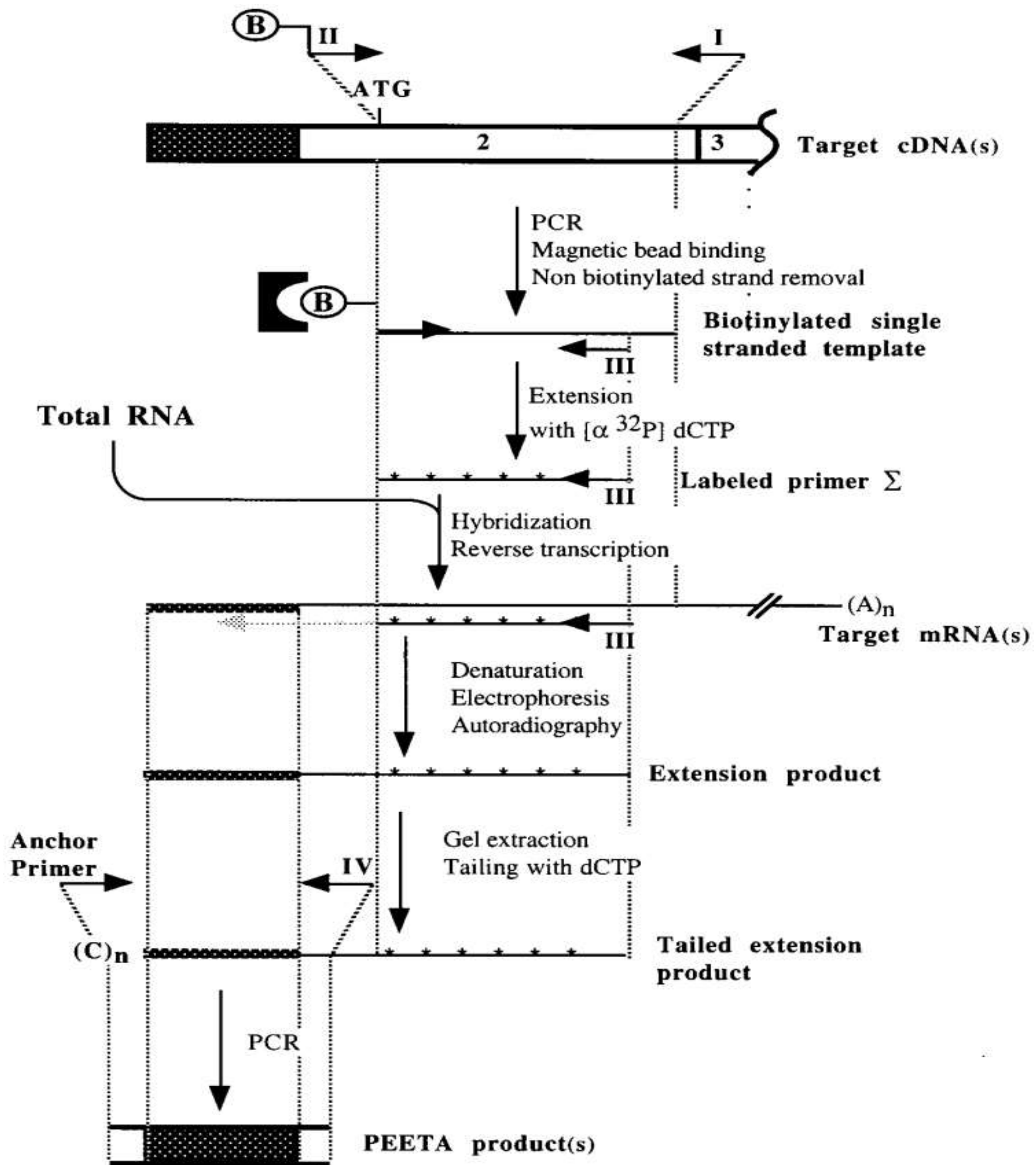


Figure 2. Schematic illustration of PEETA technique. Primer II is biotinylated, primer IV is nested to primer III which is nested to primer I. Shaded area represents an unknown region, while open rectangle represents the known fragment (Taken from Flouriot *et al.,* 1999).

2.4 CapFinder method

Usually the 5' end of genes is under-represented in cDNA populations especially if oligo(dT) is used in the cDNA synthesis and if the starting material is limited. In order to solve this problem the CapFinder method, that isolates the intact and complete 5' end of mRNA, has been developed (Schmidt and Mueller, 1999; Schramm *et al*., 2000). This method is another modification of the 5'RACE technique (Frohman *et al.,* 1988).
The CapFinder method takes advantage of template switching of reverse transcriptase. It has been observed that the reverse transcriptase adds up to four cytosine nucleotides upon reaching the 5' end of mRNA when $MgCl_2$ and $MnCl_2$ are present in the buffer (Schmidt and Mueller, 1999). When the primer containing an oligo(rG) (CapFinder oligonucleotide) is present in the reaction it base-pairs with the attached cytosine stretch. The reverse transcriptase then switches template and continues replicating the sequence of the CapFinder oligonucleotide including the complementary CapFinder oligonucleotide at the 3' end of the newly synthesized cDNA. The PCR is performed using the CapFinder specific primer and a gene specific primer. Recently this method has been modified by capturing the isolated RNA with biotinylated oligo(dT) (Schramm *et al*., 2000). The biotinylated oligo(dT) acts as a primer for cDNA synthesis and this modification decreases the background in the subsequent steps.

2.5 Step out PCR

In order to solve the problem of the high background that is produced by the RACE technique, Matz *et al*., (1999) devised the step out PCR method that combines the 5'/3' RACE technique (Frohman *et al*., 1988) with a technique called suppression PCR (Siebert *et al*., 1995). In addition, the step out PCR also takes advantage of the template switching effect that is carried out by reverse transcriptase (Schmidt and Mueller, 1999). Step out PCR uses two primers to replace a primer that tends to give a high background in PCR. It uses one short primer (SP) that has oligo(rG) for template switching effect and a long primer (LP)(~ 50 bp) with half of the 3'end identical to the short primer and in addition containing a 'heel' (flanking sequences) at the 5' end (Fig. 3). The cDNA

synthesis is performed using the gene specific primer (GSP) or oligo(dT) and the short primer is also included in the reaction. Whenever, the short primer is used in PCR in combination with a gene specific primer using the cDNA as a template, the high background is produced by a short primer alone. This occurs since during the reverse transcription the short primer is free to anneal not only at the oligo(C) stretch at the 3' end of the cDNA but non-specifically to anywhere in the RNA, and acts as a primer for reverse transcription.

In order to eliminate the high background the longer primer and the heel specific primer are designed to introduce the suppression effect. The mixture of the longer primer and the heel specific primer are used in combination with the gene specific primer in PCR (Fig. 3). Under these conditions the cDNA molecules that originated from in-strand annealing of the short primer are flanked by terminal inverted repeats that form 'pan'-like structures and their amplification is suppressed. This method applies the same principle as the chase PCR method (Lukyanov *et al*., 1995, Timblin *et al*., 1990), except chase PCR does not take advantage of template switching. To test the efficacy of step out PCR it was successfully used to amplify the interferon α receptor and interleukin 10, which are both found in low copy numbers in human placenta (Matz *et al*., 1999).

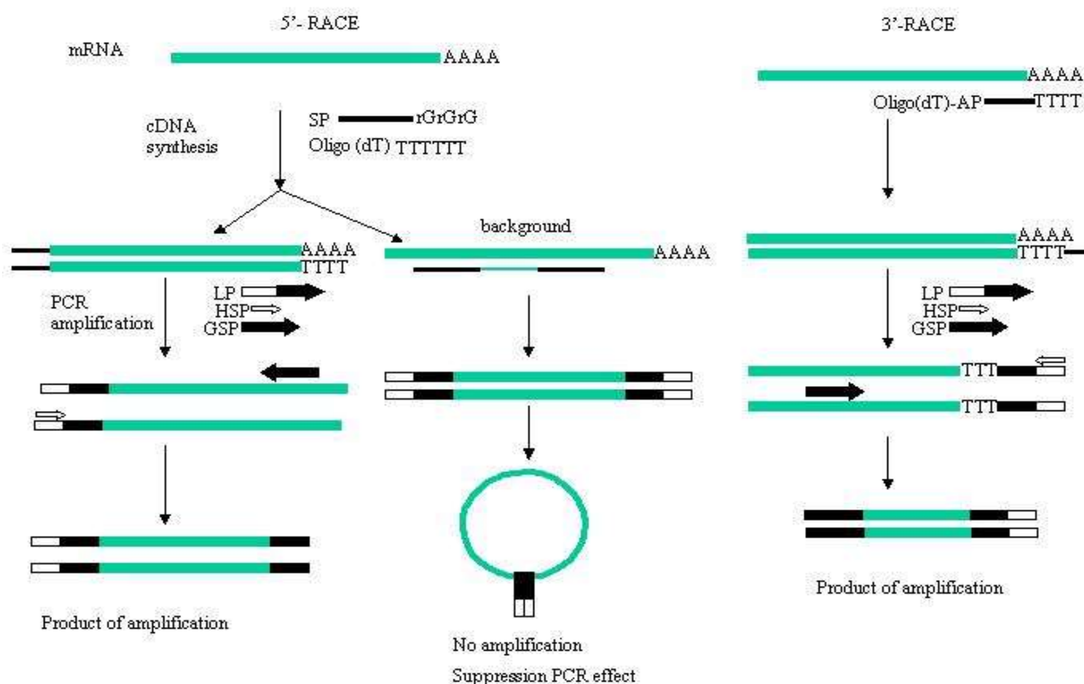


Figure 3. Schematic illustration of 5' and 3' RACE step out PCR. For 5', the cDNA synthesis is performed with oligo(dT) primer and SP (short primer). For 3', the cDNA is synthesised with oligo(dT)-AP (olig(dT) anchor primer). PCR amplification is performed with a mixture LP(long primer) and HSP (heel specific primer) plus the GSP (gene specific primer)(adapted from Matz *et al.,* 1999).

2.6 Rapid amplification of genomic ends (RAGE)

The RACE technique has been adapted to genomic DNA amplification and the modified method is referred to as rapid amplification of genomic ends (RAGE) (Bloomquist *et al*., 1992; Cormack and Somssich, 1997). This method involves the digestion of genomic DNA with any restriction enzyme. A homopolymeric A-tail is added to the 3' end with terminal transferase. The polyadenylated genomic DNA is used as a template in PCR using oligo(dT) primer and a gene specific primer. The RAGE technique can be modified by biotinylating one of the specific primers (Bloomquist *et al*., 1992). The biotinylated PCR products are bound to streptavidin- linked magnetic beads. The biotinylated PCR product is used as a template in PCR using nested gene specific primers. Biotin-RAGE overcomes the amplification of non-specific products by amplifying only the biotinylated

products in PCR. The RAGE method was used to clone the 5' promoter region of the *At*23 gene from *Arabidopsis thaliana* (Cormack and Somssich, 1997). This technique was also used to determine the exon 7 intron-exon junction of the rat gene coding for peptidylglycine α-amidating monooxygenase (*PAM*) (Bloomquist *et al*., 1992).

2.7 Inverse PCR (IPCR)

Thus far only RACE technique based methods have been discussed. In order to overcome some of the limitations of the RACE technique, Ochman *et al.* (1988) developed the inverse PCR (IPCR) method to amplify the unknown regions flanking known region. The IPCR is a simple method that does not require many enzymatic steps. It requires only a pair of sequence specific primers (Fig. 4).

The IPCR method involves digestion of genomic DNA with restriction enzymes that do not cut within the known region. Southern hybridisation is performed using the known region as a probe to determine the size of fragments obtained. The digested DNA is diluted and recircularized under conditions that favour the formation of monomeric circles (Collins and Weissman, 1984). The intramolecular mixture is used as a template in the PCR. The primers are designed in such a way that they are facing in opposite orientation to that of normal PCR, by inverting them to flank the known region of the gene fragment. The formation of monomeric circles is crucial for effectiveness of IPCR. Efficiency of monomeric circles formation is favoured whenever the ligation is carried at a very low concentration of DNA. At a high concentration of DNA the formation of concatemers have been observed (Collins and Weissman, 1984).

This method was applied in various studies. Its applications include the identification of the consensus sequences for insertion of transposable elements (Ochman *et al*., 1988; Earp *et al*., 1990; Li *et al*., 1999; Martin and Mohn, 1999), viral integration sites (Silver and Keerikatte, 1989), cloning of genes (Arand *et al*., 1999; Triglia *et al*., 1988) and screening of a YAC library for novel genes (Silverman *et al*., 1989). Even though IPCR has been used widely, it cannot be used to clone unknown genes since the primers must be designed based on known sequences. On the other hand, the end sequences of the

initial restriction fragment cannot be retrieved without going through some laborious primer walking process especially for long fragments. To overcome all these drawbacks of IPCR, Kohda and Taira, (2000) modified IPCR by introducing bridged inverse PCR (BI-PCR) (Fig. 4).

BI-PCR follows the same principle as IPCR, except that bridge DNA is included in the circularisation ligation step. Bridge DNA can be produced by digestion of a plasmid for example the polycloning sites. During ligation of digested DNA and bridge DNA, four different molecules can be generated: 1) No bridge DNA, but internal known sequences circularised, 2) No bridge DNA, no internal known sequences circularised, 3) Bridge DNA and internal known sequences circularised, and 4) Bridge DNA, but no internal known sequences circularised.

The first round of PCR is performed using gene specific primers (A and B) (Fig. 4). The PCR product is used as a template in the second round of PCR using gene specific primers (A and B) and bridge DNA specific primers (C and D respectively). In this case only the DNA molecules with a bridge DNA can be amplified. The BI-PCR has an advantage over the IPCR because it is easy to monitor the progress at each step. The efficacy of this method was tested by amplifying a region of RNA-dependent ATPase *hera* gene from *T. thermophilus* (Kohda and Taira, 2000).

Because IPCR has been proven as a useful method many researchers adapted it for cDNA synthesis (Towner and Gartner, 1992; Zeiner and Gehring, 1994; Huang *et al*., 1990; Zilberberg and Gurevitz, 1993) to obtain full-length cDNA. IPCR for cDNA synthesis involves the synthesis of a first strand of cDNA using an oligo(dT) or a gene specific primer. This is followed by the synthesis of second strand with the cDNA ends blunted by using T4 DNA polymerase. The blunt end cDNA is circularised to allow it to form intramolecular molecules.

This method was used to amplify the opsin genes from *Sphodromantis* spp. (Towner and Gartner, 1992), to clone 5’ and 3’ regions of human deoxycytidine kinase (dC kinase)

gene (Huang *et al*., 1990) and to clone the full-length cDNA of α-neurotoxins of scorpion *Leiurus quinquestriatus bebraeus* (Zilberberg and Gurevitz, 1993).

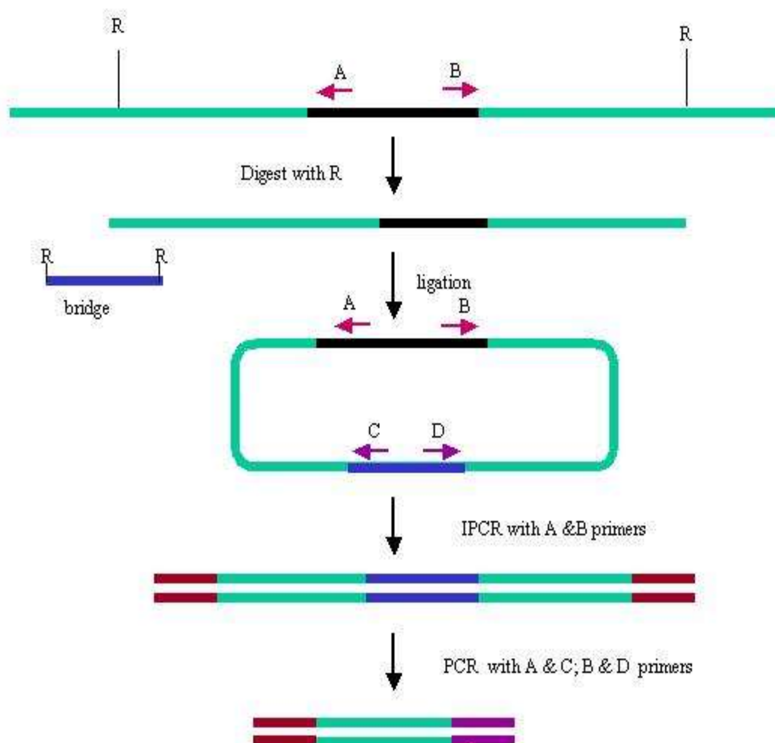


Figure 4. Schematic illustration of bridged inverse PCR (BI-PCR). The black area represents the known segment. The IPCR is performed with primers A and B. The upstream is amplified with A & C primers, while the downstream is amplified with B & D primers. R refers to restriction sites (Adapted from Kohda and Taira, 2000).

2.8 Panhandle PCR

Panhandle PCR, which can be used to amplify an unknown sequence flanking one side of a known sequence, generates a template shaped like a pan with a handle (Jones and Winistorfer, 1992; Felix *et al*., 1997).

The panhandle PCR method involves digestion of genomic DNA with restriction enzymes, followed by alkaline phosphatase treatment or Klenow fill-in. The digested DNA is ligated to a 5' phosphorylated single-stranded oligonucleotide with a 3' end complementary to the known single-stranded region ends of the digested genomic DNA (Fig. 5). The ligated mixture is denatured and allowed to reanneal resulting in the ligated synthetic oligo annealing to its complementary sequence in the genomic DNA. The strands of the genomic DNA that contain a complement of the ligated oligonucleotide form a stem-loop structure. This is followed by polymerase extension of the recessed 3' end. The polymerisation results in known DNA being appended to the end of unknown DNA contained in the loop, generating a panhandle structure. The resultant molecule is used as a template in PCR using primers based on the known sequence (primers 1 & 2) (Fig. 5). This is followed by second round of nested PCR using primers 3 and 4. The efficacy of this method was tested to amplify a region of β-globin DNA and the promoter of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene (Jones and Winistorfer, 1992). It has also been used for routine diagnosis of acute lymphoblastic leukemia to determine MLL genomic translocation breakpoints (Felix *et al*., 1997). This method has a wide range of applications such as determining the viral integration sites, amplification of fragments adjacent to cDNA such as regulatory regions, intron-exon junctions and generation of yeast artificial chromosome end points (YAC).

Due to the versatility of the panhandle PCR method, it was adapted to isolate the MLL fusion transcripts involving unknown partner genes (Megonigal *et al*., 2000). This modification involves the synthesis of cDNA using a primer containing MLL sequences at the 5' end and random hexamers at the 3' end. The second strand is generated and the double-stranded DNA is denatured and re-annealed to generate the stem-loop template followed by extension with *Taq* polymerase. Panhandle PCR is more effective when cDNA is used as template, since it does not require restriction enzymes and does not require the ligation step that could be problematic. This modification allows use of this method to investigate genes that are involved in alternative splicings (Megonigal *et al*., 2000).

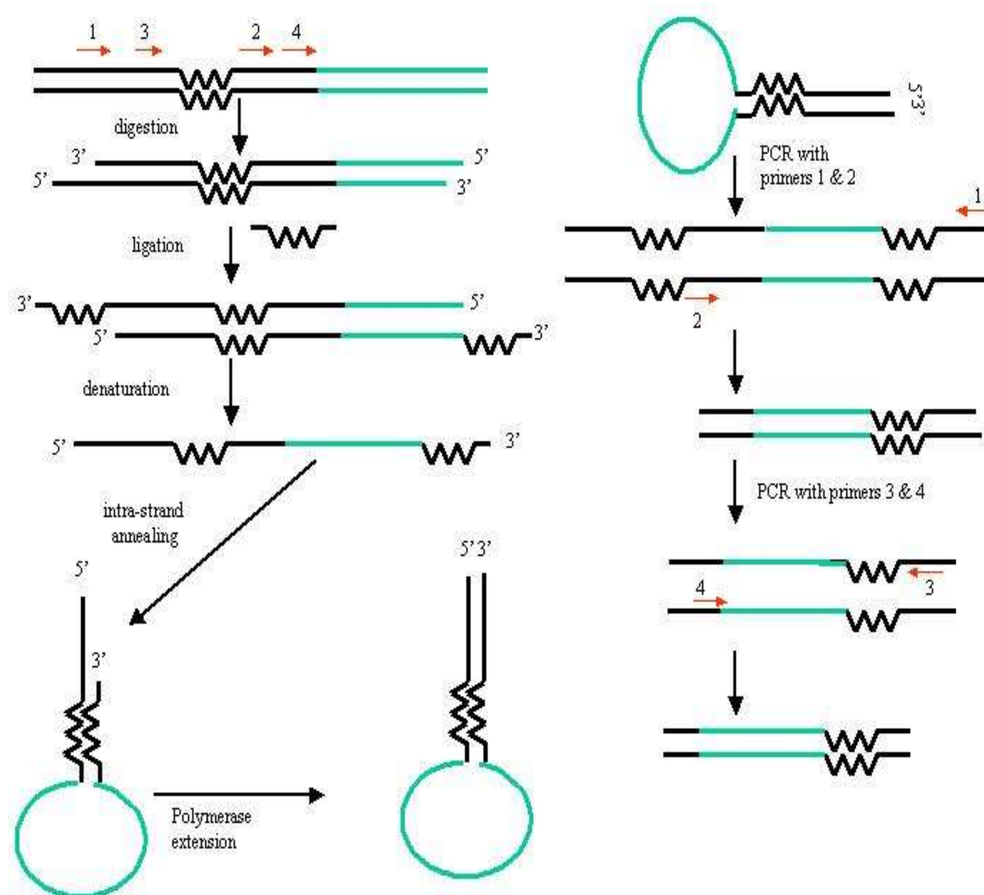


Figure 5. Schematic illustration of different steps that are involved in panhandle PCR. The jagged portion of the line represents the known fragment that is complementary to the ligated single- stranded linker. The PCR primers are shown with numbered arrows (adapted from Jones and Winistorfer, 1992)

2.9 Ligation-Mediated PCR

Ligation-mediated PCR involves a wide range of methods that can be used to clone unknown sequences that are flanking the known sequence of a gene. These methods require the ligation of adaptors or linkers to an unknown part of the sequence of digested genomic DNA or cDNA. The adaptors can be single-stranded (Troutt *et al*., 1992**;** Edwards *et al*., 1991; Fromont-Racine *et al*., 1993; Bertrand *et al*., 1993), double-stranded (Mueller and Wold, 1989; Collasius *et al*., 1991; Willems *et al*., 1998; Shyamala and Ames, 1989; Kalman *et al*., 1990; Kilstrup and Kristiansen, 2000), vectorette or bubble (Riley *et al*., 1990) or splinkerette (Devon *et al*., 1995). After ligation of adaptors

to DNA the PCR is performed using adaptor and gene specific primers. The adaptor specific primer is designed in such a way that it does not participate in amplification in the first cycle of PCR. Thus, it is only used for amplification in the second cycle after the target DNA has been extended to its end by a gene specific primer. These methods have been used to isolate promoters (Fors *et al*., 1990), determine insertional sites for transposons (Prod'hom *et al*., 1998; Collasius *et al*., 1991; Shyamala and Ames, 1989) and cloning of genes (Kilstrup and Kristiansen, 2000).

2.9.1 Biotin capture PCR

Even though ligation-mediated PCR methods have been used widely, they sometimes do not yield reliable results, especially when dealing with complex genomes. 'End repair priming' or 'filling in' of the recessed ends have been observed with these methods. These side reactions lead to non-specific amplifications that make it difficult to identify the correct product within the background (Rosenthal, 1992). To circumvent these problems, one of the gene specific primers is biotinylated and is used in a linear PCR mediated amplification (Rosenthal *et al*., 1991; Rosenthal and Jones, 1990; Rosenthal, 1992) (Fig. 6). The biotinylated PCR fragments are isolated from the complex genomic mixture using magnetic beads coated with streptavidin and then purified. The anchored single-stranded template is subsequently exponential amplified using the gene specific primer and the adaptor specific primer. Nested PCR is also performed to ensure that only the desired fragment will be further amplified. The isolation of specific biotinylated fragments is very important to ensure high specificity in the subsequent steps. It reduces the complexity of the mixture, especially when amplification is being performed in complex genomes. This method was used to amplify the nematode *unc*31 gene contained on YAC (Rosenthal *et al*., 1990) and to clone the human gene for adhesion molecule L1 (CAM-L1) (Rosenthal *et al*., 1991).

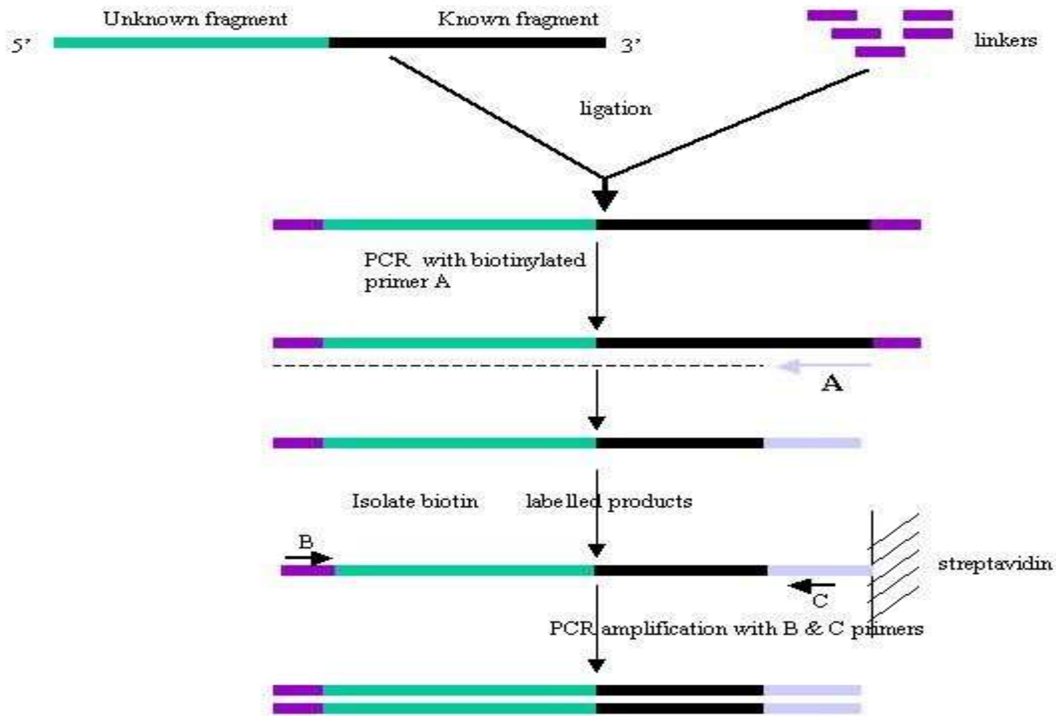


Figure 6. Schematic illustration of Capture biotin PCR. Primer A is biotinylated. Primers B and C are used for exponential amplification (adapted from Rosenthal, 1992).

2.9.2 Vectorette PCR

Due to the difficulties being experienced with other ligation-mediated PCR methods, the vectorette (bubble) PCR has been devised (Riley *et al*., 1990) and is more widely used than other methods. Vectorette units consist of double-stranded linker sequences with central mismatches and cohensive ends that are suitable for ligation to DNA fragments produced by digestion with restriction enzymes (Fig. 7). They are phosphorylated at their 5' ends. Vectorette PCR involves the digestion of genomic DNA with restriction enzymes that generate sticky overhangs. The vectorette units are ligated to the digested DNA and a vectorette PCR is performed using the vectorette and gene specific primers. The vectorette specific primer (VSP) has the identical sequence as the top strand, therefore the amplification will not start until its complementary sequence has been synthesized from the specific target DNA primer. Vectorette PCR was used to determine

wheat telomere-associated sequences (TASS) (Mao *et al*., 1997) and the exon-intron junction of the human dystrophin gene (Roberts *et al*., 1992).

However, the specificity of the vectorette PCR is limited. Vectorette PCR can generate 'end-repair priming' which involves the cohensive ends of the unligated vectorettes and the DNA that is generated by restriction enzymes that produce 5' overhangs. The overhangs are filled in during the first cycle of PCR and during the next step of the PCR reaction those ends are able to anneal to each other with sufficient stability to initiate priming. This results in a sequence complementary to the vectorette primer, allowing amplification without participation of the gene specific primer.

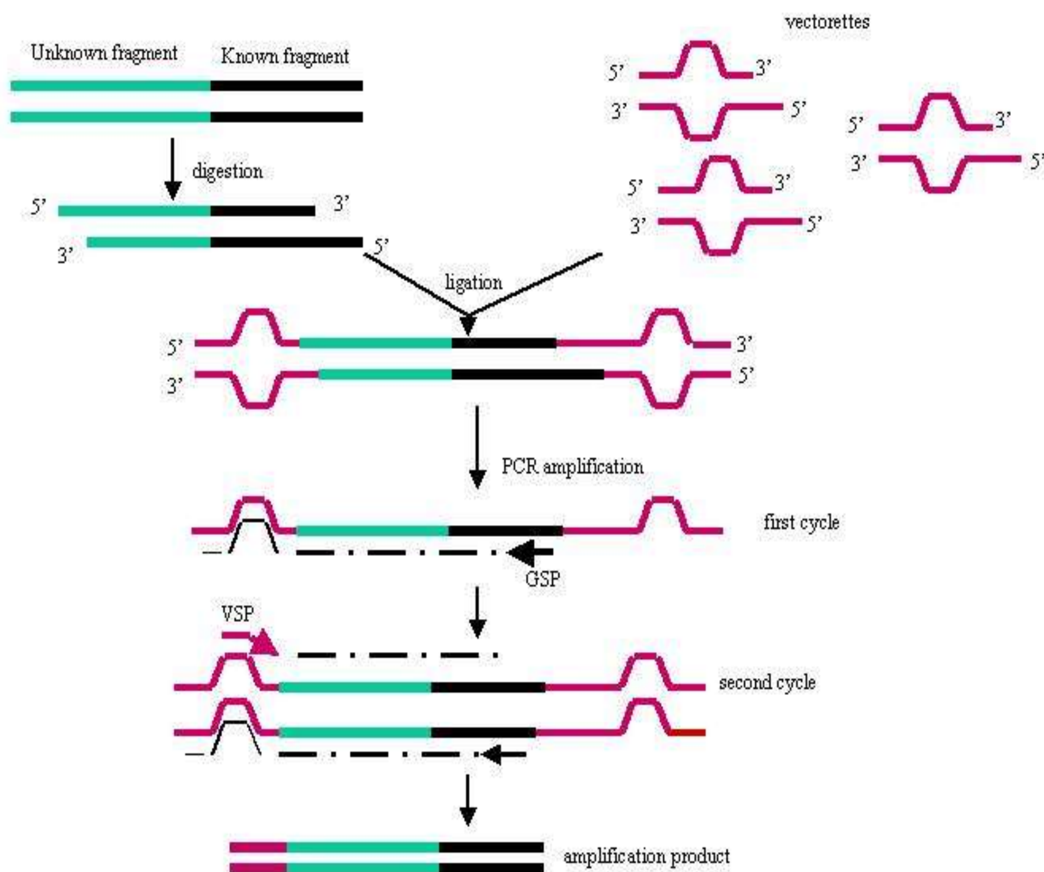


Figure 7. Schematic illustration of vectorette PCR. The vectorette units are ligated with digested genomic DNA and PCR amplification is performed with GSP (gene specific primer) and VSP (vectorette specific primer).

2.9.3 Splinkerette PCR

Due to problems associated with vectorette PCR, the splinkerette PCR technique has been designed to improve the efficiency of the vectorette PCR (Devon *et al*., 1995). The splinkerette unit contains a 'hairpin' structure on the bottom strand (Fig. 8). On the other end the splinkerettes are not kinased so that there is no covalent bonding between the splinkerettes bottom strands and the DNA to be amplified. As in the case of vectorette PCR the splinkerette specific primer is the same sequence as the top strand and thus, is unable to act as a primer until the complementary sequence has been synthesized. During PCR the free 3' ends of the bottom strands of splinkerettes flip back on themselves to form hairpin structures and begin elongation further along the bottom strands. This hairpin structures are eliminated from the reaction. Therefore, there is no chance of the 'end-repair priming' or 'filling in'. It has been observed that the splinkerette PCR is able to amplify long fragments that vectorette PCR cannot handle (Devon *et al.,* 1995). This gave the splinkerette PCR an advantage to isolate longer end fragments from a YAC library (Devon *et al.,* 1995).

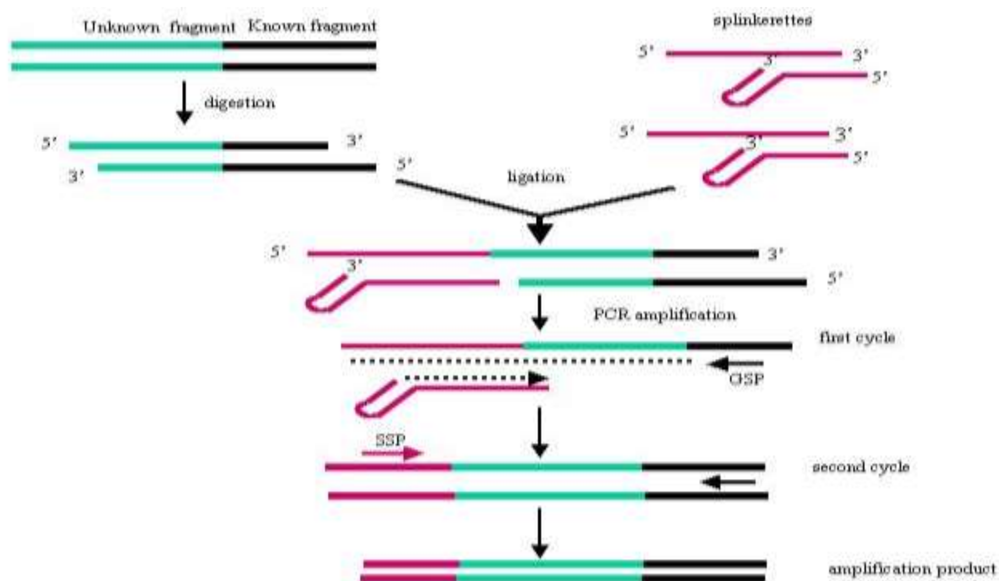


Figure 8. Schematic illustration of splinkerette PCR. The splinkerette units are ligated with digested genomic DNA and PCR amplification is performed with GSP (gene specific primer) and SSP (splinkerette specific primer).

2.9.4 Ligation –Anchored PCR (LA-PCR)

The ligation –anchored PCR (LA-PCR) combines the RACE technique (Frohman *et al.,* 1988) with ligation-mediated PCR (Mueller and Wold, 1989). However this method does not involve the homopolymeric tailing of the cDNA. The RACE technique has a potential to generate non-specific products due to the use of a homopolymer–containing primer in the PCR. Thus ligation–anchored PCR (LA-PCR) is an alternative method to RACE in which a single-stranded anchor of defined sequence is directly ligated to the 3' end of cDNA (Troutt *et al*., 1992**;** Edwards *et al*., 1991). This method involves the synthesis of cDNA by using an oligo(dT) or a gene specific primer. The cDNA is ligated by T4 RNA ligase to an anchor that is phosphorylated at the 5' end and blocked at the 3' end. The 3'end of the anchor is blocked by addition of dideoxynucleotide leaving only the 5'-phosphorylated terminal as a potential substrate for ligation. The ligation mixture is subjected to PCR amplification using a cDNA specific primer and an anchor specific primer. The LA-PCR's success depends on the T4 RNA ligase to attach the anchor to the 3' end of the cDNA. Since the 3' end of the anchor is blocked only the 5'-phoshorylated end of the anchor is available for ligation to the 3' end of the cDNA.

This method was used to amplify the IgG1 cDNA (Troutt *et al*., 1992) and to analyse the 5' end of the rat tryptophan hydroxylase (TPH) mRNA that is expressed at a very low level (Delort *et al*., 1989; Edwards *et al*., 1991).

2.9.5 Reverse ligation –mediated PCR (RLM-PCR)

The ligation-mediated PCR methods (Mueller and Wold 1989) are not limited to genomic DNA amplification, but have also been adapted for cDNA synthesis (Fromont-Racine *et al*., 1993; Bertrand *et al*., 1993). This modification is referred to as reverse ligation-mediated PCR (RLM-PCR). RLM-PCR involves the removal of the 5' cap structure from mRNA by tobacco acid pyrophosphatase (TAP), followed by ligation of an RNA linker of a known sequence to the unknown 5' end of mRNA. The reverse transcription of the desired mRNA using a gene specific primer is performed. Then the PCR is performed with a gene specific primer and a linker specific primer. This method was used to map

the 5' end of the expressed chloramphenicol acetyltransferase (CAT) mRNA (Fromont-Racine *et al*., 1993). It was also used to detect *in vivo* the iron–depletion–dependent footprints on two iron-responsive elements (IRE) of the human transferrin receptor (TfR) mRNA (Bertrand *et al*., 1993). Thus, the versatility of this method can be used to study the regulation of gene expression that takes place at RNA level.

2.10 Suppression PCR

Due to the non-specific amplifications being experienced with ligation-mediated PCR methods, suppression PCR has been designed to improve these methods (Siebert *et al*., 1995). Suppression PCR applies the same principles as ligation-mediated PCR (Mueller and Wold, 1989) except for a few modifications. The adaptor with one end blunt is ligated to the digested DNA and the PCR amplification is performed using an adaptor specific (ASP) primer and a gene specific primer (GSP) (Fig. 9). The adaptor has an amine group on the 3' end of the lower strand. This amine group blocks any extension of the lower adaptor strand, unless a gene specific primer extends a DNA strand opposite the upper strand of the adaptor. The adaptor specific primer is shorter in length than the adaptor. If non-specific amplification generates PCR products that contain the double stranded adaptor sequences at both ends; the ends of individual DNA strands form 'pan'-like structures following every denaturation step. This is possible due to the presence of inverted repeats at each end of the strands (Fig. 9). In addition these structures are more stable than the primer-template hybrid, therefore they suppress exponential amplifications of undesired fragments. However, when the gene specific primer extends through the adaptor, the products contain an adaptor sequence only at one end. Thus, it cannot form the 'pan'-like structure that can suppress the PCR and the desired products are amplified exponentially. This method was used to amplify the upstream part of exon 1 of the human tissue-type plasminogen activator gene (Siebert *et al*., 1995). This method has a great potential to be applied to isolate the flanking regions required for techniques such as RFLP (restriction fragment length polymorphism), RAPD (random amplification of polymorphic DNA) and EST (expressed sequence tags) (Schupp *et al*., 1999).

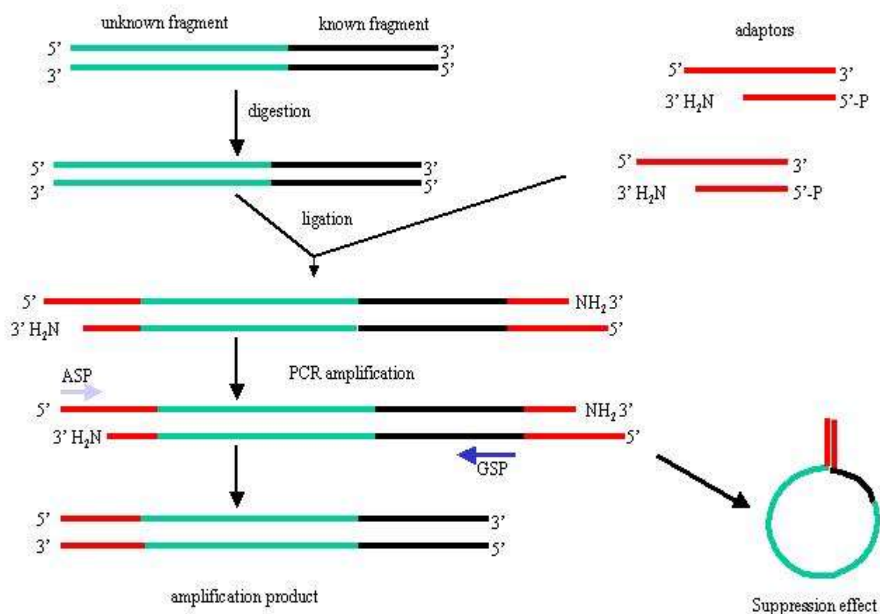


Figure 9. Schematic illustration of suppression PCR. The adaptors are ligated with digested genomic DNA and PCR amplification is performed with GSP (gene specific primer) and ASP (adaptor specific primer). The products flanked by adaptor primers are eliminated by suppression effect (adapted from Siebert *et al*., 1995).

2.11 Chase PCR

Chase PCR is a modification of suppression PCR (Siebert *et al*., 1995) that has been adapted for cDNA amplification (Lukyanov *et al*., 1995, Timblin *et al*., 1990). The method is based on the insertion of inverted terminal repeats into the amplified cDNA that permits short molecules to form 'pan'- like structures at each PCR cycle and thus escapes the annealing with primers (Fig. 10). The method involves cDNA synthesis with a oligo(dT) primer that contains flanking sequences attached to it. The cDNA is tailed with a dGTP (or dATP). The PCR amplification is performed with oligo(dT)-anchor and oligo(dC)-anchor primers. The oligo(dC)-anchor primer contains flanking sequences identical to that one of the oligo(dT)-anchor primer. The presence of inverted terminal repeats on both sides of the amplified cDNA allows cDNA strands to form 'pan'-like

structures at each cycle of PCR due to the annealing of the 3' and 5' ends of the same molecule. These 'pan'-like structures prevent effective hybridisation of primers and reduce the amplification rate as well. It has been shown that the inhibition is more effective for shorter molecules because the formation of 'pan'-like structures is faster due to shorter distances between 3' and 5' ends (Lukyanov *et al*., 1995; Timblin *et al*., 1990).

The concentration of the primers also influences the formation of the 'pan'-like structures. The lower the concentration of primers the higher the probability that the shorter molecules will form 'pan'-like structures. Thus, this allows the amplification of longer molecules and eliminates the formation of primer dimers. This method is applicable for generation of cDNA libraries that contain full-length cDNAs (Lukyanov *et al*., 1995; Timblin *et al*., 1990).

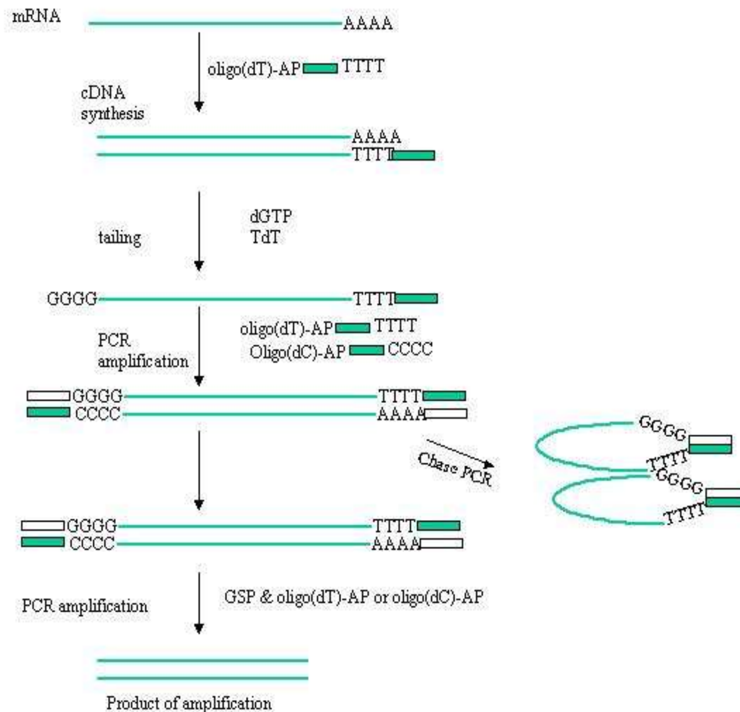


Figure 10. Schematic illustration of chase PCR. The cDNA is synthesised with oligo(dT)-AP (oligo(dT) anchor primer) and G- tailed with TdT (terminal deoxynucleotidyl- transferase). PCR amplification is performed with oligo(dC)-AP and oligo(dT)-AP. The short fragments are eliminated by chase PCR effect (adapted from Lukyanov *et al.,* 1995).

2.12 Random PCR

Many PCR based methods that are used to amplify the unknown regions of genes involve numerous enzymatic steps that could produce a number of artefacts. Many researchers found alternative methods to amplify the unknown fragments without using the methods that involve enzymatic steps. These methods are termed random PCR (Trueba and Johnson, 1996; Parker *et al*., 1991), unpredictably PCR (Dominguez and Lopez-Larrea, 1994) and low stringency or non-specific PCR (Parks *et al*., 1991; Malo *et al*., 1994).

Even though these methods sound complicated they apply the same principles involving the synthesis of single stranded DNA using either a gene specific primer or a non-specific primer. The PCR is performed using a gene specific primer that is used in the first reaction together with non-specific primer at low stringency. The second round of nested PCR is performed using a gene specific primer and a non-specific primer at high stringency.

The non-specific primers are oligonucleotides of defined and artificial sequences. These primers are designed to bind anywhere for the first hybridisation and bind selectively in the subsequent PCR annealings. The first non-specific primer is longer than the second non-specific primer and their sequences have an overlap. It has been observed that as far as there is a partial homology at the 3' end of the non-specific primer and the correct pairing of the last three nucleotides at the 3' end, the amplification is possible (Parker *et al*., 1991; Sommer *et al*., 1989). Thus, this method can be applied to search for polymorphic and endonuclease sites (Parker *et al*., 1991). This method was used to determine the $V\alpha$ and $J\alpha$ sequences of human T cell receptors (Struck and Collins, 1994).

2.13 Non-specific primed, Nested suppression PCR (NSPS-PCR)

In the case of non-specific or random PCR one primer is used that is specific for the known region while the other binds non-specifically on an uncharacterised region (Parks *et al*., 1991). However this method can generate a number of undesirable short PCR products. For suppression PCR, a single primer is used to amplify DNA fragments that

are flanked by terminal inverted repeats (Siebert *et al.,* 1995). Thus, it discourages the formation of short PCR fragments. When the PCR fragments are short the inverted repeats have a tendency to block the primer binding and DNA synthesis. Tamme *et al*., (2000) developed a method to overcome this problem by combining the random or non-specific PCR with suppression PCR.

This method involves non-specific primed PCR where a PCR amplification is performed using only one primer, which is based on the known sequence. The PCR is performed under low stringency using a polymerase such as *Taq* that lacks 3'exonuclease activity. When the non-specific priming creates a specific sequence within the synthesized strands with inverse terminal repeats at the ends the DNA can be amplified exponentially.

In the nested suppression PCR the non-specific PCR product is reamplified using a nested primer. The nested primer is identical to the primer used in the non-specific PCR, except the nested primer has been extended by 6 bases at the 3' end. In this case a polymerase with exonuclease activity such as *Pfu* is used. In the first few cycles the DNA is synthesised only from the complementary sites in the known sequence. During those cycles the polymerase with its exonuclease activity has truncated the complementary part of the primer at the 3' end so that it is completely complementary to the sequence at the opposite end of the fragment. Thus, the exponential amplification can occur. This method was used to clone the promoter of the tyrosine gene of the zebrafish (Tamme *et al*., 2000).

NSPS-PCR has been adapted further for cDNA synthesis. The cDNA is synthesised with a gene specific primer biotinylated at 5' end. The desired cDNA is purified using the streptavidin–coated magnetic particles. The NSPS-PCR is performed using the purified cDNA as a template. To test the efficacy of this method on cDNA amplification, it was used to clone a *Notch* gene of *Branchiostoma floridae* which is expressed at a very low level (Tamme *et al*., 2000).

2.14 Conclusions

The invention of PCR has revoluntionalised molecular biology research. The application of PCR was initially limited due to its inability to amplify unknown sequence. However, many clever adaptations have been developed to overcome this limitation. Thus PCR is probably the most important tool for biotechnological research. PCR based methods have in recent years replaced the use of genomic or cDNA libraries for the discovery of new genes. The creation of genomic libraries using conventional methods involves digestion of genomic DNA with restriction enzymes and ligation into for example bacteriophage λ vectors. The λ vectors are packaged into phage particles that infect *E. coli* cells. The creation of cDNA libraries involves the synthesis of cDNA which is ligated into bacteriophage λ vectors. The screening of these libraries is also laborious and frustrating. It can take months or even years to find a gene of interest. With PCR based methods the results can be obtained within a few hours or days.

With the study of eukaryotic organisms the presence of introns in coding regions necessitated the development of methods that allow the isolation of genes using mRNA as initial template. For example, the 5'/3 'RACE technique has been devised to amplify the unknown 5' and 3' end sequences flanking known cDNA sequences. The modification of this method led to the discovery of template switching activity of reverse transcriptase that has unravelled the understanding of eukaryotic gene expression (Schmidt and Mueller, 1999). However, many problems such as generation of cDNA fragments with truncated 5' ends have been encountered with cDNA based methods. To solve these problems additional methods were developed such as the Cap-finder method. This method allows the isolation of intact 5' ends of mRNA taking advantage of a template-switching effect. Since the reverse transcriptase adds some cytosine residues upon reaching the 5' end of mRNA, the primer with oligo(rG) can be included in the cDNA synthesis reaction. The reverse transcriptase switches templates and continues with cDNA synthesis using the oligo(rG) primer. Upon finishing the cDNA synthesis there is a known sequence at the 3' end of the cDNA. PCR can be performed using the oligo(rG) specific primer and a gene specific primer. Another method that is useful to

obtain the intact 5' end of mRNA is ligation-anchored PCR (LA-PCR). A single-stranded anchor of defined sequence is directly ligated to the 3' end of the cDNA and PCR is performed using an anchor specific primer and a gene specific primer.

With the use of genomic DNA as starting material various problems can be encountered. Sometimes the genomic DNA to be amplified can be too long for amplification. Different methods have been developed to circumvent this problem. The DNA is digested with restriction enzymes and linkers or vectorettes are ligated onto the digested DNA. PCR is performed using the linker specific primer and a gene specific primer.

Digested genomic DNA can also be used to clone unknown sequences using inverse PCR. In this case the DNA is digested with one enzyme and then allowed to re-ligate on itself. The primers for PCR are designed in such way that they are facing in opposite orientation so that they can amplify the unknown sequence. However, there are critical steps that need to be treated with caution. The efficiency of forming intramolecular circles depends on the concentration of the DNA. At high concentration of DNA the concatamers are more favoured than the intramolecular circles. The efficiency of PCR amplification also depends on the size of the circles. The restriction enzymes should be chosen in such a way that they generate the fragments of reasonable size for amplification.

Digested genomic DNA can also be used for panhandle PCR. In the panhandle PCR the known DNA sequence is positioned on both sides of the unknown flanking sequence that results in amplification of unknown sequences.

Some methods give undesirable results due to non-specific amplifications. Methods such as suppression PCR, chase PCR and step-out PCR have been developed to eliminate background amplification. The primers or adaptors are designed with inverted terminal repeats. During amplification the fragments that are flanked by inverted repeats form 'pan'-like structures that are very stable. The shorter fragments form 'pan'-like structures faster than the longer fragments, because the distance between the 3' and 5' ends is

shorter. The concentrations of primers also play a role in formation of the ‘pan’-like structures. The lower the concentration of primers the higher the probability for the shorter molecules to form ‘pan’-like structures and this favours the amplification of longer fragments. These methods help to eliminate the primer dimers from the PCR amplification.

The use of biotinylated primers in linear PCR can also reduce high background. The biotinylated fragments are isolated from the complex genomic DNA by using magnetic beads coated with streptavidin. The purified biotinylated fragments are used in subsequent PCR. To ensure that only the desired fragments are amplified, nested primers are designed to increase the PCR specificity.

Even though many methods look simple, they are very difficult to execute and to obtain the desired results. Non-specific amplifications and the introduction of errors in DNA sequences are often observed. This led some researchers to discover the “desperate methods” such as random or non-specific PCR.

PCR based methods are not only used for the direct cloning of novel genes but are also applied for identification of the consensus sequences for insertion of transposable elements, viral integration sites and screening of YAC libraries for novel genes. With the complete human genome sequenced, these methods can also be used as diagnostic tools for genetic linked diseases.

2.15 References

Arand, M., Hemmer, H., Durk, H., Baratti, J., Archelas, A., Furstoss, R., and Oesch, F. (1999). Cloning and molecular characterization of a soluble epoxide hydrolase from *Aspergillus niger* that is related to mammalian microsomal epoxide hydrolase. *Biochem J*. **344**: 273-80.

Berchtold, MW. (1989). A simple method for direct cloning and sequencing cDNA by the use of a single specific oligonucleotide and oligo(dT) in a polymerase chain reaction (PCR). *Nucleic Acids Res*. **17**: 453.

Bertrand, E., Fromont-Racine, M., Pictet, R. and Grange, T. (1993). Visualization of the interaction of a regulatory protein with RNA *in vivo*. *Proc Natl Acad Sci U S A*. **90**: 3496-3500.

Bloomquist, BT., Johnson, RC. and Mains, RE. (1992). Rapid isolation of flanking genomic DNA using biotin-RAGE, a variation of single-sided polymerase chain reaction. *DNA Cell Biol*. **11**: 791-797.

Collasius, M., Puchta, H., Schlenker, S. and Valet, G. (1991). Analysis of unknown DNA sequences by polymerase chain reaction (PCR) using a single specific primer and a standardized adaptor. *J Virol Methods*. **32**: 115-119.

Collins, FS. and Weissman, SM. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularisation method. *Proc Natl Acad Sci U S A*. **81**: 6812-6816.

Cormack, RS. and Somssich, IE. (1997). Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. *Gene*. **194**: 273-276.

Delort, J., Dumas. JB., Darmon, MC. and Mallet, J. (1989). An efficient strategy for cloning 5' extremities of rare transcripts permits isolation of multiple 5'-

untranslated regions of rat tryptophan hydroxylase mRNA. *Nucleic Acids Res*. **17**: 6439-6448.

Devon, RS., Porteous, DJ. and Brookes, A.J. (1995). Splinkerettes-improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res*. **23**: 1644-1645.

Dominguez, O. and Lopez-Larrea, C. (1994). Gene walking by unpredictably primed PCR. *Nucleic Acids Res*. **22**: 3247-3248.

Earp, DJ., Lowe, B. and Baker, B. (1990). Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. *Nucleic Acids Res*. **18**: 3271-3279.

Edwards, JB., Delort, J. and Mallet, J. (1991). Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5' ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. *Nucleic Acids Res*. **19**: 5227-5232.

Felix, CA., Kim, CS., Megonigal, MD., Slater, DJ., Jones, DH., Spinner, NB., Stump, T., Hosler, MR., Nowell, PC., Lange, BJ. and Rappaport, EF. (1997). Panhandle polymerase chain reaction amplifies MLL genomic translocation breakpoint involving unknown partner gene. *Blood*. **90**: 4679-4686.

Flouriot, G., Brand, H. and Gannon, F. (1999). Identification of differentially expressed 5'-end mRNA variants by an improved RACE technique (PEETA). *Nucleic Acids Res*. **27**: e8.

Fors, L., Saavedra, RA. and Hood, L. (1990). Cloning of the shark Po promoter using a genomic walking technique based on the polymerase chain reaction. *Nucleic Acids Res*. **18**: 2793-2799.

Frohman, MA., Dush, MK. and Martin, GR. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*. **85**: 8998-9002.

Fromont-Racine, M., Bertrand, E., Pictet, R. and Grange, T. (1993). A highly sensitive method for mapping the 5' termini of mRNAs. *Nucleic Acids Res*. **21**: 1683-1684.

Harvey, RJ. and Darlison, MG. (1991). Random-primed cDNA synthesis facilitates the isolation of multiple 5'-cDNA ends by RACE. *Nucleic Acids Res*. **19**: 4002.

Huang, SH., Hu, YY., Wu, CH. and Holcenberg, J. (1990). A simple method for direct cloning cDNA sequence that flanks a region of known sequence from total RNA by applying the inverse polymerase chain reaction. *Nucleic Acids Res*. **18**: 1922.

Jones, DH. and Winistorfer, SC. (1992). Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucleic Acids Res*. **20**: 595-600.

Kalman, M., Kalman, ET. and Cashel, M. (1990). Polymerase chain reaction (PCR) amplification with a single specific primer. *Biochem Biophys Res Commun*. **167**: 504-506.

Kilstrup, M. and Kristiansen, KN. (2000). Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. *Nucleic Acids Res*. **28**: e55.

Kohda, T. and Taira, K. (2000). A simple and efficient method to determine the terminal sequences of restriction fragments containing known sequences. *DNA Res*. **7**: 151-155.

Kriangkum, J., Vainshtein, I. and Elliott, JF. (1992). A reliable method for amplifying cDNA using the anchored-polymerase chain reaction (A-PCR). Nucleic Acids Res. **20**: 3793-3794.

Li, ZH., Liu, DP. and Liang, CC. (1999). Modified inverse PCR method for cloning the flanking sequences from human cell pools. *Biotechniques* **27**: 660-662.

Loh, EY., Elliott, JF., Cwirla, S., Lanier, LL. and Davis, MM. (1989). Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain. *Science*. **243**: 217-220.

Lukyanov, KA., Launer, GA., Tarabykin, VS., Zaraisky, AG. and Lukyanov, SA. (1995). Inverted terminal repeats permit the average length of amplified DNA fragments to be regulated during preparation of cDNA libraries by polymerase chain reaction. *Anal Biochem*. **229**: 198-202.

Malo, MS., Srivastava, K., Andresen, JM., Chen, XN., Korenberg, JR. and Ingram, VM. (1994). Targeted gene walking by low stringency polymerase chain reaction: assignment of a putative human brain sodium channel gene (*SCN3A*) to chromosome 2q24-31.*Proc Natl Acad Sci U S A*. **91**: 2975-2979.

Mao, L., Devos, K.M., Zhu, L. and Gale, MD. (1997). Cloning and genetic mapping of wheat telomere-associated sequences.*Mol Gen Genet*. **254**: 584-591.

Martin VJ. and Mohn, WW. (1999). An alternative inverse PCR (IPCR) method to amplify DNA sequences flanking Tn5 transposon insertions. *J Microbiol Methods*. **35**: 163-166.

Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A. (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res*. **27**: 1558-1560.

Megonigal, MD., Rappaport, E.F., Wilson, RB., Jones, DH., Whitlock, JA., Ortega, JA., Slater, DJ., Nowell, PC. and Felix, CA. (2000). Panhandle PCR for cDNA: a rapid method for isolation of MLL fusion transcripts involving unknown partner genes. *Proc Natl Acad Sci U S A*. **97**: 9597-9602.

Mueller, PR. and Wold, B. (1989). *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science*. **246**: 780-786.

Ochman, H., Gerber, AS. and Hartl, DL. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics*. **120**: 621-623.

Ohara, O., Dorit, RL. and Gilbert, W. (1989). One-sided polymerase chain reaction: the amplification of cDNA. *Proc Natl Acad Sci U S A*. **86**: 5673-5677.

Parker, JD., Rabinovitch, PS. and Burmer, GC. (1991). Targeted gene walking polymerase chain reaction. *Nucleic Acids Res*. **19**: 3055-3060.

Parks, CL., Chang, LS. and Shenk, T. (1991). A polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor protein coding region. *Nucleic Acids Res*. **19**: 7155-7160.

Prod'hom, G., Lagier, B., Pelicic, V., Hance, AJ., Gicquel, B. and Guilhot, C. (1998). A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences. *FEMS Microbiol Lett*. **158**: 75-81.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, JC. and Markham, AF. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res*. **18**: 2887-2890.

Roberts, RG., Coffey, AJ., Bobrow, M. and Bentley, DR. (1992). Determination of the exon structure of the distal portion of the dystrophin gene by vectorette PCR. *Genomics*. **13**: 942-950.

Rosenthal A. (1992). PCR amplification techniques for chromosome walking. *Trends Biotechnol*. **10**: 44-48.

Rosenthal, A. and Jones, DS. (1990). Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res*. **18**: 3095-3096.

Rosenthal, A., MacKinnon, RN. and Jones, DS. (1991). PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1). *Nucleic Acids Res*. **19**: 5395-53401.

Saiki, RK., Gelfand, DH., Stoffel, S., Scharf, SJ., Higuchi, R., Horn, GT., Mullis, KB. and Erlich, HA. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. **239**: 487-491.

Schaefer, BC. (1995). Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal Biochem*. **227**: 255-273.

Schmidt WM. and Mueller, MW. (1999). CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res*. **27**: e31.

Schramm, G., Bruchhaus, I. and Roeder, T. (2000). A simple and reliable 5'-RACE approach. *Nucleic Acids Res*. **28**: e96.

Schupp, JM., Price, LB., Klevytska, A. and Keim, P. (1999). Internal and flanking sequence from AFLP fragments using ligation-mediated suppression PCR. *Biotechniques*. **26**: 905-910, 912.

Shyamala, V. and Ames, GF. (1989). Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene*. **84**: 1-8.

Siebert, PD., Chenchik, A., Kellogg, DE., Lukyanov, KA. and Lukyanov, SA. (1995). An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res*. **23**: 1087-1088.

Silver, J. and Keerikatte, V. (1989). Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J Virol*. **63**: 1924-1928.

Silverman, GA., Ye, RD., Pollock, KM., Sadler, JE. and Korsmeyer, SJ. (1989). Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc Natl Acad Sci U S A*. **86**: 7485-7489.

Sommer, R. and Tautz, D. (1989). Minimal homology requirements for PCR primers. *Nucleic Acids Res*. **17**: 6749.

Struck, F. and Collins, J. (1994). Simple and rapid 5' and 3' extension techniques in RT-PCR. *Nucleic Acids Res*. **22**: 1923-1924.

Tamme, R., Camp, E., Kortschak, RD. and Lardelli, M. (2000). Nonspecific, nested suppression PCR method for isolation of unknown flanking DNA. *Biotechniques*. **28**: 895-899, 902.

Timblin, C., Battey, J. and Kuehl, WM. (1990). Application for PCR technology to subtractive cDNA cloning: identification of genes expressed specifically in murine plasmacytoma cells. *Nucleic Acids Res*. **18**: 1587-1593.

Towner, P. and Gartner, W. (1992). cDNA cloning of 5' terminal regions. *Nucleic Acids Res*. **20**: 4669-4670.

Triglia, T., Peterson, MG. and Kemp, DJ. (1988). A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res*. **16**: 8186.

Troutt, AB., McHeyzer-Williams, MG., Pulendran, B. and Nossal, GJ. (1992). Ligation-anchored PCR: a simple amplification technique with single-sided specificity. *Proc Natl Acad Sci U S A*. **89**: 9823-9825.

Trueba, GA. and Johnson, RC. (1996). Random primed gene walking PCR: a simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences. *Biotechniques*. **21**: 20.

Willems, H. (1998). Adaptor PCR for the specific amplification of unknown DNA fragments. *Biotechniques*. **24**: 26, 28.

Zeiner, M. and Gehring, U. (1994). Cloning of 5' cDNA regions by inverse PCR. *Biotechniques*. **17**: 1050-1052, 1054.

Zilberberg, N. and Gurevitz, M. (1993). Rapid isolation of full length cDNA clones by "inverse PCR:" purification of a scorpion cDNA family encoding alpha-neurotoxins. *Anal Biochem*. **209**: 203-205.

CHAPTER 3

MOLECULAR CLONING AND CHARACTERIZATION OF A NOVEL CYTOCHROME P450 GENE OF *Rhodotorula* SP. CBS 8446

3.1 Introduction

Cytochrome P450s are heme-containing monooxygenases that are widely distributed in nature being found in bacteria, archae, yeasts, plants and animals (Käppeli, 1986; Nelson, 1999; Ohkuma *et al.,* 1991a; Mauersberger *et al.*, 1996). Bacterial cytochrome P450s are found in soluble form localized in cytosolic fractions. On the other hand, the eukaryotic P450s are anchored in membranes of the endoplasmic reticulum (ER) by a hydrophobic N-terminal (Black, 1992; Sanglard *et al*., 1989; Scheller *et al*., 1994; Menzel *et al*., 1996; Nelson and Strobel, 1988). Cytochrome P450s require molecular oxygen (O_2) and transfer of electrons from NADPH via flavin adenine dinucleotide (FAD) by a flavoprotein NADPH oxidoreductase (Capdevila *et al*., 1984; Sutter *et al*., 1990; Vogel *et al.*, 1992; Kargel *et al.*, 1996; Gilewicz *et al.*, 1979).

In mammals the P450 system has been observed in multiple forms that play an important role in metabolic functions such as steroid synthesis and detoxification of xenobiotics and drugs (Sanglard *et al.*, 1989; Mathews and van Holde, 1990; Gonzalez, 1988; 1990; Käppeli, 1986, Nebert *et al.*, 991). In yeasts cytochrome P450s are involved in the metabolism of *n*-alkanes and fatty acids. Two different enzymatic steps have been observed in yeasts: the mono-terminal hydroxylation of alkanes that is regarded as a first and rate–limiting step in alkane degradation as well as ω-hydroxylation of fatty acids (Eschenfedt *et al*., 2003; Craft *et al*., 2003; Scheller *et al.,* 1998; Fukui and Tanaka, 1981; Schunck *et al*., 1987, Blasig, 1988; Zimmer *et al*., 1996). The *n*-alkanes are hydroxylated to fatty alcohols that are subsequently oxidized to fatty aldehydes and fatty acids. The

hydroxylation of fatty acids at the omega position forms the dioic acids, which are useful chemical products (Eschenfedt *et al*., 2003; Craft *et al*., 2003; Picataggio *et al*., 1992, Scheller *et al.,* 1998).

The alkane hydroxylating P450s from alkane utilizing ascomycetous yeasts are genetically well characterised. They have been classified in the CYP52 family and multiple genes are present in most yeasts. There are eight genes in *Yarrowia lipolytica* (Iida *et al.,* 1998; 2000); two in *Candida apicola* (Lottermoser *et al*., 1996); eight in *Candida maltosa* (Schunck *et al.,* 1989; Ohkuma *et al.*, 1991a, 1991b); ten in *Candida tropicalis* (Craft *et al*., 2003; Sanglard *et al*., 1987; Sanglard and Loper 1989; Seghezzi *et al*., 1991; 1992) and two in *Debaryomyces hansenii* (Yadav and Loper, 1999). Some of these genes are arranged in tandem, an indication that they might have arisen as a result of gene duplication of the ancestral gene (Ohkuma *et al*., 1991b; Seghezzi *et al*., 1991).

However, no *n*-alkane or fatty acid hydroxylase encoding genes have yet been isolated from any *Rhodotorula* sp. The only P450 encoding gene isolated from a *Rhodotorula* sp. is the *benzoate-para-hydroxylase* gene (*CYP53B1*) that was cloned from *Rhodotorula minuta* (Fujii *et al.,* 1997). Studies performed in our laboratory have shown that some *Rhodotorula* spp. can utilize not only *n*-alkanes and benzoic acid but also monoterpenes as carbon sources (Moleleki, 1998). A question arose as to how many different P450 encoding genes are present in *Rhodotorula* spp. and which one(s) code for the P450s involved in monoterpene utilization. An ability to utilize *n*-alkanes is an indication that the *Rhodotorula* spp. contains *n*-alkane inducible hydroxylases. In this study, we report the isolation and sequencing of a novel P450 gene *CYP557A1* from *Rhodotorula* sp. CBS 8446.

3.2 Materials and Methods

3.2.1 Strains, vectors and media

Rhodotorula sp. CBS 8446 was obtained from the Yeast Culture Collection of the University of the Free State and was maintained on YEPD agar [1% (w/v), yeast extract,

2% (w/v) peptone, 2% (w/v) glucose and 2% (w/v) agar]. *Escherichia coli* TOP 10 (F^- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80 *lacZ*ΔM15 Δ*lacX74 recA1 deoR araD139* Δ(*ara-leu*)*7697 galU galK rpsL* (Str^r) *endA1 nupG* (Invitrogen) was used for plasmid preparations. *E. coli* carrying plasmids were grown on Luria-Bertani (LB) agar [1% (w/v), tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl and 2% (w/v) agar] supplemented with 100 µg/ml Ampicillin. The cloning of PCR products was performed using pGEM-T Easy vector (Promega).

3.2.2 Isolation of total genomic DNA

Rhodotorula sp. CBS 8446 was grown in 10 ml YEPD broth in 100 ml shake flasks at 30°C on a rotary shaker (600 rpm) until stationary phase. Cells were harvested by centrifugation at 5000 x g for 10 minutes and washed twice with 5 ml of double distilled water. The pellet was resuspended in 500 µl lysis buffer (0.1 M Tris pH 8, 50 mM EDTA, 1% SDS) by vortexing. Acid–washed glass beads (400-500 microns) were added to about 2 mm below meniscus of the suspension and vortexed for 30 seconds, then 25 µl of 5 M NaCl was added and vortexed again for 30 seconds. The cells were centrifuged at 10 000 x g for 2 minutes. The supernatant was transferred to a new eppendorf tube and extracted with 400 µl TE saturated phenol. The resultant supernatant was extracted with 400 µl of phenol: chloroform (4:1). The DNA was precipitated with 1 ml 95% ethanol, centrifuged at 12 000 x g for 10 minutes and washed with 70% ethanol, dried and dissolved in 50 µl 1xTE. To remove contaminating RNA, the DNA sample was treated with 1 µl 10 mg/ml RNase and incubated at 37°C for 60 minutes. The concentration of DNA was determined spectrophotometrically at 260 nm.

3.2.3 PCR amplification of P450 gene fragments

PCR amplification of the isolated genomic DNA was performed under standard reaction conditions in an automated 2400 PCR system (Perkin Elmer). The Mol2 and Mol3 primers were used (Table 1). The PCR reaction mixture of 50 µl contained 1x polymerase

buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCL, pH 8.3), 10 mM of each dNTP, 100 pmol of each primer, 100 ng of DNA template and 2.5 U *Taq* DNA polymerase (Roche). The following cycle profiles were used: an initial DNA denaturation at 94°C for 2 minutes, followed by 35 cycles with denaturation at 94°C for 30 seconds; annealing at 48°C for 30 seconds; primer extension at 72°C for 1 minute, and finally primer extension at 72°C for 7 minutes.

The PCR products were separated by electrophoresis on a 1% (w/v) agarose gel containing 0.5 μg/ml ethidium bromide. The appropriate size bands were sliced from the gel and purified using the GFX™ PCR DNA and Gel band Purification kit according to supplier's instructions (Amersham Biosciences). The purified DNA samples were ligated into a pGEM-T Easy vector according to the standard protocol provided by the supplier (Promega). The ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells. The positive clones were grown in LB medium containing 100 μg/ml ampicillin. Plasmid was isolated from each clone by an alkaline lysis miniprep method (Sambrook *et. al*., 1989). The presence of appropriate sized insert was determined by cleavage of each isolated plasmid with *Eco*RI*,* which cleaves on either side of the ligation site of the pGEM-T Easy vector.

3.2.4 Nucleotide sequence analysis

The insert in the pGEM-T Easy vector was sequenced using T7 and Sp6 universal primers. The sequencing was based on the dideoxynucleotide chain termination method using fluorescence labeled BigDye™ Terminator v3.0 DNA sequencing kit according to the supplier’s protocol (Applied Biosystems). The samples were sequenced by ABI PRISM ™ 377 DNA sequencer (Perkin Elmer) and the DNA sequences obtained were analyzed using Sequence Navigator (Perkin Elmer). The sequences obtained were translated into amino acid sequences using translation tools at http://www.expasy.ch/tools. The sequences were compared with sequences in the GenBank database using the "Gapped Blast and PSI-BLAST" search program at http://www.ncbi.nlm.nih.gov/cgi-bin/blast/nph-newblast (Altschul *et al*., 1997).

Table 1. Primers used in this study.

Name	**Sequence (5'-3')**
Mol2	GGTAGAGATACCACTGC
Mol3	GACCCAAACAAATTCTTGGACCACC
Mol246	CGACACCACTGCACAAAGCCTG
Mol346	CAAACATATTCTTGGTCCGCCGTT
Mol46Fnest	ATTGGAAACGGCATCTTCGTCACAGAC
Mol46Rnest	GTCTGTGACGAAGATGCCGTTTCCAAT
Mol46RC-5nest	CAATCGAATGAACTCGCCAGCCTTGATC
MoliF	AAGCCTTCGCGCTGGATCGATGAC
MoliR	GTTCCGTACCACCGGCAAGCTGG
Mol*Bam*HI	CGGGATCCATGCTCGCGCTCG
Mol*Avr*II	GGCCTAGGTCATCGTCGGGATATTGTTACC
Mol46FNATG	CCGACCTCGCAGGCCCGAA
Mol46ATG3	ATGCTCGCGCTCGTCTGCG
Sub1R1	GTCATGAACATGGTCAGGAGGTC
Sub1F1	TGGCCTCTGTAAGCTTCTCTCG
Sub1F2	CTTTCAGGAATAGACTGAGCGACAT
Sub1FR1	GTCGGATGTCGCTATGGCAT
Sub2F1	ATATCCCGACGATGATGCGT
Sub2RF1	ACTTGATAGGAGCAGGATGAACG

3.2.5 Southern Hybridization

Five microgram of total genomic DNA was digested overnight with *Eco*RI, *Bam*HI, *Hind*III, *Pst*I and *Xba*I (Roche) at 37ºC. The digested DNA was electrophoresed on a 1% agarose gel at 80 V for 4 hours. The DNA was depurinated by washing the gel in 0.25 M HCl for 7 minutes. Denaturation was performed by washing the gel two times for 15 minutes in denaturation solution (0.5 M NaOH, 1.5 M NaCl). The gel was neutralized in 0.5 M Tris pH 7.0, 1.5 M NaCl for 30 minutes. The DNA was blotted onto 0.22 micron

nylon membrane (Separations) by capillary transfer. The membrane was washed in 10x SSC (1.5 M NaCl, 0.15 M Na-citrate, pH 7.0) for five minutes, and the DNA was cross-linked to the membrane by exposure to UV light using GS Gene Linker (Bio-Rad).

The PCR product from *Rhodotorula* sp. CBS 8446 was digoxigenin labeled using a DIG DNA labeling kit according to the supplier's protocol (Roche) and hybridization was performed at 55ºC for 24 hours. All post hybridization washes were done under the following low stringency conditions: 2 x 5 minutes in 2 x SSC, 0.1% SDS at room temperature, followed by washing of the membrane 2 x 15 minutes in 0.1% SSC, 0.1% SDS at room temperature. Detection was performed as described by the supplier's protocol (Roche).

3.2.6 Inverse Polymerase Chain Reaction (IPCR) amplification of *Pst*I fragment

The genomic DNA was prepared using a method described by Arand *et al*., (1999). Fifty microgram of genomic DNA was digested with *Pst*I. The digested DNA was extracted once with phenol: chloroform: isomyl alcohol (25:24:1). The aqueous layer was further extracted with chloroform and the DNA was precipitated with ethanol. The concentration of the recovered DNA was determined spectrophotometrically at 260 nm. The purified DNA was diluted to final concentrations of 7 ng/μl and 14 ng/μl respectively. The T4 DNA ligase (Roche) was added to a final concentration of 1 U/μl and the samples were incubated at 14ºC for 16 hours.

To perform the IPCR the primers were designed in such a way that they were facing in opposite direction to each other to include the helix I (HI) and heme-binding (HR2) regions in amplification. The IPCR reaction mixture of 50 μl contained 1x polymerase buffer 3 containing 2.25 mM $MgCl_2$, 25 mM of dNTP, 100 pmol of MoliF and MoliR primers (Table 1), 14 or 28 ng DNA template and 2.6 U *Taq* DNA expand long template (Roche). The following cycle profiles were used: an initial DNA denaturation at 94°C for 2 minutes, followed by 10 cycles with denaturation at 94°C for 30 seconds; annealing at 65°C for 30 seconds; primer extension at 68°C for 3 minutes. These cycles were followed

by 20 cycles with denaturation at 94°C for 30 seconds; annealing at 65°C for 30 seconds; primer extension at 68°C for 3 minutes with 20 seconds increase per cycle, then finally primer extension at 68°C for 7 minutes.

The IPCR product was run on a 1% agarose gel and bands of interest were sliced out from the gel and purified. The purified products were ligated into pGEM-T Easy vector, transformed into *E. coli* and sequenced. To facilitate sequencing the IPCR product was cut out of pGEM-T Easy vector with *Eco*RI and *Pst*I and subcloned into pGEM-3Z (Promega). Since the subcloned fragments were too long to be sequenced once in one reaction, the walking sequencing primers were designed.

3.2.7 Isolation of total RNA and preparation of poly $(A)^+$ mRNA

Fresh cells were streaked on YNB agar plates [2% agar and 0.67% YNB (Difco)]. The plates were placed in a desiccator together with a glass vial containing 200 µl of decane or limonene (Fluka) and a second vial containing 1 ml of water. The desiccator was tightly sealed and incubated at 25°C for 9 days. The cells were scraped from the surface of the agar and washed twice with DEPC treated water.

Total RNA was isolated using the modified TRIzol (Invitrogen) method. The cells were frozen in liquid nitrogen and transferred to a pre-cooled mortar and ground to a fine powder. The fine powder was transferred to a pre-cooled eppendorf tube to approximately 0.5 ml. The ground cells were homogenized in 1 ml of TRIzol and incubated at room temperature for 5 minutes. Two hundred microliters of chloroform was added, the mixture was vortexed vigorously for 15 seconds and incubated for 3 minutes at the room temperature. The samples were centrifuged at 12000 x g for 15 minutes at 4˚C and the supernatant was transferred to new tubes. The RNA was precipitated by adding 500 µl of isopropanol to the supernatant, vortexed and incubated at room temperature for 10 minutes. The samples were centrifuged at 12000 x g for 10 minutes and washed with 1 ml 70% ethanol. The samples were dried in a speedyVac and

dissolved in 50 μl of formamide. The poly$(A)^+$ mRNA was isolated from total RNA by using the mRNA Isolation kit according to the supplier's instructions (Roche).

3.2.8 RT-PCR amplification of mRNA isolated from decane and limonene grown cells

The primers used for RT-PCR were designed from the specific sequences of helix I (HI) and heme-binding (HR2) regions of the sequenced *Pst*I fragment of IPCR. The RT-PCR reaction was performed using Access RT-PCR Introductory System kit (Promega). The RT-PCR reaction mixture of 50 μl contained 1 x AMV/*Tf1* reaction buffer, 10 mM of each dNTP, 50 pmol each of Mol246 and Mol346 primers, 1 mM of $MgSO_4$, 5 U AMV reverse transcriptase and 5 U *Tf1* DNA polymerase. A negative control reaction was set up where reverse transcriptase (AMV) has been omitted. The following cycle profiles were used: the synthesis of the first strand cDNA at 48°C for 45 minutes, the AMV reverse transcriptase inactivation and cDNA denaturation at 94°C for 2 minutes. This cycle was followed by 40 cycles with denaturation at 94°C for 30 seconds; annealing at 57°C for 1 minute; primer extension at 68°C for 1 minute and finally primer extension at 68°C for 7 minutes. The RT-PCR products were cloned into pGEM-T Easy vector and sequenced.

3.2.9 Amplification of flanking regions of cDNA using the 5'/3' RACE technique

a) 3' end amplification of cDNA: cDNA was synthesized in a 20 μl reaction. The reaction components consisted of 1x cDNA synthesis buffer (50 mM Tris-HCl, 8 mM $MgCl_2$, 30 mM KCl, 1 mM dithiotreitol, pH 8.5), 34 ng of poly$(A)^+$ RNA in 12 μl of water, 12.5 μM of oligo(dT)-anchor primer, 20 mM of each dNTP and 20 U AMV (Roche). The reaction mixture was incubated at 55°C for 60 minutes and further incubated at 65°C for 10 minutes. The PCR was performed using 1 μl of the synthesized cDNA template. The PCR reaction mixture of 50 μl consisted of 1x polymerase buffer 1 containing 1.75 mM $MgCl_2$, 10 mM of each dNTP, 12.5 μM each of Mol246 and anchor primers, and 2.6 U

Taq DNA expand long template (Roche). The following cycle profiles were used: an initial DNA denaturation at 94°C for 2 minutes, followed by 10 cycles with denaturation at 94°C for 15 seconds; annealing at 62°C for 30 seconds; primer extension at 68°C for 40 seconds. These cycles were followed by 20 cycles with denaturation at 94°C for 15 seconds; annealing at 62°C for 30 seconds; primer extension at 68°C for 40 seconds with 20 seconds increase per cycle, then finally primer extension at 68°C for 7 minutes. The RT-PCR products were cloned and sequenced.

b) 5' end amplification of cDNA: The cDNA synthesis was performed as described in 3'RACE technique, except Mol346 was used as a reverse primer. The synthesized cDNA was purified using a High Pure PCR Product Purification kit (Roche). The purified cDNA was A-tailed in a volume of 25 μl. The reaction components consisted of 1x reaction buffer (100 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 19 μl of purified cDNA, 5 mM of dATP (Roche). The reaction mixture was incubated at 94°C for 3 minutes and chilled on ice. The reaction was started through the addition of 10 U terminal transferase (Roche) and incubated at 37°C for 20 minutes. Finally the reaction was incubated at 70°C for 10 minutes to inactivate the terminal transferase. The PCR was performed using 5 μl of the A-tailed cDNA as a template. The PCR reaction mixture of 50 μl consisted of 1x polymerase buffer1 containing 1.75 mM $MgCl_2$, 10 mM of each dNTP, 12.5 μM each of Mol346 and oligo(dT)-anchor primers, and 2.6 U *Taq* DNA expand long template (Roche). The PCR cycle profiles were the same as in 3' end amplification. The resulted PCR product was diluted twenty-fold and 1 μl was used as a template in the second round of nested PCR using MoliR and anchored primers.

3.2.10 RT- PCR amplification of 5' end of cDNA using walking primers

The 5'end sequence of the genomic DNA was translated into three frames and compared with *Arabidopsis thaliana* cytochrome P450 sequence that showed highest homology by BLAST search. The conserved amino acids were used to design forward walking primers (Mol46Fnest, Mol46ATG3 and Mol46FNATG) (Table 1). RT-PCR profiles were the same as in section 3.2.8. Mol46Fnest was used in combination with Mol46RC-5nest reverse primer. Mol46ATG3 and Mol46FNATG primers were used in two different RT-PCRs in combination with Mol46Rnest reverse primer. The RT-PCR products were cloned and sequenced.

3.2.11 Isolation of full-length cDNA by RT-PCR amplification

The primers were designed to amplify the full-length cDNA. The cDNA was synthesized using 3' RACE technique (Frohman *et al.,* 1988). PCR was performed using 1 µl of synthesized cDNA. The PCR components were the same as described in section 3.2.9a, except 12.5 µM of each Mol*Bam*HI and Mol*Avr*II primers were used. A control reaction containing genomic DNA as a template was included. The following cycle profiles were used: an initial DNA denaturation at 94°C for 2 minutes, followed by 10 cycles with denaturation at 94°C for 15 seconds; annealing at 55°C for 30 seconds; primer extension at 68°C for 40 seconds. These cycles were followed by 20 cycles with denaturation at 94°C for 15 seconds; annealing at 55°C for 30 seconds; primer extension at 68°C for 40 seconds with 20 seconds increase per cycle, then finally primer extension at 68°C for 7 minutes.

The PCR products were cloned and sequenced. The sequences were translated to obtain the open reading frames (ORF). The nucleotide and amino acid sequences were compared with other sequences in the database by Blast search program. The nucleotide sequences were submitted into Genbank under the accession number AY316198.

3.3 Results

3.3.1 PCR amplification and sequence analyses of cytochrome P450 gene fragments

The primers (Mol2 and Mol3) used in this study were identical to the primers used by Lottermoser *et al.* (1996) to isolate fatty acid hydroxylating genes from *C. apicola*. They designed the Mol2 and Mol3 primers based on the conserved sequences of the helix I (HI) and heme-binding (HR2) domains of 15 CYP52 proteins from *C. tropicalis* and *C. maltosa*. The primer nucleotide sequences were based on the sequences of *C. maltosa* *CYP52A3* gene. This primer pair amplified three fragments of 900 bp, 600 bp and 400 bp from *Rhodotorula* sp. CBS 8446 and a 460 bp fragment from *C. maltosa*, which was used as a control (Fig. 1).

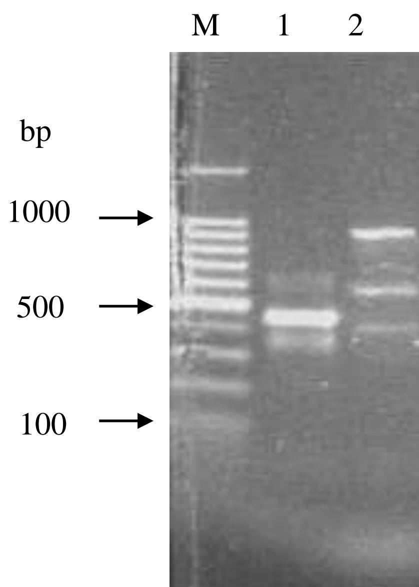


Figure 1. PCR products obtained with primer pair Mol2 and Mol3 from genomic DNA of *C. maltosa* CBS 5611T (lane 1) and *Rhodotorula* sp. CBS 8446 (lane 2). Ethidium bromide stained 1% agarose gel. M = 100 bp ladder.

These PCR products were ligated into the pGEM-T Easy vector and the ligation mix was transformed into *E. coli* cells. Positive colonies were grown in liquid LB medium and the

plasmids were isolated from the clones. To confirm the presence of the inserts in the vector, the clones were digested with *Eco*RI that cleaves on either side of the ligation site.

The clones with inserts of the correct sizes were sequenced in both directions using the T7 and Sp6 universal primers. The sequence analyses showed that some clones, particularly those containing the 900 bp and 400 bp PCR fragments, were on both sides flanked by the sequence of the Mol3 primer. This could be ascribed to nonspecific binding of the primers during PCR amplification. More clones were screened with PCR using only the Mol3 primer. The sequence analyses showed that clones containing the 600 bp fragment were flanked by both primers. The sequences of these clones were identical with the exception of possible sequencing errors and were analyzed further.

The nucleotide sequence of the 600 bp fragment were compared with sequences in the Genbank database using the BLAST search program, but did not show significant similarity to any P450 gene in the Genbank. The nucleotide sequence was subsequently translated into amino acid sequences in the three reading frames corresponding with the orientation of the primers. The nucleotide sequence of fragment did not code for a single open reading frame (ORF). Several stop codons occurred in each of the reading frames, indicating the presence of introns in the fragment. When the relatively long amino acid sequences from different frames were compared to sequences in the Genbank by BLAST search, they showed a homology with several P450 proteins from a range of organisms including plants, animals and alkane utilizing yeasts. Thus, it was possible to conclude that the isolated gene fragment belonged to a cytochrome P450 gene, possibly an alkane or fatty acid hydroxylase.

3.3.2 Southern hybridization

Inverse PCR was considered an attractive method for obtaining the sequence of the complete P450 gene. With this in mind the genomic DNA isolated from *Rhodotorula* sp. CBS 8446 was digested with different restriction enzymes and transferred to a nylon membrane. The sequenced 600 bp fragment was labeled and used as a probe for detection

of restriction fragments containing possible CYP genes. Restriction enzymes *Eco*RI, *Bam*HI and *Xba*I gave fragments of approximately 21 kb in size, while *Hind*III and *Pst*I gave fragments of approximately 2 kb and 3 kb respectively (Fig. 2). The latter fragments had reasonable sizes for cloning and construction of a library. Since only one band was observed for each enzyme used, it can be concluded that there is probably only one copy of the gene of interest present in the genome and that there are also no other closely related CYP genes present.

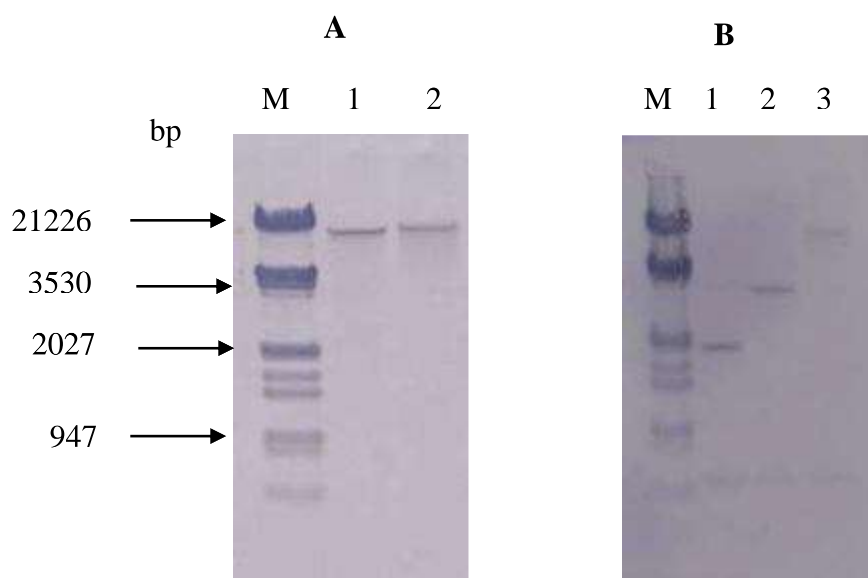


Figure 2. Southern blot analysis of DNA isolated from *Rhodotorula* sp. CBS 8446 using the 600 bp HI-HR2 CYP fragment as a probe. (**A**) DNA was digested with *Bam*HI (lane 1), and *Eco*RI (lane 2). **B**) DNA was digested with *Hind*III (lane 1), *Pst*I (lane 2) and *Xba*I (lane 3). M = λ DNA digested with *Hind*III/*Eco*RI. Hybridization was performed at 55 ºC and all washes were performed at room temperature.

3.3.3 Inverse Polymerase Chain Reaction (IPCR) amplification of the putative CYP gene from the *Pst*I fragment from *Rhodotorula* sp. CBS 8446

Inverse PCR requires the digestion of genomic DNA with restriction enzymes followed by re-ligation of digested DNA in diluted form on itself to form monomeric circles (Arand *et al.*, 1999). The primers MoliF and MoliR were designed facing in opposite

orientation from the known sequence so that they could amplify the unknown sequence. Genomic DNA from *Rhodotorula* sp. CBS8446 was digested with *Pst*I, purified and treated with ligase. The primers MoliF and MoliR were subsequently used in IPCR, and amplified a fragment of ~3 kb (Fig. 3). This result correlated with the result obtained from Southern hybridization where a band of ~ 3 kb was obtained from *Pst*I digested DNA (Fig. 2).

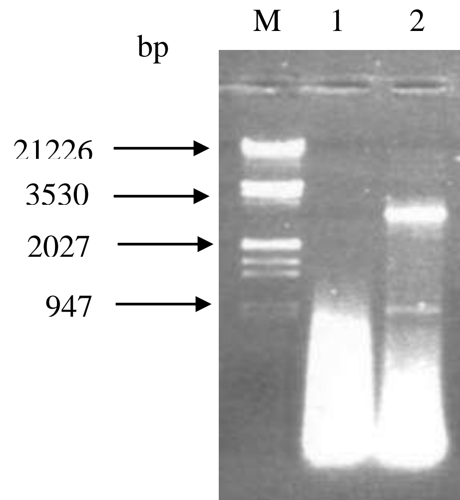


Figure 3. Products obtained from inverse PCR amplification on the *Rhodotorula sp.CBS 8446 Pst*I self-ligated DNA fragment using MoliF and MoliR primers. PCR was performed using 14 ng (lane 1) and 28 ng (lane 2) ligated DNA templates. M = λ DNA digested with *Hind*III/*Eco*RI. Ethidium bromide stained 1% agarose gel is shown.

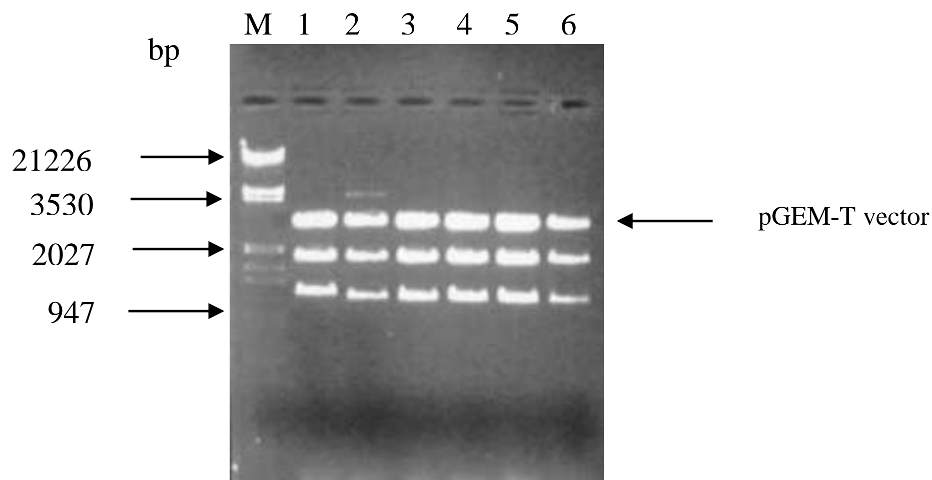


Figure 4. Fragments of 1.1 and 1.9 kb obtained when the IPCR product from *Rhodotorula* sp. CBS 8446 *Pst*I fragment was subcloned into pGEM-T Easy vector and resultant clones were digested with *Pst*I/*Eco*RI. Ethidium bromide stained 1% agarose gel showing the digested clones (lanes 1-6). Lane M = λ DNA digested with *Hind*III/*Eco*RI.

The IPCR product was cloned into pGEM-T Easy vector and then double digested with *Eco*RI and *Pst*I. *Eco*RI cuts on both sides of the vector to release the cloned insert and *Pst*I cuts within the insert. Two *Eco*RI/*Pst*I fragments of 1.9 kb and 1.1 kb were obtained (Fig. 4). These fragments were subsequently subcloned into pGEM-3Z and sequenced by using walking primers (Fig. 5). The sequence analyses of the fragments showed that they were interrupted by introns, since the translated amino acid sequences showing similarity to CYP sequences shifted from one reading frame to another.

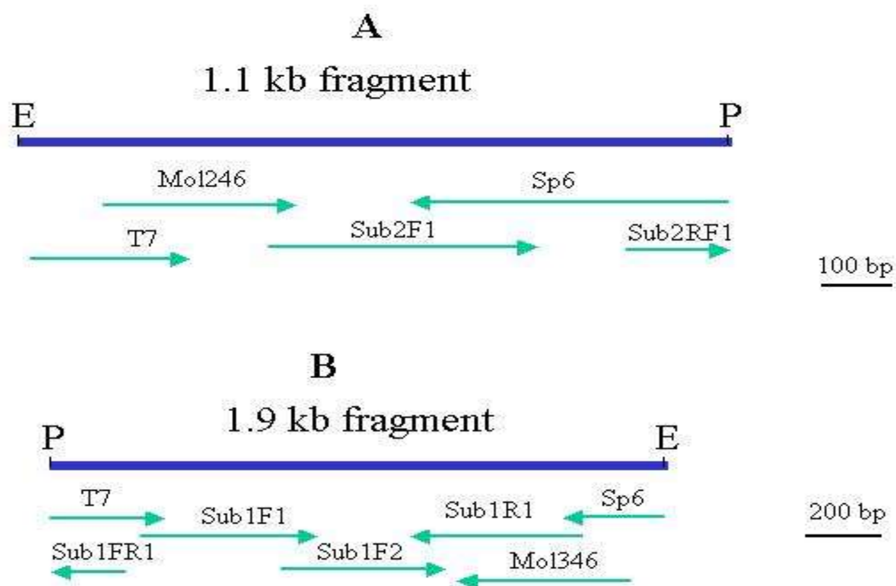


Figure 5. Sequencing strategy of IPCR product. The obtained IPCR product from *Rhodotorula* sp. CBS 8446 was cloned into pGEM-T Easy vector and the resultant clones were digested with *Pst*I/*Eco*RI to yield 1.1 and 1.9 fragments. The fragments were subcloned into pGEM-3Z and sequenced by using walking primers. Arrows in the figure show the direction of the sequencing primers. E and P refer to *Eco*RI and *Pst*I restriction sites respectively.

The real nucleotide sequences for the HI and HR2 regions were determined from the sequences of the IPCR fragments since in the initial 600 bp fragment these sequences were the same as those of the original primers, which were based on the sequences from *C. maltosa* (Table 1). Comparison of these sequences showed that Mol2 has 10 out of 17 nucleotides identical to the corresponding sequence in the HI region, while Mol3 has 18 out of 25 nucleotides identical to the corresponding HR2 sequence. The high nucleotide identity of the primers to the corresponding sequences in the gene fragments; explains why these primers amplified a CYP fragment from *Rhodotorula* sp. CBS 8446.

3.3.4 RT-PCR amplification of mRNA isolated from cells grown on limonene and decane

Many *Rhodotorula* spp. can grow on non-carbohydrate carbon sources such as *n*-alkanes, phenols and benzoate (Durham *et al*., 1984; Durham, 1984; Wright and Ratledge, 1991). *Rhodotorula* sp. CBS 8446 had been isolated from enrichments done with cyclohexane dicarboxylic acid as only carbon source and it was subsequently shown that it could grow on decane and limonene when these substrates were supplied in the vapor phase (Moleleki, 1998). mRNA for RT-PCR was isolated from cultures grown on YNB agar plates without carbon source, placed in desiccators saturated with limonene or decane.

The primers Mol246 and Mol346 used for RT-PCR amplification were designed based on the specific sequences of the HI and HR2 regions obtained from the sequenced *Pst*I fragment amplified by IPCR. A cDNA fragment of approximately 490 bp was amplified from the mRNA that was isolated from cells grown on decane (C-10) as a carbon source (Fig. 6). There was no amplification from the mRNA isolated from glucose grown cells and where reverse transcriptase (RT) was omitted (data not shown). The absence of a product from the sample where RT was omitted was an indication that the amplification was from mRNA and not as the result of DNA contamination. In addition, the obtained RT-PCR product was smaller than the genomic DNA PCR product. The same primers amplified a fragment of approximately 545 bp from mRNA isolated from limonene grown cells (Fig. 6).

The amplified cDNA fragments were cloned into pGEM-T easy vector and sequenced. Translation of this sequence obtained from the cDNA amplified from decane grown cells revealed a single ORF. On the other hand, a single ORF was not obtained from cDNA amplified from limonene grown cells. The nucleotide sequences were compared with the genomic DNA sequences and the positions of introns were determined. The amplified fragment of the genomic DNA is interrupted by three introns. All introns were spliced out in the cDNA amplified from decane grown cells, while the second intron (~55 bp) remained in the cDNA amplified from limonene grown cells. A BLAST search against

the NCBI protein database using the translated RT-PCR fragment obtained from decane grown cells as query indicated significant homology to a number of P450 proteins. The study of mRNA from limonene grown cells was discontinued, since a single ORF could not be obtained from cDNA amplified from mRNA isolated from limonene grown cells. All further cDNA synthesis and RT-PCR was carried out on mRNA isolated from decane grown cells.

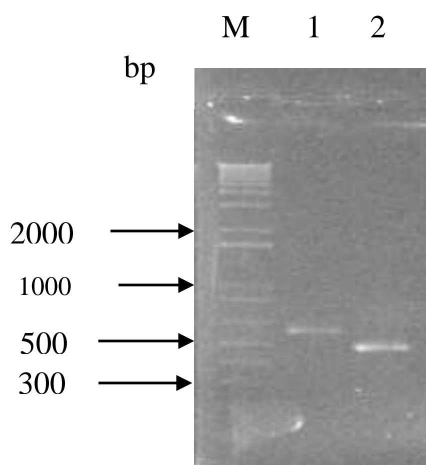


Figure 6. RT-PCR amplification of mRNA. Ethidium bromide stained 1 % (w/v) agarose gel showing the RT-PCR amplification of mRNA from limonene grown cells (lane 1) and decane grown cells (lane 2). M = 1 kb ladder marker.

3.3.5 Application of Rapid Amplification of cDNA Ends (5'/3' RACE) Technique

The 5'/3' RACE technique amplifies the unknown sequences flanking a known sequence using cDNA as a template (Frohman *et al.,* 1988). It utilizes a sequence specific primer in combination with an adaptor specific primer that binds on homopolymeric A-tails located at the ends of the unknown parts of the sequence. Poly(A)$^+$ RNA was isolated from total RNA in order to decrease the background during cDNA synthesis and in the subsequent PCR amplification. To amplify the 3' end of the cDNA an oligo(dT) primer was used. The oligo(dT) primer consisted of 15 dT residues with flanking adaptor sequences. The dT residues of the oligo dT primer bind on the poly (A)$^+$ tail of the mRNA and with reverse transcriptase the mRNA is transcribed into the corresponding cDNA. The

flanking adaptor sequence of the primer places a unique sequence at the unknown end of the cDNA.

For amplification of the 3' end the Mol246 primer was used in combination with the adaptor specific primer. Mol246 represents the conserved HI sequence. Thus, amplification specificity of the PCR is determined by the Mol246 primer. After PCR amplification three distinct bands of 700 bp, 500 bp and 300 bp respectively were obtained (Fig. 7). These bands were purified from the gel, cloned and were successfully sequenced using T7 and Sp6 universal primers.

Only the sequence of the 700 bp fragment corresponded to the sequence in the genomic DNA. The other fragments did not match the known sequence and probably resulted from nonspecific amplification. Nonspecific amplification has been observed when the RACE technique was used (Frohman *et al*., 1988). It happens because during cDNA synthesis the dT residues bind to sequences that are A-rich.

To amplify the 5' end, the cDNA synthesis was carried out using Mol46RC-5nest reverse primer. This primer was based on the cDNA sequence just downstream of the HI sequence. The primer-extended products were separated from the excess primers and a homopolymeric A-tail was added at the 3' end of the cDNA.

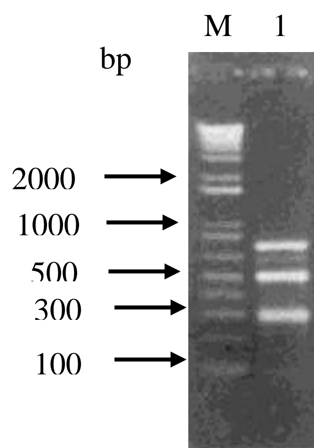


Figure 7. Application of 3' RACE technique to obtain downstream part of the gene fragment**.** Ethidium bromide stained 1 % (w/v) agarose gel showing the amplified PCR product of the 3' of the gene fragment (lane1). M = 1 kb ladder marker.

The first amplification PCR was carried out using the oligo(dT) primer in combination with the Mol46RC-5nest primer. Nested PCR was performed using the adaptor sequence specific primer and a gene specific primer MoliR, that binds upstream of the primer Mol46RC-5nest. The second nested specific primer should increase the specificity and efficiency of the amplification, since it should bind only to the cDNA of interest. After several attempts of 5' end amplification of cDNA there was no clear bands observed (data not shown). Several explanations could be offered for this. The fragment to be amplified might have been too long or very few copies of the target transcript might have been present. Another possibility was that a poly (A)–tail had not been introduced at the 3' end of the cDNA.

3.3.6 RT-PCR amplification of 5' end of cDNA using walking primers

After several attempts the RT-PCR product could not be obtained from the 5' end of the sequence by using the 5'RACE technique. The genomic DNA corresponding to the 5' end was translated in three reading frames and a BLAST search was performed using the 'exact short matches' option. A cytochrome P450 sequence from *Arabidopsis thaliana* showed the highest homology to the submitted sequence. A number of amino acid motives that were identical in the two sequences were used to design forward walking primers (Mol46ATG3, Mol46FNATG and Mol46Fnest respectively) (Table 1) (Fig. 8). When Mol46Fnest was used in combination with Mol46RC-5nest reverse primer two fragments of approximately 1000 and 800 bp were amplified (Fig. 9A). These fragments were cloned and sequenced. The band of 1000 bp did not correspond to the known sequence, while the sequence of the 800 bp fragment corresponded to the genomic DNA sequence. This fragment is 400 bp smaller than the corresponding genomic fragment that was amplified by the same primers.

Since Mol46Fnest had given a correct cDNA product, its sequence was reverse complemented to give a reverse primer (Mol46Rnest). When Mol46FNATG and Mol46Rnest primers were used in RT-PCR, two bands of 400 and 350 bp were obtained

(Fig. 9B). Sequencing showed that only the 350 bp fragment corresponded to the genomic DNA sequence.

When Mol46ATG3 and Mol46Rnest primers were used in RT-PCR, three bands of 600, 550 and 450 bp respectively were obtained (Fig. 9C). The sequence of the 600 bp fragment did not correspond to the known sequence and does not contain an ORF. The sequence analyses showed that the sequences of the other two bands corresponded to the genomic sequence. However, the sequence of the 550 bp fragment does not contain an ORF while the 450 bp fragment gave an ORF. An intron of 100 bp had remained in the 550 bp fragment, but was spliced out from the 450 bp fragment.

CTGCAGGTGGCAAGGACGCCGCGGCAAGCGGCATCAACGCCGCTCCTCGAGGGTCGTTGCGTTTCTCAGT -71
CCTGCTTCTACATAGACTCCATTTCGTCACACCTGTCGCCTCCTGCGCGTCTCGGTTGTGCCCTTCCCTG -1

— Mol46ATG3 & Mol*BAMH1* →

ATGCTCGCGCTCGTCTGCGTTGGTCTCTTCCTGTACCTCCTGTGCAAGTAGTAAGCAATGCCATTGACAA 70
ATCGCTCGCCCGTTGACGCTGAGCCGTCTTGCAGTCGGTCCCATGCCATAGCGACATCCGACAGGA**CCGA** 140

— Mol46FNATG →

CCTCGCAGGCCCGAAGGGGTGGCCTCTGTAAGCTTCTCTCGAGCTCGTCCCGTTGTGGCGGCGCATCGGG 210
//
TACCTCGCCCTGCATATGCACTCGCGGGCGTCCCGTTCATCGTTTCGTGCTTGCTCCAGCTTTCAGGAAT 630

← Mol46Fnest & Mol46Rnest →

AGACTGAGCGACATT**ATTGGAAACGGCATCTTCGTCACAGAC**GGGTAAACCAATGTGATCTATCGGGATG 700
AACAGAGCTTACTCGATCAGGGCTTCGCAGGCCGCGTATGGCAGTTTCAACGCAAGGTGACCAGGTCATT 770
//
TACCGAGCGCCTAATTTCCTCCGACGCCCAGGTATGATGACTCGCGTCCACTCTCGCGCGTCGAGATCCG 1470
GGACGCGATCATCAACCTCCTTTTGGCAGGGTACACGAGACATCCTTGTGTTGACGTGGAGACAGAGAGG 1540

— Mol246 →

ACTGACTCTGTCTTTGCAGGCG**CGACACCACTGCACAAAGCCTG**ACGTGGGCGCTATATCGCCTTGTGGA 1610

← MoliR —

CAATCCGGCACACCAACGACTCGTGCGGGAAGAGATCTGTGG**CCAGCTTGCCGGTGGTACGGAAC**CCGTG 1680
ACATTGGAGAACGTCAAGAATCTGGTACAGACGCAGGCAGCCCACCTGGAGAGTCTCCGACTCCATCCGC 1750
CGGTCCCACGGATAGTCAAGTACGTCTACTGGGTCACAAGACTGGCTTCCTGTGCCTGATACAAAGCAAC 1820

← MolRC-5nest —

AGGCAAGCTGTGAAAGACGACGCTTTGCCCAACGGAGGGCCGCT**GATCAAGGCTGGCGAGTTCATTCGAT** 1890
TGTGTGAGCTCTTGCACTGCGCACCGCTGGAGATTCGATCTCATGGACGGATAATCACCGGACTGGGCTT 1960

— MoliF →

TGGGGCGCAACGAAGAAGTGTGGGGAGCCGACGCGAAAGAGTGG**AAGCCTTCGCGCTGGATCGATGAC**GA 2030
AGGACGTCCGATCCAGTACAGCCAGTGGAAGGTAAACAGTGTCCCTCTCCGAGTCTGTCTCACAACGCAT 2100

← Mol346 —

GACGGACCAGAGCGTACACAGGCTCACTTTTTC**AACGGCGGACCAAGAATATGTTTGGGCA**AGAGTCTGG 2170
CGACTCTTGAAGGTGCGGAACAGATAGCGAGTCTCGTTTCGACGCTCAACTAATTTTGCGCACCAGGCGT 2240
CGCCGTGATTGCAAATTTGTTGCACCGATACAATATCGCGTTCGCGCCAGGCTGGTGGGAAAATGTCGAA 2310
AAGACCGGACGCATTGCGGGCGACCTCGAAGAAACTCCTCTCTATGGCCCGGCGCTGACGCTTCCGATGA 2380

← Mol*AvrII* —

AATCTCCTTTCTGG**GTAACAATATCCCGACGATGA**TGCGTTCTGAGTGACATTTCCGCCGTGTGCTGTTT 2450
//
ATTTTTATGCCAACGTTACAAGAAGGCACGCACGTCGTTGCCAGCGGGAAGCCCGCCGCAAGCCTTGTAA 3150
GACTTGACAACTGCAG 3166

Figure 8. Genomic DNA sequence determined from IPCR product. The positions of primers that were used for RT-PCR are shown. The arrows show the direction of the primers. // indicates that part of the sequence that has been omitted.

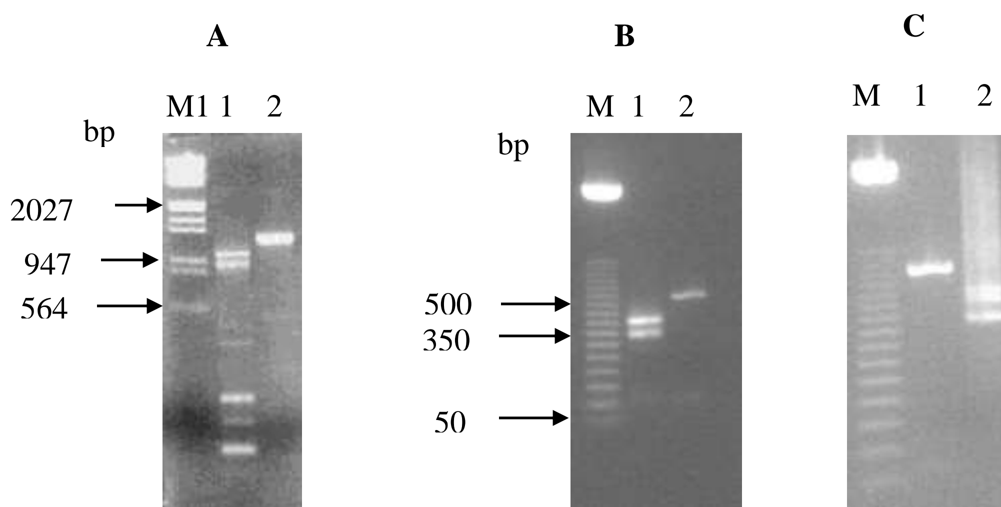


Figure 9. RT-PCR amplification of 5' end of cDNA using walking primers. Ethidium bromide stained 1.8 % (w/v) agarose gel showing the amplified RT-PCR products. **A**) Mol46RC-5nest and Mol46Fnest were used as primers. Lane 1 = RT-PCR products, lane 2 = PCR on genomic DNA. M1= λ DNA digested with *Hind*III/*Eco*RI. **B**) Mol46FNATG and Mol46Rnest were used as primers. Lane 1 = RT-PCR product, lane 2 = PCR on genomic DNA. **C**) Mol46ATG3 and Mol46Rnest were used as primers. Lane 1= PCR on genomic DNA, lane 2= RT-PCR products. M = 50 bp ladder marker.

3.3.7 Isolation of full-length cDNA

The cDNA sequences obtained from the 3' RACE technique and walking RT-PCR were put together to give the full-length cDNA. The primers Mol*Bam*HI and Mol*Avr*II were designed based on the full-length cDNA sequence. The Mol*Bam*HI primer contains a *Bam*HI restriction site and includes the ATG start codon, while the Mol*Avr*II primer contains an *Avr*II restriction site and includes the TGA stop codon. cDNA was synthesized using the oligo dT(15) primer that binds on the poly $(A)^+$ of mRNA. The primers Mol*Bam*HI and *Avr*II amplified the expected bands of approximately 1608 bp and 2400 bp from cDNA and genomic DNA respectively (Fig. 10). The cDNA product was cloned into pGEM-T easy vector and sequenced.

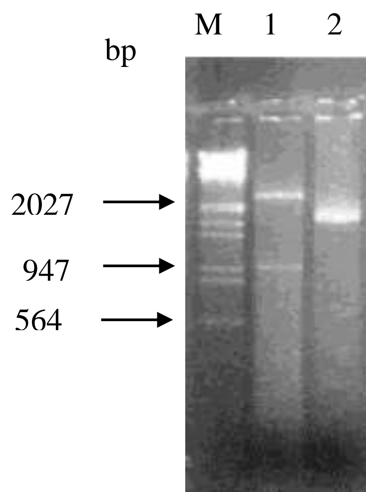


Figure 10. Isolation of full-length cDNA by RT-PCR using mRNA isolated from decane grown cells. Ethidium bromide stained 1 % (w/v) agarose gel showing the amplified genomic DNA PCR product (lane 1) and RT-PCR product (lane 2) using Mol*Bam*HI and Mol*Avr*II primers. M= λ DNA digested with *Hind*III/*Eco*RI.

3.3.8 Sequence and alternative splice variants analyses

The genomic DNA sequence showed that the putative ATG codon is at position +1 and TGA stop codon is at position 2416 (Fig. 11). Comparison of the cDNA sequence and genomic DNA sequence show that the gene is interrupted by 13 introns with sizes ranging from 53 bp to 110 bp (Fig. 11). Nine of the thirteen introns conform to the canonical rule GT/AG (Mount *et al*., 1982). The exceptions are introns 4, 7, 8, and 11 (Fig. 11 and Table 2). The TACTAAC branch point sequence was not detected in these introns. This consensus has been found in introns of *S. cerevisiae* (Davis *et al*., 2000, Langford *et al*., 1984, Teem *et al*., 1984) and plays a role in splicing efficiency. All three types of introns (type 0, I and II) (Smith, 1988) were observed in the sequence (Table 2). Type 0 appears when the intron is between codons, while type I and II appear when the intron comes after the first and second base respectively.

Sequence analyses have shown some variations in the clones that contained the amplified cDNA. Some clones retained introns number 2 and 6, and some clones were fully spliced.The clones that retained intron number 6 appeared with a higher frequency than

the clones that retained intron number 2. When limonene was used as a carbon source it was observed that intron number 11 was retained. However, when decane was used as carbon source intron number 11 was not retained. The variant clones did not give any ORF. These variations could have arisen as the result of incomplete processing of mRNA or as a result of alternative splicing. The possibility of incomplete processing can be ruled out, since this splicing event was observed repeatedly. The similar pattern of splicing has been observed for the *CYP63A1* gene of the fungus *Phanerochaete chrysosporium* where two alternatively spliced transcripts were observed (Yadav *et al.,* 2003). One transcript was free of all introns while the other transcript retained one intron and could not give any ORF.

3.3.9 Analyses of flanking sequences

There was not much information on the sequences upstream of the start codon. Only 140 nucleotides were obtained (Fig. 11). Sequence analyses of this sequence at http://www.fruitfly.org/seq_tools/promoter.html and by manual inspection, did not reveal any regulatory elements such as TATA box, CCAAT box, repeated elements and palindrome sequences as was found for other P450 genes (Lottermoser *et al*., 1996; Yadav and Lopper, 1999, Yadav *et al*., 2003). The repetitive elements have been implicated in control of transcription in *S. cerevisiae* (Rudolf and Hinnen, 1987; Giniger *et al*., 1985). The consensus sequence G/ANNATG which has been found in highly expressed genes in yeast (Hamilton *et al*, 1987) including some P450 genes (Seghezzi *et al*., 1991; Yadav and Lopper, 1999, Yadav *et al*., 2003), was not in this case present in the sequence preceding the ATG codon.

In the 3' non-coding region the tripartite signal [5'- TAG-TA(T/A)GT-TTT-3'] was detected (Fig. 11) but no AATAAA box was found. These sequences are implicated in transcription termination (Guo *et al*., 1995; Guo and Sherman, 1996; Zaret and Sherman, 1982).

```
CTGCAGGTGGCAAGGACGCCGCGGCAAGCGGCATCAACGCCGCTCCTCGAGGGTCGTTGCGTTTCTCAGT -71
CCTGCTTCTACATAGACTCCATTTCGTCACACCTGTCGCCTCCTGCGCGTCTCGGTTGTGCCCTTCCCTG -1
M  L  A  L  V  C  V  G  L  F  L  Y  L  L  C  K  Y
ATGCTCGCGCTCGTCTGCGTTGGTCTCTTCCTGTACCTCCTGTGCAAGTAgtaagcaatgccattgacaa  70
           intron 1                R  S  H  A  I  A  T  S  D  R  T  D
atcgctcgcccgttgacgctgagccgtcttgcagTCGGTCCCATGCCATAGCGACATCCGACAGGACCGA  140
 L  A  G  P  K  G  W  P  L
CCTCGCAGGCCCGAAGGGGTGGCCTCTgtaagcttctctcgagctcgtcccgttgtggcggcgcatcggg  210
                     Intron 2                                        M
cgtgcgcgtttcgtcctcccgtagctgacgccatcgcgtcgacgcgtggggctaaacctggccacagGAT  280
 G  V  F  K  F  P  L  P  T  L  S  V  T  Y  V  V  V  D  V  T  E  K  T
GGGCGTCTTTAAGTTCCCTCTTCCGACATTGAGTGTAACTTATGTGGTCGTTGACGTGACCGAAAAAACC  350
L  S  G  N  L  I  F  Y  A  T  Q  S  S  T  R  G  L  E  W  G  V  R  F  V
CTCAGCGGGAACCTCATCTTCTACGCCACCCAGAGCTCCACTCGCGGACTCGAGTGGGGTGTTCGCTTCG  420
  K  T  Y  G  H  A  W  T  L  S  V  P  G  L  R  L  I  D  I  S  T  R  P
TGAAGACATACGGCCACGCGTGGACGCTCAGCGTGCCTGGTTTGCGCCTCATCGACATCAGCACGCGCCC  490
 D  W  I  E  Y  V  Q  K  T  N  Y  H  N  Y  V  K  G  S  A
AGACTGGATCGAGTACGTTCAGAAGACGAACTACCACAACTATGTCAAAGGGTCCGgttcgtcgtttttc  560
                   Intron 3                                     F  R  N
tacctcgccctgcatatgcactcgcgggcgtcccgttcatcgtttcgtgcttgctccagCTTTCAGGAAT  630
R  L  S  D  I  I  G  N  G  I  F  V  T  D  G
AGACTGAGCGACATTATTGGAAACGGCATCTTCGTCACAGACGGgtaaaccaatgtgatctatcgggatg  700
          Intron 4                 R  V  W  Q  F  Q  R  K  V  T  S
aacagagcttactcgatcagggcttcgcaggCCGCGTATGGCAGTTTCAACGCAAGGTGACCAGgtcatt  770
                    intron 5                                 H  I  F  T  G
ctttcttacaccaccccagtcccaactttccactcatgaccaaccatcccttacagTCACATCTTTACTG  840
  R  S  F  Q  D  A  I  C  P  A  I  A  E  E  L  R  S  L  N  K  L  L  D
GGCGGTCCTTCCAGGACGCCATATGTCCTGCAATCGCCGAAGAGCTCCGATCGCTGAACAAGTTGCTGGA  910
 N  Y  A  D  T  G  E  T  V  D  L  Q  D  V  F  Y  R  F  T  F  Q  A  F
CAACTACGCCGACACGGGCGAGACTGTCGACCTTCAAGACGTCTTCTACCGCTTCACATTTCAAGCTTTT  980
G  R  F  A                 intron 6                                 F
GGGAGATTCGCgttcatacgagtacttcgtcaggcgaggtcgtcgaggctgatcataggatgcccagGTT  1050
 N  L  D  L  R  N  L  E  I  D  H  K  P  V  P  F  T  T  A
CAACCTCGACCTGCGGAACCTCGAAATCGACCACAAACCGGTCCCGTTCACGACAGgcgagttgatcgcc  1120
                 Intron 7                    F  D  Q  C  Q  K  P  S  L
acagctttgaagttgactgtccacgctgacattgggagtgcagCGTTCGACCAATGCCAGAAGCCATCAT  1190
  D  A  S  S  T  R  S  G  A  F  A  S  A  L  P  L  P  A  V  E  Y  Q  T
TGGACGCTTCTTCAACCCGTTCTGGCGCATTCGCGAGCGCTTTACCTCTGCCAGCCGTAGAATATCAGAC  1260
 P  P  S  C  F  R  L  C  L  R  N  D  D  K  R  A  A  L  L  A  K  K  D
GCCGCCGTCTTGTTTCCGACTTTGCCTACGAAATGATGACAAACGCGCCGCCTTACTGGCAAAAAAAGAC  1330
A  K  V  P  S  D  L  L  T  M  F  M  T  A  R
GCCAAGGTGCCCTCGGACCTCCTGACCATGTTCATGACAGCGAGgggagtgttcgccaacagagtcgtgg  1400
              Intron 8          Y  D  D  S  R  P  L  S  R  V  E  I  R
taccgagcgcctaatttcctccgacgcccagGTATGATGACTCGCGTCCACTCTCGCGCGTCGAGATCCG  1470
 D  A  I  I  N  L  L  L  A  G                 intron 9
GGACGCGATCATCAACCTCCTTTTGGCAGGgtacacgagacatccttgtgttgacgtggagacagagagg  1540
                   R  D  T  T  A  Q  S  L  T  W  A  L  Y  R  L  V  D
actgactctgtctttgcagGCGCGACACCACTGCACAAAGCCTGACGTGGGCGCTATATCGCCTTGTGGA  1610
```

```
  N  P  A  H  Q  R  L  V  R  E  E  I  C  G  Q  L  A  G  G  T  E  P  V
CAATCCGGCACACCAACGACTCGTGCGGGAAGAGATCTGTGGCCAGCTTGCCGGTGGTACGGAACCCGTG  1680
T  L  E  N  V  K  N  L  V  Q  T  Q  A  A  H  L  E  S  L  R  L  H  P  P
ACATTGGAGAACGTCAAGAATCTGGTACAGACGCAGGCAGCCCACCTGGAGAGTCTCCGACTCCATCCGC  1750
  V  P  R  I  V  K                    intron 10
CGGTCCCACGGATAGTCAAgtacgtctactgggtcacaagactggcttcctgtgcctgatacaaagcaac  1820
   Q  A  V  K  D  D  A  L  P  N  G  G  P  L  I  K  A  G  E  F  I  R  L
agGCAAGCTGTGAAAGACGACGCTTTGCCCAACGGAGGGCCGCTGATCAAGGCTGGCGAGTTCATTCGAT  1890
  S                    intron 11                             D  W  A  L
TGTgtgagctcttgcactgcgcaccgctggagattcgatctcatggacggataatcacCGGACTGGGCTT  1960

  G  R  N  E  E  V  W  G  A  D  A  K  E  W  K  P  S  R  W  I  D  D  E
TGGGGCGCAACGAAGAAGTGTGGGGAGCCGACGCGAAAGAGTGGAAGCCTTCGCGCTGGATCGATGACGA  2030

 G  R  P  I  Q  Y  S  Q  W  K                  intron 12
AGGACGTCCGATCCAGTACAGCCAGTGGAAGgtaaacagtgtccctctccgagtctgtctcacaacgcat  2100

                       A  H  F  F  N  G  G  P  R  I  C  L  G  K  S  L  A
gacggaccagagcgtacacagGCTCACTTTTTCAACGGCGGACCAAGAATATGTTTGGGCAAGAGTCTGG  2170
  T  L  E  G                        intron 13                        V
CGACTCTTGAAGgtgcggaacagatagcgagtctcgtttcgacgctcaactaattttgcgcaccagGCGT  2240
 A  V  I  A  N  L  L  H  R  Y  N  I  A  F  A  P  G  W  W  E  N  V  E
CGCCGTGATTGCAAATTTGTTGCACCGATACAATATCGCGTTCGCGCCAGGCTGGTGGGAAAATGTCGAA  2310
K  T  G  R  I  A  G  D  L  E  E  T  P  L  Y  G  P  A  L  T  L  P  M  K
AAGACCGGACGCATTGCGGGCGACCTCGAAGAAACTCCTCTCTATGGCCCGGCGCTGACGCTTCCGATGA  2380
  S  P  F  W  V  T  I  S  R  R  *
AATCTCCTTTCTGGGTAACAATATCCCGACGATGATGCGTTCTGAGTGACATTTCCGCCGTGTGCTGTTT  2450
GATACTCTCGTGTAGAATATGTTGCGGGCATTCGGATCTCAAATTCATTCAAAACCTTTGCGATGGCCTG  2520
TTATATGGGCGAGAGGTGACTTTTTCAGGACAGACTAGAGACTGGGCGTTGGTCCTCTAGGGAGATTCAA  2590
GGCAAGAAACCCTCATGGCGCCAGCAGCTCAACCAGCCCGCGAAATCCTTGGAGTCGAGCACAGGCAACG  2660
AAGCCATTCGTACGCGAGAGCGACTCGGCCAACTTCGTCTTCATGCGAGGAATACCAATGGCCTCGCGCC  2730
CGGGCGTCCCTCCAGTCAAGGGCTTCGCCTTGTAGTACGAGAGCAACTTGACAGCAGGGTTCGCGCTCGC  2800
GTCCGCAGCCTCAGGCGACCCAAGCTTCGCGATCCAGACATCGAAGGGGATGACAGGGACATTAAGGTCC  2870
TTGGCCAGGTAAGCAATCATCGTGCCCCACCTGCGGCAAGGTCGGTTAGTTCGCAACAAGTGAGGCAGGA  2940
GTTTCATACGCACTTGATAGGAGCAGGATGAACGAGGTGCAAAGTGCCCGGGAGCTTCGTCGCCTTCGTG  3010
AAGTCCAGAATTATCTCGGCCGCCGTGTCGGTGGGAACCCAGCATATATCCTGCGCGCGTATTTCAGAAC  3080
ATTTTTATGCCAACGTTACAAGAAGGCACGCACGTCGTTGCCAGCGGGAAGCCCGCCGCAAGCCTTGTAA  3150
GACTTGACAACTGCAG  3166
```

Figure 11. Nucleotide and deduced amino acid sequences of *CYP557A1* gene from *Rhodotorula* sp. CBS 8446. Base numbering for the coding region begins with the putative ATG codon. Intron bases are indicated in lower case letters. Amino acid residues are indicated as a single letter notation above the first base of each codon. Sequences coinciding with the initial PCR product is underlined in the coding region. The start and stop codons are shown in boldface. In the 3' non-coding region the tripartite signal is underlined. Two putative N-terminal hydrophobic domains are double underlined in the amino acid sequence. The helix I and heme-binding amino acids are dotted underlined.

Table 2. Analysis of the intron sequences

Number	Size (bp)	Multiple of 3	Intron sequence	Conform to canonical GT/AG rule	Intron type(0,I,II)	Spliced in all transcripts
1	54	Yes	gtaagcaatg //cgtcttgcag	Yes	II	Yes
2	110	No	gtaagcttct //ctggccacag	Yes	II	No
3	73	No	gttcgtcgtt //cttgctccag	Yes	I	Yes
4	57	Yes	gtaaaccaat // gcttcgcagg	No	II	Yes
5	62	No	gtcattcttt // tcccttacag	Yes	II	Yes
6	56	No	gttcatacga // ggatgcccag	Yes	II	No
7	57	Yes	gcgagttgat // gggagtgcag	No	I	Yes
8	57	Yes	gggagtgttc // cgacgcccag	No	II	Yes
9	59	No	gtacacgaga // gtctttgcag	Yes	II	Yes
10	53	No	gtacgtctac // aaagcaacag	Yes	II	No
11	55	No	gtgagctctt // ggataatcac	No	I	No
12	60	Yes	gtaaacagtg // gcgtacacag	Yes	0	No
13	54	Yes	gtgcggaaca // tgcgcaccag	Yes	I	Yes

3.3.10 Deduced protein

The isolated cDNA has an ORF of 1608 bp encoding a protein of 535 amino acids. It has an estimated molecular weight of 60 kDa and a calculated isoelectric point of 8.85 (Bjellqvist *et al.,* 1993). A Blastp search of the protein database of GenBank using the deduced amino acid sequence as query yielded 95 amino acid sequences of confirmed and putative P450 monooxygenases which gave blast scores of more than 100. Sixty sequences which gave blast scores of between 200 and 162 all belonged to plants, mainly

Arabidopsis thaliana and *Vicia sativa*. The sequences amongst these with more detailed descriptions or proven activity were assigned to the CYP86 and CYP94 families, which represent the omega hydroxylases in plants (Benveniste *et al*., 1998; Tijet *et al*., 1998; Le Bouquin *et al*., 1999). The sequences which gave blast scores between 147 and 127 mainly (with the exception of a few plant sequences) belonged to the CYP52 family of alkane and fatty acid omega hydroxylases from alkane utilizing yeasts belonging to the ascomycetes.

A ClustalW alignment with representative sequences belonging to different families of alkane and fatty acid omega hydroxylases showed that the *Rhodotorula* P450 has the highest homology (24-27%) with the CYP family from plants, while it showed 17-22% identity with the CYP52 family from yeasts (Table 3). The *Rhodotorula* P450 also showed 28-33% identity to newly sequenced putative P450s from *Ustilago maydis* (accession numbers, AACP01000066, AACP01000070, AACP01000250 and AACP01000252). Because the amino acid sequence of the *Rhodotorula* P450 shows less than 40% sequence similarity with P450s belonging to established families, the International P450 Nomenclature Committee assigned this protein to a new family CYP557. Being the first member of the family it was assigned the name *CYP557A1* (Nelson, DR., personal communication). At nucleotide level *CYP557A1* showed the highest identity with *Ustilago maydis* sequences (6-42 %), while it showed identity of less than 5 % with CY52, CYP86 and CYP94 families

Most eukaryotic P450s are lodged in the endoplasmic reticulum. A hydrophobicity analysis was done using the algorithm of Kyte and Doolittle (1982) with a window of 19 amino acids. Two hydrophobic segments were detected in the N-terminal (Fig. 11) indicating that CYP557A1 is a membrane bound protein.

Alignment with other P450s clearly indicated the conserved helix I (HI) and heme-binding regions (HR2) (Fig. 11 &12). Helix I has conserved Ala, Gly and Thr residues while the HR2 has conserved amino acids Gly, Arg, Cys and Gly. The Cys is regarded as a fifth ligand of the heme iron (Werck-Reichhart *et al.*, 2000a, 2000b). A comparison of

CYP557A1 with omega hydroxylases of which the catalytic sites have already been characterized, showed that CYP557A1 has the conserved Phe^{142} and Leu^{520} residues which are distinguishing features of omega hydroxylases (Fig. 12). These conserved Phe and Leu residues in omega hydroxylases were modeled from the sequence of CYP102 of *Bacillus megaterium* and proven to be important in the catalytic site (Ravichandran *et al*., 1993; Kahn *et al.* 2001). CYP102 is the only bacterial omega hydroxylase which is closer to eukaryotic omega hydroxylases than to bacterial P450s. The X-ray structure of CYP102 (Ravichandran *et al*., 1993) showed that the Leu^{437} residue is situated in the SRS-6 (substrate recognition site) domain (Gotoh, 1992), while the aromatic side chain of Phe^{87} limits and/or controls substrate access to the heme iron and thus determines regioselectivity of oxygen attack. When Phe^{87} was replaced with Val, CYP102 lost its ability to perform hydroxylation of saturated carbons (ω and ω-1 hydroxylations) but became a very good catalyst for epoxidation of arachdonic acid (Graham-Lorence *et al.*, 1997). It is interesting that alignment of the fatty acid hydroxylases belonging to the CYP52, CYP86, CYP94, CYP557 and CYP102 families revealed a further 18 amino acid residues that are conserved in these fatty acid hydroxylases but not in P450s with other activities such as benzoate and phenyl acetate hydroxylation or monoterpene hydroxylation (Fig. 12). Noteworthy are the 6 amino acid residues which are conserved in the other fatty acid hydroxylases, but not in CYP557A1. Once activity of CYP557A1 has been characterized these amino acid residues might be of interest for future modeling studies. Three of these residues have replaced residues in the highly conserved Phe-Arg-Pro-Glu-Arg-Trp (FRPERW) sequence upstream from the HR2 domain which is referred to as the "meander" (Werck-Reichhart *et al.*, 2000a; 2000b). This sequence is highly conserved in many P450s from plants and fungi and is probably involved in heme-binding. In CYP557A1 this sequence is significantly different namely Trp-Lys-Pro-Ser-Arg-Trp (WKPSRW). In the rat fatty acid hydroxylases belonging to the CYP4A family this sequence is Phe-Asp-Pro-Ser-Arg-Phe (FDPSRF), thus indicating that it is possible for an active fatty acid hydroxylase to have different amino acid residues in this region (Hardwick *et al.*, 1987).

Table 3. Comparison of plant and fungal omega-hydroxylases (% amino acids identity)

	CYP86	CYP94	CYP52	CYP63	CYP557	UST-MAY
CYP86	>40					
CYP94	30-40	>40				
CYP52	20-25	19-22	>40			
CYP63	18-20	18-20	27-30	>40		
CYP557	24-26	25-27	17-22	17-19	100	
UST-MAY	23-26	25-28	22-24	19-21	28-33	26-38

```
cyp94a2   1         MEL------------ETLISWLL----FSTSLFWFLFLATK-----TKSKPPKTPSST 37
cyp94a1   1      MFQFHL------------EVLLPYLLPL--LLLILPTTIFFLTK-----PNNKVSSTSTNN 42
cyp52a4   1   MSVSFVHNVLEVVTPYVEYYQENLTKWYILIPTILLTLPFLSILHTKY--LEYKFNAKPLTNFA 62
cyp52a13  1      MTVHDI--IATY--------FTKWYVIVPLALIAYRVLDYFYGRY--LMYKLGAKPFFQKQ 49
cyp86a1   1      MEALNS-------------ILTGYAVA---ALSVYALWFYFLSR--------R----LTGP 33
cyp557a1  1      MLALVCVG-----------LFLYLLCKYRSHAIATSDRTDLAGPKGWPLMGVFKFPLPTLS 50

cyp94a2   38  TNTPIPKSYPIFGSAFS-LLANFHRR--IQWTSDILQTIPSSTFVLHRPFGARQVFTAQPAVVQ 98
cyp94a1   43  NIITLPKSYPLIGSYLS-FRKNLHRR--IQWLSDIVQISPSATFQLDGTLGKRQIITGNPSTVQ 103
cyp52a4   63  QDYSFGVITPLMLMYFK-WHGTVMEFACNVWNNKFLVLNGNVRTVGLRIMGLNIIETTDPENVK 125
cyp52a13  50  TDGCFGFKAPLELLKKK-SDGTLIDFTLQRIHDLD---RPDIPTFTFPVFSINLVNTLEPENIK 109
cyp86a1   34  KVLPFVGSLPYLIANRSRIHDWIADN--LRATGGTYQTCTMVIPFVAKAQG-FYTVTCHPKNVE 94
cyp557a1  51  VTYVVVDVTEKTLSGNLIFYATQSSTRGLEWGVRFVKTYGHAWTLSVPGLR-LIDISTRPDWIE 113

                                          ↓
cyp94a2   99  HILRTNFTCYGKGLTFYQSINDFLGDGIFNADGESWKFQRQISSHEFNTRSLRKFVETVVDVEL 162
cyp94a1   104 HILKNQFSNYQKGTTFTNTLSDFLGTGIFNTNGPNWKFQRQVASHEFNTKSIRNFVEHIVDTEL 167
cyp52a4   126 AILATQFNDFSLG-TRHDFLYSLLGDGIFTLDGAGWKHSRAMLRPQF----AREQVAHVKLLEP 184
cyp52a13  110 AILATQFNDFSLG-TRHSHFAPLLGDGIFTLDGAGWKHSRSMLRPQF----AREQISHVKLLEP 168
cyp86a1   95  HILKTRFDNYPKGPMWRAAFHDLLGQGIFNSDGDTWLMQRKTAALEFTTRTLRQAMARWVNGTI 158
cyp557a1  114 YVQKTNYHNYVKGSAFRNRLSDIIGNGIFVTDGRVWQFQRKVTSHIFTGRSFQDAICPAIAEEL 177
                □                     ■ ■ ■      ■          ■

cyp94a2   163 SDRLVPVLSQASNSQTTLDFQDILQRLTFDNICMIAFGYDPEYLLP-SLPEIPFAKAFD---ES 222
cyp94a1   168 TNRLIPILTSSTQTNNILDFQDILQRFTFDNICNIAFGYDPEYLTP-STNRSKFAEAYE---DA 227
cyp52a4   185 HVQVLFKHVRKSQ-GKTFDIQELFFRLTVDSSTEFLFGGSVESLRDASIGMVPSTKNIAGR-EE 246
cyp52a13  169 HVQVFFKHVRKAQ-GKTFDIQELFFRLTVDSATEFLFGESVESLRDESIGMSINALDFDGK-AG 230
cyp86a1   159 KNRLWLILDRAVQNNKPVDLQDLFLRLTFDNICGLTFGKDPETLSL-DLPDNPFSVAFD---TA 218
cyp557a1  178 RSLNKLLDNYADT-GETVDLQDVFYRFTFQAFGRFAFNLDLRNLEI-DHKPVPFTTAFDQCQKP 239
                                         ■ □             ■□

cyp94a2   223 SQLSIERLNALIPLLWKVKRFLNIGVERQLKEAVAEVRGLATKIVKNKKKELKEKALQSESESV 286
cyp94a1   228 TEISSKRFRLPLPIIWKIKKYFNIGSEKRLKEAVTEVRSFAKKLVREKKRELEEKSS---LETE 288
cyp52a4   247 FADAFNYSQTYNAYRFLLQQFYWILNGSKFNKSIKTVHKFADFYVQKALSLTDDDLE---KQEG 307
cyp52a13  231 FADAFNYSQNYLASRAVMQQLYWVLNGKKFKECNAKVHKFADYYVNKALDLTPEQLE---KQDG 291
cyp86a1   219 TEATLKRL-LYTGFLWRIQKAMGIGSEDKLKKSLEVVETYMNDAIDARKNSPSDDLL-----SR 276
cyp557a1  240 SLDASSTRSGAFASALPLPAVEYQTPPSCFRLCLRNDDKRAALLAKKDAKVPSDLLT------M 297

cyp94a2   287 DLLSRFLSSGHSDESFVTDMVISIILAGRDTTSAALTWFFWLLSKHSHVENEILKEITGKS--- 347
cyp94a1   289 DMLSRFLSSGHSDEDFVADIVISFILAGKDTTSAALTWFFWLLWKNPRVEEEIVNELSKKS--- 349
cyp52a4   308 YVFLYELAKQTRDPKVLRDQLLNILVAGRDTTAGLLSFLFFELSRNPTVFEKLKEEIHNRFGAK 371
cyp52a13  292 YVFLYELVKQTRDKQVLRDQLLNIMVAGRDTTAGLLSFVFFELARNPEVTNKLREEIEDKFGLG 355
cyp86a1   277 FLKKRDVNGNVLPTDVLQRIALNFVLAGRDTSSVALSWFFWLVMNNREVETKIVNELSMVLKET 340
cyp557a1  298 FMTARYDDSRPLSRVEIRDAIINLLLAGRDTTAQSLTWALYRLVDNPAHQRLVREEICGQLAGG 361
                                        ▲▲  ▲    ■                   ■
```

```
cyp94a2    348 ----------ETVGYDEVKDMVYTHAALCESMRLYPPLPVDTKVAVHDDVLPDG--------TL 393
cyp94a1    350 ----------ELMVYDEVKEMVYTHAALSESMRLYPPVPMDSKEAVNDDVLPDG--------WV 395
cyp52a4    372 -----EDARVEEITFESLKLCEYLKACVNEALRVYPSVPHNFRVATRNTTLPRGGGKDGMSPIA 430
cyp52a13   356 -----ENASVEDISFESLKSCEYLKAVLNETLRLYPSVPQNFRVATKNTTLPRGGGKDGLSPVL 414
cyp86a1    341 RGNDQEKWTEEPLEFDEADRLVYLKAALAETLRLYPSVPQDFKYVVDDDVLPDG--------TF 396
cyp557a1   362 ---------TEPVTLENVKNLVQTQAAHLESLRLHPPVPRIVKQAVKDDALPNGG-------PL 409

cyp94a2    394 VKKGWRVTYHIYAMGRSEKIWGPDWAEFRPERWLSRD-----EV-GKWSFVGIDYYSYPVFQAG 451
cyp94a1    396 VKKGTIVTYHVYAMGRMKSLWGDDWAEFRPERWLEKD-----EVNGKWVFVGRDSYSYPVFQAG 454
cyp52a4    431 IKKGQNVMYTISATHRDPSIYGEDANVFRPERWFEPE-----TRKLGWAYVP--------FNGG 481
cyp52a13   415 VRKGQTVIYGVYAAHRNPAVYGKDALEFRPERWFEPE-----TKKLGWAFLP--------FNGG 465
cyp86a1    397 VPRGSTVTYSIYSIGRMKTIWGEDCLEFRPERWLTADGERFETPKDGYKFVA--------FNAG 452
cyp557a1   410 IKAGEFIRLSDWALGRNEEVWGADAKEWKPSRWIDDE-----GRPIQYSQWKAHF-----FNGG 463

cyp94a2    452 PRVCIGKEMAFLQMKRVVAGIMGRFRVVPAM-------------VEGIEPE-YTAHFTSVMKGG 501
cyp94a1    455 PRVCLGKEMAFMQMKRIVAGIVGKFKVVPEA-------------HLAQEPG-FISFLSSQMEGG 504
cyp52a4    482 PRICLGQQFALTEASYVTVRLLQEFHTLTQD-------------ANTRYPPRLQNSLTVSLCDG 532
cyp52a13   466 PRICLGQQFALTEASYVTVRLLQEFAHLSMD-------------PDTEYPPKKMSHLTMSLFDG 516
cyp86a1    453 PRTCLGKDLAYNQMKSVASAVLLRYRVFPVP-------------GHRVEQK---MSLTLFMKNG 500
cyp557a1   464 PRICLGKSLATLEGVAVIANLLHRYNIAFAPGWWENVEKTGRIAGDLEETPLYGPALTLPMKSP 527

cyp94a2    502 FPVKIEKRSPLV   513
cyp94a1    505 FPVTIQKRDS     514
cyp52a4    533 --ANIQMY       538
cyp52a13   517 --ANIEMY       522
cyp86a1    501 LRVYLQPRGEVLA  513
cyp557a1   528 FWVTISRR       535
```

Figure 12. Alignment of *Rhodotorula* CYP557A1 with fatty acid hydroxylases from plants (CYP94A1 and CYP94A2 from *Vicia sativa*; CYP86A1 from *Arabidopsis thaliana*) and yeasts (CYP52A4 and CYP52A13 from *Candida maltosa* and *Candida tropicalis* respectively). Conserved sequences are shown in red and sequences that are similar are shown in dark green. Closed triangles indicate amino acids that were conserved when 18 cytochrome P450 monooxygenases from plants, fungi and bacteria with fatty acid hydroxylase activity (CYP52, CYP86, CYP94, CYP557 and CYP102), benzoate hydroxylase activity (CYP53), monoterpene hydroxylase activity (CYP71) and phenyl acetic acid hydroxylase activity were aligned. Closed squares indicate amino acids that were conserved in ten different proven fatty acid hydroxylases from yeasts (CYP52) and plants (CYP86 and CYP94) as well as in CYP557A1 and in CYP102 from *Bacillus megaterium*. Open squares indicate amino acids that were conserved in the other fatty acid hydroxylases but not in CYP557A1. The underlined sequence shows the position for SRS-6 implicated in chain length specificity. The diamond headed arrow indicates the position of the highly conserved phenylalanine implicated in chemoselectivity in CYP102 (Ravichandran *et al*., 1993). The triangle headed arrow shows the position for the conserved Leu residue implicated in regioselectivity in CYP94s (Kahn *et al*., 2001). The sequences were aligned by DNAssist version 2.0.

3.4 Discussion

Rhodotorula sp. CBS 8446 is a basidiomycetous yeast which can grow on decane and limonene vapours. No alkane or fatty acid hydroxylase has been cloned from a basidiomycetous yeast. The alkane and fatty acid hydroxylating P450s from 5 alkane utilizing ascomycetous yeasts have been sequenced and classified into the CYP52 family (Iida *et al.,* 1998; 2000; Lottermoser *et al*., 1996; Schunck *et al* 1989; Ohkuma *et al.*, 1991a, 1991b; Sanglard *et al*., 1987; Sanglard and Loper, 1989; Seghezzi *et al*., 1991; 1992, Craft *et al*., 2003; Yadav and Loper, 1999; http://drnelson.utmem.edu/biblioC.html). Primers based on the sequences of the helix I (HI) and heme-binding (HR2) domains of 15 different members of the CYP52 family were used to amplify a gene fragment from *Rhodotorula* sp. CBS 8446. Sequence analyses of the amplified gene fragment showed that it belonged to a cytochrome P450.

The ordinary PCR technique has limitation, allowing only DNA amplification of fragments that are flanked by primers based on known sequences. Therefore inverse PCR (IPCR) was used to amplify the full-length gene sequence of the new P450 gene. IPCR is based on inversion of the sequence of interest by circularisation and thus can amplify a sequence that lies outside the boundaries of a known sequence using primers based on the known sequence but orientated in opposite directions (Triglia *et al*., 1988; Arand *et al*., 1999).

Since it was found that the gene fragment was interrupted by introns, the 3' RACE technique (Frohman *et al*., 1988) was used to isolate the downstream part of the cDNA. RT-PCR using gene specific walking primers yielded the full-length cDNA sequence that included the start and stop codons. Comparison with other P450s showed that the sequenced gene codes for a novel fatty acid hydroxylase. Because the deduced protein showed less that 40% homology with previously identified P450s it was classified into a new P450 family and given the number CYP557A1.

Sequence analysis of the cDNA revealed that the genomic DNA is interrupted by 13 introns. Four of the 13 introns did not conform to the canonical GT/AG rule (Mount *et al*., 1982). This observation is not surprising, since it has been observed that the epoxide hydrolase encoding gene from *Rhodotorula glutinis* also deviated from the canonical rule (Visser *et al*., 2000). Recently it was observed that in *Saccharomyces cerevisiae* software predictions identified intron containing regions, but failed to specify the correct splice sites and branch points (Davis *et al.,* 2000). Thus, the correct prediction of introns using software remains a problem, limiting our understanding of the relationship between the structure and function of eukaryotic genomes.

Alternatively spliced variants were observed in transcripts of the *CYP557A1* gene. However, the transcripts that retained introns failed to give an ORF. If the complete ORF can be restored under specific conditions; it might yield proteins with different substrate specificities. Alternative splicing has been reported for the *MATa1* gene of *S. cerevisiae* (Ner and Smith, 1989) and for the *Prp10* (pre-mRNA processing) gene of *S. pombe* (Habara *et al.*, 1998) as well as for a putative fatty acid hydroxylase from the fungus *Phanerochaete chrysosporium* (Yadav *et al.,* 2003). In the fungus *Aspergillus niger* the glucoamylases G1 and G2 are synthesized from the same mRNA transcript by differential splicing (Boel *et al.,* 1984). In this case the G1 transcript retains a stretch of 169 nucleotides that can be spliced out to produce G2. Alternative splicing is a common phenomenon in mammalian P450s where it generates the functional diversity of the P450s (Christmas *et al*., 2001) and where exon scrambling and skipping have been observed (Zaphiropoulos *et al*., 1996; 1997; Caldas *et al*., 1998). In the case of exon scrambling the exons are spliced at the correct sites but joined in an order different from that present in genomic DNA; while in exon-skipping, some exons are absent or skipped during differential splicing. The significance of alternative splicing in *Rhodotorula* genes remains to be determined.

Southern blot analysis indicated that there is only one copy of the *CYP557A1* gene present in the genome of *Rhodotorula* sp. CBS 8446 and that there is also no similar genes present. The presence of only a single fatty acid hydroxylase in *Rhodotorula* sp.

CBS 8446 is unusual, since in alkane-assimilating ascomycetous yeasts such as *C. maltosa*, *C. tropicalis* and *Y. lipolytica* eight to ten CYP52 isoforms have in each case been sequenced (Iida *et al.,* 1998; 2000; Schunck *et al* 1989; Ohkuma *et al.*, 1991a, 1991b; Sanglard *et al.*, 1987; Sanglard and Loper, 1989; Seghezzi *et al.*, 1991; 1992, Craft *et al.*, 2003). P450s belonging to the CYP52 multigene family can hydroxylate either fatty acids or alkanes or both. In the basidiomycetous white rot fungus two putative fatty acid hydroxylases have been identified and assigned to the CYP63 family (Yadav *et al.,* 2003). The results obtained from this study coincide with the results obtained from *S. cerevisiae* where the P45014DM (lanosterol 14α-demethylase) and ERG5 (C-22 sterol desaturase) that are involved in ergosterol biogenesis are encoded by only a single gene each (Kalb *et al.*, 1987; Skaggs *et al.*, 1996).

Blast searches of the NCBI databases showed that the CYP557A1 protein has a low but significant homology with omega hydroxylases from plants and yeasts. The fully characterized P450s with which it has the highest amino acid identity (24 - 27%) is the CYP86s and CYP94s from *A. thaliana* and *V. sativa* respectively. CYP557A1 has 17 – 22 % amino acid identity with the CYP52s from yeasts (Table 3). The CYP86s and CYP94s from plants differ from the CYP52s from ascomycetous yeasts in that they hydroxylate only fatty acids and show no activity towards n-alkanes (Benveniste *et al.*, 1998; Tijet *et al.*, 1998). It is therefore possible that CYP557A1 is also exclusively a fatty acid hydroxylase. Blast searches of the NCBI databases with CYP557A1 also yielded four unidentified proteins from the genome of *Ustilago maydis*. These four proteins showed higher amino acid identity with CYP557A1 (28 - 33 %) than any characterized P450 in the database. However, these four putative fatty acid hydroxylases showed only 26 – 38 % amino acid with each other, indicating that they probably belong to an additional four different P450 families.

A detailed analysis of the CYP557A1 protein revealed all the conserved residues of the helix I (HI) and heme-binding (HR2) domains which characterize P450 proteins. The HR2 has a conserved Cys that is regarded as a fifth ligand of the heme iron. The HI has been implicated in substrates recognition in P450s (Gotoh, 1992). A hydrophobicity

prediction revealed two hydrophobic segments in the N-terminal. These transmembrane segments are responsible for anchoring the cytochrome P450 proteins in the endoplasmic reticulum (ER) (Scheller *et al*., 1994; Menzel *et al*., 1996, Nelson and Strobel, 1988). Taken together, these features indicate that CYP557A1 is probably associated with microsomes. It has been observed that some P450 proteins have two transmembrane segments while some contain only one (Seghezzi *et al*., 1991; 1992). This feature gives the P450s a wide range of the substrate specificities.

Alignment of CYP557A1 with fatty acid hydroxylases belonging to the CYP52, CYP86, CYP94 and CYP102 families revealed 18 amino acid residues that are conserved in CYP557A1 and in these fatty acid hydroxylases, but not in P450s with other activities. One of these residues is Phe^{142} which corresponds to Phe^{87} in CYP102. The X-ray structure of CYP102 (Ravichandran *et al*., 1993) showed that the aromatic side chain of Phe^{87} limits and/or controls substrate access to the heme iron and thus determines regioselectivity of oxygen attack. The SRS-6 domain which is also involved in substrate binding could also be recognized and contains in CYP557A1 Leu^{520}-Thr^{521}-Leu^{522}. The Leu-Thr-Leu motive is also present in CYP102 (Ravichandran *et al*., 1993) and is common in fatty acid hydroxylases. It has been shown with the CYP94s of *V. sativa* and the CYP52s of *C. maltosa*, where there are changes in this conserved motive (Kahn *et al*., 2001; Zimmer *et al*., 1998), that these residues play a role in determining chain length specificity and regioselectivity (whether terminal or in-chain hydroxylation takes place). The alignment with the other fatty acid hydroxylases also revealed 6 amino acid residues which are conserved in the other fatty acid hydroxylases, but not in CYP557A1. Once activity of CYP557A1 has been characterized these amino acid residues might be of interest for future modeling studies.

This is the first report of an omega hydroxylase encoding gene from a basidiomycetous yeast. Its regulation and substrate- and regiospecificity have not yet been determined. Heterologous expression of this gene in *Yarrowia lipolytica* is in progress and might answer some of these questions.

3.5 References

Altschul, S F., Madden, T L., Schäffer, A A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D J. (1997). “Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs”. *Nucleic acids Res*. **25**: 3389-3402.

Arand, M., Hemmer, H., Dürk, H., Baratti, J., Archelas, A., Furstoss, R. and Oesch, F. (1999). Cloning and molecular characterization of a soluble epoxide hydrolase from *Aspergillus niger* that is related to mammalian microsomal epoxide hydrolase. *Biochem J* **344**: 273-280.

Benveniste, I., Tijet, N., Adas, F., Philipps, G., Salaun, JP. and Durst, F. (1998). *CYP86A1* from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid omega-hydroxylase. *Biochem Biophys Res Commun*. **243**: 688-693.

Bjellqvist, B., Hughes, GJ., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J-Ch., Frutiger, S. and Hochstrasser, DF. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis.* **14**: 1023-1031.

Black, SD. (1992). Membrane topology of the mammalian P450 cytochromes**.** *FASEB J*. **6**: 680-685.

Blasig, R., Mauersberger, S., Riege, P., Schunck, WH., Jockisch, W. Franke, P. and Müller, HG. (1988). Degradation of long- chain n-alkanes by yeast *Candida maltosa*. *Appl. Microbiol Biotechnol.* **28**: 589-597.

Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, KE. and Fiil, NP. (1984). Glucoamylases G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs. *EMBO J*. **3**: 1097-1102.

Caldas, C., So, CW., MacGregor, A., Ford, AM., McDonald, B., Chan, LC., Wiedemann, LM. (1998). Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. *Gene*. **208**: 167-176.

Capdevila, J., Saeki, Y. and Falck, JR. (1984). The mechanistic plurality of cytochrome P-450 and its biological ramifications. *Xenobiotica*. **14**: 105-118.

Christmas, P., Jones, JP., Patten, CJ., Rock, DA., Zheng, Y., Cheng, SM., Weber, BM., Carlesso, N., Scadden, DT., Rettie, AE. and Soberman, RJ. (2001). Alternative splicing determines the function of CYP4F3 by switching substrate specificity. J Biol Chem. **276**: 38166-38172.

Craft, DL., Madduri, KM., Eshoo, M. and Wilson, CR. (2003). Identification and characterization of the CYP52 family of *Candida tropicalis* ATCC 20336, important for the conversion of fatty acids and alkanes to alpha, omega-dicarboxylic acids. *Appl Environ Microbiol*. **69**: 5983-5991.

Davis, CA., Grate, L., Spingola, M. and Ares, M Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res*. **28**: 1700-1706.

Durham, DR. (1984). Initial reactions involved in the dissimilation of mandelate by *Rhodotorula graminis*. *J Bacteriol*. **160**: 778-780.

Durham, DR., McNamee, CG. and Stewart, DB. (1984). Dissimilation of aromatic compounds in *Rhodotorula graminis*: biochemical characterization of pleiotropically negative mutants. *J Bacteriol*. **160**: 771-777.

Eschenfeldt, WH., Zhang, Y., Samaha, H., Stols, L., Eirich, LD., Wilson, CR. and Donnelly, MI. (2003). Transformation of fatty acids catalyzed by cytochrome P450

monooxygenase enzymes of *Candida tropicalis*. *Appl Environ Microbiol*. **69**: 5992-5999.

Frohman, MA., Dush, MK., Martin, GR. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*. **85**: 8998-9002

Fujii, T., Nakamura, K., Shibuya, K., Tanase, S., Gotoh, O., Ogawa, T. and Fukuda, H. (1997). Structural characterization of the gene and corresponding cDNA for the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. *Mol Gen Genet*. **256**: 115-120.

Fukui, S. and Tanaka, A. (1981). Metabolism of alkanes by yeasts. *Adv. Biochem. Eng*. **19**: 217-237.

Gilewicz, M., Zacek, M., Bertrand, J-C. and Azoul ay E. (1979). Hydroxylase regulation in *Candida tropicalis* grown on alkanes. *Can. J. Microbiol*. **25**: 201-206.

Giniger, E., Varnum, SM. and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*. **40**: 767-774.

Gonzalez, FJ. (1988). The molecular biology of cytochrome P450s. *Pharmacol Rev*. **40**: 243-288.

Gonzalez, FJ. (1990). Molecular genetics of the P-450 superfamily. *Pharmacol Ther*. **45**: 1-38.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem*. **267**: 83-90.

Graham-Lorence, S., Truan, G., Peterson, JA., Falck, JR., Wei, S., Helvig, C. and Capdevila, JH. (1997). An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase. *J Biol Chem*. **272**: 1127-1135.

Guo, Z. and Sherman, F. (1996). 3'-end-forming signals of yeast mRNA. *Trends Biochem Sci*. **21**: 477-481.

Guo, Z., Russo, P., Yun, DF., Butler, JS. and Sherman, F. (1995). Redundant 3' end-forming signals for the yeast CYC1 mRNA. *Proc Natl Acad Sci U S A*. **92**: 4211-4214.

Habara, Y., Urushiyama, S., Tani, T. and Ohshima, Y. (1998). The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor, SAP155. *Nucleic Acids Res*. **26**: 5662-5669.

Hamilton, R., Watanabe, CK. and de Boer, HA. (1987). Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs. *Nucleic Acids Res*. **15**: 3581-3593.

Hardwick, JP., Song, BJ., Huberman, E. and Gonzalez, FJ. (1987). Isolation, complementary DNA sequence, and regulation of rat hepatic lauric acid omega-hydroxylase (cytochrome P-450LA omega). Identification of a new cytochrome P-450 gene family. *J Biol Chem*. **262**: 801-810.

Iida, T, Ohta, A. and Takagi, M. (1998). Cloning and characterization of an *n*-alkane-inducible cytochrome P450 gene essential for *n*-decane assimilation by *Yarrowia lipolytica*. *Yeast*. **14**: 1387-1397.

Iida, T., Sumita, T., Ohta, A. and Takagi, M. (2000). The cytochrome P450ALK multigene family of an *n*-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning

and characterization of genes coding for new CYP52 family members. *Yeast*. **16**: 1077-1087.

Kahn, RA., Le Bouquin, R., Pinot, F., Benveniste, I. and Durst, F. (2001). A conservative amino acid substitution alters the regiospecificity of CYP94A2, a fatty acid hydroxylase from the plant *Vicia sativa*. *Arch Biochem Biophys*. **391**: 180-187.

Kalb, VF., Woods, CW., Turi, TG., Dey, CR., Sutter, TR. and Loper, JC. (1987). Primary structure of the P450 lanosterol demethylase gene from *Saccharomyces cerevisiae*. *DNA*. **6**: 529-537.

Käppeli O. (1986). Cytochromes P-450 of yeasts. *Microbiol Rev*. **50**: 244-258.

Kargel, E., Menzel, R., Honeck, H., Vogel, F., Bohmer, A. and Schunck, WH. (1996). *Candida maltosa* NADPH-cytochrome P450 reductase: cloning of a full-length cDNA, heterologous expression in *Saccharomyces cerevisiae* and function of the N-terminal region for membrane anchoring and proliferation of the endoplasmic reticulum. *Yeast*. **12**: 333-348.

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

Langford, CJ., Klinz, FJ., Donath, C. and Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell*. **36**: 645-653.

Le Bouquin, R., Pinot, F., Benveniste, I., Salaun, JP. and Durst, F. (1999). Cloning and functional characterization of *CYP94A2*, a medium chain fatty acid hydroxylase from *Vicia sativa*. *Biochem Biophys Res Commun*. **261**:156-162.

Lottermoser, K., Schunck, WH. and Asperger, O. (1996). Cytochrome P450 of the sophorose lipid-producing yeast *Candida apicola*: heterogeneity and polymerase chain reaction - mediated cloning of two genes. *Yeast*. **12**: 565-575.

Mathews, C. K. and van Holde, K. E. (1990). Biochemistry, The Benjamin/ Cummings Publishing Company, Inc., CA.

Mauersberger, S., Ohkuma, M., Schunck, W-H. and Takagi, M. (1996). *Candida maltosa* In Wolf K (ed.), Nonconventional Yeasts in Biotechnology, Springer - Verlag Berlin pp. 411- 579.

Menzel, R., Kargel, E., Vogel, F., Bottcher, C. and Schunck, WH. (1996). Topogenesis of a microsomal cytochrome P450 and induction of endoplasmic reticulum membrane proliferation in *Saccharomyces cerevisiae*. *Arch Biochem Biophys*. **330**: 97-109.

Moleleki, N. (1998). The involvement of cytochrome P450 monooxygenases in the hydroxylation of monoterpenes by yeasts (M.Sc. thesis).

Mount, SM. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res*. **10**: 459-472.

Nebert, DW., Nelson, DR., Coon, MJ., Estabrook, RW., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, FJ., Guengerich, FP., Gunsalus, IC., Johnson, EF., Loper, JC., Sato, R., Waterman, MR. and Waxman, DJ. (1991). The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. DNA *Cell Biol*. **10**:1-14.

Nelson, DR. (1999). Cytochrome P450 and the individuality of species. *Arch Biochem Biophys*. **369**: 1-10.

Nelson, DR. and Strobel, HW. (1988). On the membrane topology of vertebrate cytochrome P-450 proteins. *J Biol Chem.* **263**: 6038-6050.

Ner, SS. and Smith, M. (1989). Role of intron splicing in the function of the *MATa*1 gene of *Saccharomyces cerevisiae*. *Mol Cell Biol***. 9**: 4613-4620.

Ohkuma, M., Hikiji, T., Tanimoto, T., Schunck, W-H., Müller, H-G., Yano, K. and Takagi, M. (1991a). Evidence that more than one gene encodes n-alkane-inducible cytochrome P-450s in *Candida maltosa*, found by two-step gene disruption. *Agric. Biol. Chem*. **55**: 1757-1764.

Ohkuma, M., Tanimoto, T., Yano, K. and Takagi, M. (1991b). *CYP*52 (cytochrome P450alk) multigene family in *Candida maltosa*: molecular cloning and nucleotide sequence of the two tandemly arranged genes. *DNA Cell Biol*. **10**: 271-282.

Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J. and Eirich, LD. (1992). Metabolic engineering of *Candida tropicalis* for the production of long-chain dicarboxylic acids. *Biotechnology* (N Y). **10**: 894-898.

Ravichandran, KG., Boddupalli, SS., Hasermann, CA., Peterson, JA. and Deisenhofer J. (1993). Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science*. **261**: 731-736.

Rudolph, H. and Hinnen, A. (1987). The yeast PHO5 promoter: phosphate-control elements and sequences mediating mRNA start-site selection. *Proc Natl Acad Sci U S A*. **84**: 1340-1344.

Sambrook, J., Fritsch, EF. and Maniates F. (1989). Molecular cloning: a laboratory manual, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanglard, D. and Loper, C. J. (1989). Characterization of the alkane-inducible cytochrome P450 (P450alk) gene from the yeast *Candida tropicalis*: identification of a new P450 gene family. *Gene*. **76**: 121-136.

Sanglard, D., Chen, C. and Loper, JC. (1987). Isolation of the alkane inducible cytochrome P450 (P450alk) gene from the yeast *Candida tropicalis*. *Biochem Biophys Res Commun*. **144**: 251-257.

Scheller, U., Kraft, R., Schroder, KL., Schunck, WH. (1994). Generation of the soluble and functional cytosolic domain of microsomal cytochrome P450 52A3. *J Biol Chem*. **269**:12779-12783.

Scheller, U., Zimmer, T., Becher, D., Schauer, F. and Schunck, WH. (1998). Oxygenation cascade in conversion of n-alkanes to alpha,omega-dioic acids catalyzed by cytochrome P450 52A3. *J Biol Chem*. **273**: 32528-32534

Schunck, WH., Kargel, E., Gross, B., Wiedmann, B., Mauersberger, S., Kopke, K., Kiessling, U., Strauss, M., Gaestel, M. and Muller, HG. (1989). Molecular cloning and characterization of the primary structure of the alkane hydroxylating cytochrome P-450 from the yeast *Candida maltosa*. *Biochem Biophys Res Commun*. **161**: 843-850.

Schunck, WH., Mauersberger, S., Huth, J., Riege, P., and Müller, HG. (1987). Function and regulation of cytochrome P-450 in alkane- assimilating yeast. *Arch Microbiol*. **147**: 240-244.

Seghezzi, W., Meili, C., Ruffiner, R., Kuenzi, R., Sanglard, D. and Fiechter, A. (1992). Identification and characterization of additional members of the cytochrome P450 multigene family CYP52 of *Candida tropicalis*. *DNA Cell Biol.* **11**: 767-780.

Seghezzi, W., Sanglard, D. and Fiechter, A. (1991). Characterization of a second alkane - inducible cytochrome P450- encoding gene, *CYP52A2*, from *Candida tropicalis*. *Gene*. **106**: 51-60.

Skaggs, BA., Alexander, JF., Pierson, CA., Schweitzer, KS., Chun, KT, Koegel, C., Barbuch, R. and Bard M. (1996). Cloning and characterization of the *Saccharomyces cerevisiae* C-22 sterol desaturase gene, encoding a second cytochrome P-450 involved in ergosterol biosynthesis**.** *Gene*. **169**: 105-109.

Smith, MW. (1988). Structure of vertebrate genes: a statistical analysis implicating selection*.* *J Mol Evol*. **27**: 45-55.

Sutter, TR., Sanglard, D., Loper, JC. and Sangard, D. (1990). Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from *Candida tropicalis*. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins. *J Biol Chem*. **265**: 16428-16436.

Teem, JL., Abovich, N., Kaufer, NF., Schwindinger, WF., Warner, JR., Levy, A., Woolford, J., Leer, RJ., van Raamsdonk-Duin, MM., Mager, WH., Planta, RJ., Schultz, L., Friesen, JD., Fried, H. and Rosbash, M. (1984). A comparison of yeast ribosomal protein gene DNA sequences. *Nucleic Acids Res*. **12**: 8295-8312.

Tijet, N., Helvig, C., Pinot, F., Le Bouquin, R., Lesot, A., Durst, F., Salaun, JP. and Benveniste, I. (1998). Functional expression in yeast and characterization of a clofibrate-inducible plant cytochrome P-450 (*CYP94A1*) involved in cutin monomers synthesis. *Biochem J*. **332**: 583-589.

Triglia, T., Peterson, MG. and Kemp, D. (1988). A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res*. **16**: 8186-8187.

Visser, H., Vreugdenhil, S., de Bont, JA. and Verdoes, JC. (2000). Cloning and characterization of an epoxide hydrolase-encoding gene from *Rhodotorula glutinis*. *Appl Microbiol Biotechnol.* **53**: 415-419.

Vogel, F., Gengnagel, C., Kärgel, E., Müller, H. and Schunck, W-H. (1992). Immunocytochemical localization of alkane - inducible cytochrome P-450 and its NADPH-dependent reductase in the yeast *Candida maltosa*. *Eur. J. Cell Biol*. **57**: 285-291.

Werck-Reichhart, D., Feyereisen, R. (2000a). Cytochromes P450: a success story. *Genome Biology*. **1**: 3003.1-3003.9.

Werck-Reichhart, D., Hehn, A. and Didierjean, L. (2000b). Cytochromes P450 for engineering herbicide tolerance. *Trends Plant Sci*. **5**: 116-23.

Wright, JD. and Ratledge C. (1991). Isolation of two *Rhodotorula rubra* strains showing differences in the degradation of aromatic compounds. *Appl. Microbiol. Biotechnol*. **35**: 94-99.

Yadav, JS. and Loper, JC. (1999). Multiple p450alk (cytochrome P450 alkane hydroxylase) genes from the halotolerant yeast *Debaryomyces hansenii*.*Gene*. **226**: 139-146.

Yadav, JS., Soellner, MB., Loper, JC. and Mishra, PK. (2003). Tandem cytochrome P450 monooxygenase genes and splice variants in the white rot fungus *Phanerochaete chrysosporium*: cloning, sequence analysis, and regulation of differential expression. *Fungal Genet Biol*. **38**: 10-21.

Zaphiropoulos, PG. (1996). Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc Natl Acad Sci U S A*. **93**: 6536-6541.

Zaphiropoulos, PG. (1997). Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol Cell Biol*. **17**: 2985-2993.

Zaret, KS. and Sherman, F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell*. **28**: 563-573.

Zimmer, T., Scheller, U., Takagi, M. and Schunck, WH. (1998). Mutual conversion of fatty-acid substrate specificity by a single amino-acid exchange at position 527 in P-450Cm2 and P-450Alk3A. *Eur J Biochem*. **256**: 398-403.

Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. and Schunck, WH. (1996). The CYP52 multigene family of *Candida maltosa* encodes functionally diverse *n*-alkane-inducible cytochromes P450. *Biochem Biophys Res Commun*. **224**:784-789.

CHAPTER 4

FUNCTIONAL EXPRESSION OF CYTOCHROME P450 MONOOXYGENASES FROM *Rhodotorula* SPP. IN *Yarrowia lipolytica*

4.1 Introduction

Cytochrome P450 proteins absorb UV radiation at 450 nm when carbon monoxide (CO) is bound to the heme group (Omura and Sato, 1964). The cytochrome P450 protein family comprises a superfamily of ubiquitous heme-thiolate proteins that oxidize a wide range of endogenous and xenobiotic substrates (Porter and Coon, 1991). These monooxygenases contribute to vital processes in the cell such as carbon source assimilation, biosynthesis of hormones and detoxification of drugs and carcinogens (Werck-Reichhart and Feyereisen, 2000). In prokaryotes the P450s are soluble proteins, while in eukaryotes they are usually bound to the endoplasmic reticulum or inner mitochondrial membranes.

Many researchers are working on cytochrome P450s to determine their functions and improve their catalytic activities. Such studies often involve heterologous expression in a foreign host. In order to express a cytochrome P450 in a foreign host, an efficient electron transfer system is required that is compatible with the host as well as the expressed enzyme. NADPH cytochrome P450 reductases (CPR) transfer electrons from NADPH to the cytochrome P450s (Sutter *et al*., 1990). Thus, in many cases the co-expression of a NADPH-reductase and a P450 is necessary. In order for the heterologous expression system to be functional the P450 and NADPH reductase (CPR) should be in the correct ratio (Dong and Porter, 1996; Backes and Kelley, 2003).

Most studies report the concentration of cytochrome P450 protein, the level of reductase activity as well as the hydroxylation activity of the reconstituted membrane fraction. The concentration of cytochrome P450 protein is determined by measuring the reduced CO-difference spectrum in microsomal fractions or whole cells (Omura and Sato, 1964). The concentration of P450 reductase is determined by following the disappearance of NADPH as cytochrome *c* is reduced. Monooxygenase activity of the cytochrome P450 is determined by monitoring the disappearance of NADPH in a reconstituted system at 340 nm or alternatively by measuring the rate of product formation.

Many studies have concentrated on expression of mammalian cytochrome P450 genes in *Escherichia coli*. The bacterial expression system was found to be easy to handle, relatively inexpensive and yielded high protein concentrations. However, *E. coli* does not have endogenous reductases (Dong and Porter, 1996; Blake *et al*., 1996). Thus in two independent studies the bovine CYP17α and a rat omega hydroxylase CYP4A1 were fused to the cDNA from a rat CPR and expressed in *E. coli* (Fisher *et al.,* 1992; Barnes *et al*., 1991). It was observed that the membrane fractions expressed the NADPH dependent activity for 17α-hydroxylation of steroids and hydroxylation of fatty acids without any addition of NADPH-reductase flavoproteins, an indication of a functional fused enzyme system.

As more human cytochrome P450s were co-expressed with human CPR in *E. coli* and biotransformations were carried out with whole cells and cell fractions (Iwata *et al*., 1998; Blake *et al*., 1996), it became evident that the P450 protein expression level was often decreased by co-expression with CPR. It was speculated that the capacity in *E. coli* was not enough to synthesize two proteins or that the P450 might be degraded by active oxygen produced by the CPR. The expression level of CPR differed depending on the P450 with which it was being co-expressed. In the co-expressed systems the oxidation of the relevant substrates were however efficient and comparable with reconstituted systems (Iwata *et al*., 1998). These results demonstrated that mammalian P450 systems were catalytically active in *E. coli* and that the system could be used in biocatalysis and in drug development (Blake *et al*., 1996; Iwata *et al*., 1998).

Mammalian cytochrome P450s were not only expressed *in E. coli*, several expression studies were also performed in the yeast *Saccharomyces cerevisiae*. Bovine CYP17α and rat CYP1A1 were fused with yeast CPR and expressed in *S. cerevisiae* (Sakaki *et al*., 1994; Murakami *et al*., 1990). It was observed that the activity increased up to 25 times as compared with clones expressing the P450 alone.

In one such a study Liu *et al*, (1998) fused CYP1A1 and yeast CPR and expressed it in *Saccharomyces diastaticus*. Immobilized whole cells were used for bioconversion of acetanilide to *p*-acetaminophene within reticulated polyurethane foam biomass support particles (BSP) in a bioreactor. It was observed that the expression of the P450 and CPR in cells immobilized within the BSP was very high in comparison with cells freely suspended. This study offers the potential for development of efficient processes for the production of a wide variety of useful compounds.

The cytochrome P450s belonging to the CYP52 family are responsible for alkane and fatty acid hydroxylation in alkane-assimilating yeasts. These monooxygenases represent one family of fungal P450s which has been studied extensively. To determine the substrate specificities and regio-selectivities of some of these enzymes the cytochrome P450s and P450 reductases were co-expressed in *S. cerevisiae* and insect cells. The enzymes displayed different specificities and regioselectivities towards alkanes and fatty acids, and demonstrated an overlap in activities (Zimmer *et al*., 1998; 1995; 1996; Scheller *et al*., 1998; Schunck *et al.,* 1991; Craft *et al*., 2003; Eschenfeldt *et al*., 2003). However, it was observed that some proteins obtained from *S. cerevisiae* have higher molecular weight than the native proteins. This was possibly due to hyperglycosylation in *S. cerevisiae* (Sanglard and Loper, 1989).

Yarrowia lipolytica is a non-conventional yeast which can be explored as an alternative host for heterologous expression of cytochrome P450s. It is a dimorphic yeast that secretes enzymes such as proteases, lipases, esterases and RNase (Barth and Gaillardin, 1996; 1997) and which has successfully been used for the expression of various extracellular enzymes (Pignede *et al*., 2000a; Park *et al.*, 1997; 2000; Nicaud *et al*., 1989; 2002). A large number

of genetic tools are available for *Y. lipolytica* (Barth and Gaillardin, 1997). Genes with strong promoters such as *XPR2, ICL,* and *POX2* have been characterized in *Y. lipolytica* (Davidow *et al*., 1987; Nicaud *et al*., 1989; Barth and Scheuber, 1993; Wang *et al*., 1999) and these promoters are used in expression vectors. The isocitrate lyase (*ICL*) promoter is induced by acetate, ethanol and fatty acids, while the *POX2* promoter is induced by alkanes and fatty acids (Pignede *et al*. 2000b; Barth and Scheuber, 1993). *XPR2* has a strong promoter that is induced whenever the pH of the medium is above 6 and if there is carbon and nitrogen limitation (Davidow *et al*., 1987; Nicaud *et al*., 1989). It also requires a high level of peptones for full induction. To improve the *XPR2* promoter the constitutive synthetic promoter hp4d was developed (Madzak *et al*., 1999; 2000). The hp4d promoter consists of four tandem copies of the upstream activating sequence 1 (UAS1 region) from the *XPR2* promoter cloned upstream of a minimal *LEU2* promoter. This promoter shows quasi-constitutive activity with the highest induction in stationary phase.

A large number of vectors contain zeta sequences (retrotransposon sequences) that are tandem repeated in the genome of some *Y. lipolytica* strains (Juretzek *et al*., 2001). The zeta sequences are used as targeting sequences for integration during yeast transformation. Before transformation, the vectors are digested with *Not*I that cuts within zeta sequences so that the flanking regions could be generated for integration.

Nthangeni *et al*. (2004) recently used the expression vectors JMP62 and JMP64 to study the heterologous expression of the human *CYP1A1* gene in *Y. lipolytica*. The expression vector JMP62 is a single copy plasmid, while JMP64 is a multicopy plasmid. These vectors contain the strong, inducible *POX*2 promoter (Pignede *et al*., 2000a; 2000b) and are carrying two selective markers, *URA*3 and kanamycin. The kanamycin marker is used for selectionof transformants into bacteria. JMP62 has a non-defective *ura3*d1 marker while JMP64 has a defective *ura3*d4 marker. The *ura3d*1 marker is used for single integration while the defective *uar3*d4 marker is used for multiple integrations during yeast transformation. The copy numbers for JMP64 can range from 3-39 copies with an average of 10-13 copies/cell (Juretzek *et al*., 2001; Le Dall *et al*., 1994) when used in strains containing zeta sequences. Positive colonies are normally observed on uracil free plates after 3 days for cells

transformed with JMP62 while the transformants with JMP64 appear after 10 days. Nthangeni *et al*. (2004) used these vectors in the strain PO1d that does not contain zeta sequences. In such cases integration occurs randomly in the genome. The JMP21-*CPR* vector was used to add an additional copy of the *Y. lipolytica CPR* gene. This vector contains the strong *ICL* (isocitrate lyase) promoter driving the *CPR* gene and has *LEU*2 as a selection marker.

There is currently an increasing number of P450 monooxygenase encoding genes, also from other fungi, becoming available in databases. The functions and substrate specificities of these cytochrome P450s are in most cases not known. It was our purpose to investigate the expression of other fungal cytochrome P450s in *Y. lipolytica*. *Rhodotorula* spp. are basidiomycetous yeasts which can utilize substrates such as alkanes, alkylbenzenes, benzoate and monoterpenes as only carbon sources (Durham *et al*., 1984; Durham, 1984; Wright and Ratledge, 1991; Fujii *et al*., 1997; Moleleki, 1998). The sequences of two CYP genes are available from *Rhodotorula* spp., *CYP53B1* encoding a benzoate *para*-hydroxylase (Fujii *et al*., 1997) and *CYP557A1* the putative fatty acid omega-hydroxylase, which we recently isolated. In this study we investigated the expression of these two genes in *Y. lipolytica* E150, a strain containing zeta sequences, using the vectors JMP62 and JMP64. The vector JMP21-*CPR* was used to introduce an additional copy of the reductase under the strong *ICL* promoter.

4.2 Materials and Methods

4.2.1 Strains, vectors and media

Rhodotorula sp. CBS 8446 and *Yarrowia lipolytica* strain E150 (*Matb, his1, ura3-302, leu2-270, xpr2-322, XPR2^{P}::SUC2)* were maintained in the MIRCEN Yeast Culture Collection of the University of the Free State, South Africa. *Yarrowia lipolytica* E150-CPR was a generous gift from Dr. E. Setati, University of the Free State. The yeast cultures were maintained on YEPD agar [1% (w/v), yeast extract, 2% (w/v) peptone, 2% (w/v) glucose and 2% (w/v) agar]. *Escherichia coli* TOP 10 (F^{-} *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80

*lacZ*ΔM15 Δ*lacX74 recA1 deoR araD139* Δ(*ara-leu*)*7697 galU galK rpsL* (Str^r) *endA1 nupG* (Invitrogen) was used for plasmid preparations. *E. coli* carrying plasmids were grown on Luria-Bertani (LB) [1% (w/v), tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl and 2% (w/v) agar] supplemented with 100 μg/ml ampicillin or 50 μg/ml kanamycin. The cloning of PCR products was performed using pGEM-T Easy vector (Promega). Expression vectors JMP62, JMP64 and JMP21-CPR (Fig. 1) were a generous gift from Dr. J-M. Nicaud, Laboratoire Microbiologie et Génétique Moléculaire, Institute National Agronomique Paris-Grignon, France.

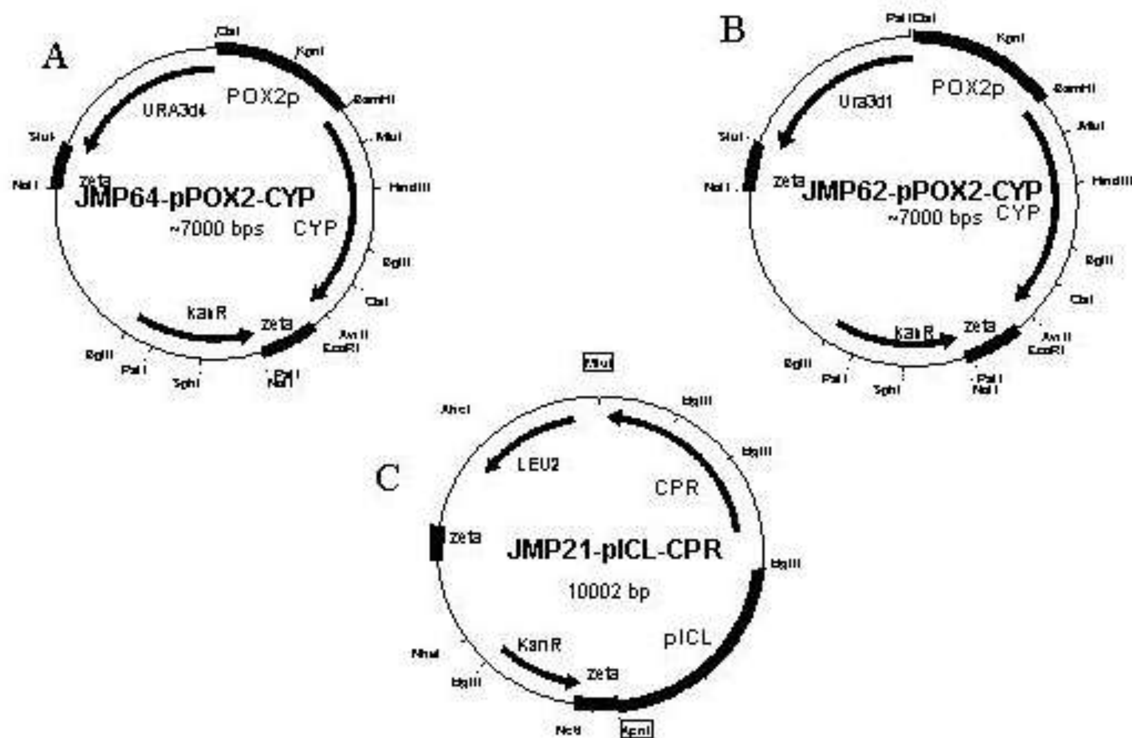


Figure 1. Schematic drawings showing the restriction maps of expression vectors. (**A**) JMP64-pPOX2, (**B**) JMP62-pPOX2 (**C**) JMP21-pICL-CPR. POX2 and ICL promoters are shown in the figures. The positions of zeta sequences are also shown in the figures. *LEU*2, *ura3d1, ura3d4* are used as selection markers during transformation of yeasts, while kanamycin is used as a selection marker during transformation of bacteria.

4.2.2 Isolation of total RNA from *Rhodotorula minuta* grown in chemical defined medium (CD) supplemented with L-phenylalanine

The *R. minuta* CBS 2177 cells were pre-grown in YEPD (1% yeast extract, 2% peptone, 2% glucose) for 16 hours. The cells were washed with physiological solution (0.9% NaCl) and transferred to chemically defined (CD) medium (Fujii *et al*., 1987): CD medium contained (per liter of deionized water): 20 g glucose, 5 g $(NH_4)_2SO_4$, 5.74 g KH_2PO_4, 9.26 g Na_2HPO_4, 1 g $MgSO_4.7H_2O$, 0.1 g $CaCl_2.H_2O$), 0.1 g NaCl, 0.5 mg *p*-aminobenzoic acid, 0.5 g thiamine-HCl, 50 mg Na-citrate, 30 mg $MnCl_2.4H_2O$, 20 mg $ZnCl_2$, 20 mg $FeCl_2.6H_2O$, 2 mg $CuSO_4.5H_2O$, 2 mg $CoCl_2.6H_2O$, 1 mg $Na_2MoO_4.2H_2O$ and 1 mg $K_2B_4O_7.xH_2O$, pH 7. For *CYP53B1* induction, glucose was replaced with L-phenylalanine as a carbon source at a concentration of 1 g/l.

The total RNA was isolated using the modified TRIzol (Invitrogen) method. The cells were frozen in liquid nitrogen and transferred to a pre-cooled mortar and ground to a fine powder. The fine powder was transferred to a pre-cooled eppendorf tube to approximately 0.5 ml. The ground cells were homogenized in 1 ml of TRIzol and incubated at room temperature for 5 minutes. Two hundred microliters of chloroform was added, vortexed vigorously for 15 seconds and incubated for 3 minutes at room temperature. Samples were centrifuged at 12000 x g for 15 minutes at 4°C and the supernatant was transferred to new tubes. The RNA was precipitated by adding 500 µl of isopropanol to the supernatant, vortexed and incubated at room temperature for 10 minutes. Samples were centrifuged at 12000 x g for 10 minutes and washed with 1 ml 70% ethanol. Samples were dried in the speedyVac and dissolved in 50 µl of formamide.

4.2.3 Isolation of total RNA and preparation of poly $(A)^+$ mRNA from *Rhodotorula* sp. CBS 8446 grown on decane

Fresh cells were streaked on YNB agar plates [2% agar and 0.67% YNB (Difco)]. The plates were placed in a desiccator together with a glass vial containing 200 µl of decane (Fluka) and a second vial containing 1 ml of water. The desiccator was tightly sealed and incubated

at 25°C for 9 days. The cells were scraped from the surface of the agar and washed twice with diethyl pyrocarbonate (DEPC) (Sigma) treated water. Total RNA was isolated using the modified TRIzol (Invitrogen) method as described in section 4.2.2. The poly(A)$^+$ mRNA was isolated from total RNA by using an mRNA Isolation kit according to the supplier's instructions (Roche).

4.2.4 RT-PCR amplification of mRNA isolated from *Rhodotorula minuta* grown in chemical defined medium (CD) supplemented with L-phenylalanine

Primers were designed to amplify the full-length coding sequence of the *CYP53B1* gene. *Avr*II restriction sites were included at the beginning of primers just before the start and stop codons. The primers used were: MbF 5'ggcctaggATGGGCATAGTCCAAGAAG3' and MbR 5'ggcctaggCTAGGCATCAATGGATCTG3'. The sequences in the lower case letters represent the *Avr*II recognition sequence.

RT-PCR was performed using the Access RT-PCR Introductory System kit (Promega). The RT-PCR reaction mixture of 50 μl contained 1 x AMV/*Tfl* reaction buffer, 10 mM of each dNTP, 25 pmol each of MbR and MbF primers, $MgSO_4$ was added to the final concentration of 1.5 mM- 2.5 mM, 5 U AMV reverse transcriptase and 5 U *Tfl* DNA polymerase. A sample with RNA supplied with the kit was included as a positive control and a sample with RNA isolated from glucose grown cells was included as a negative control. The following cycle profiles were used: the synthesis of the first strand cDNA at 48°C for 45 minutes, the AMV reverse transcriptase inactivation and cDNA denaturation at 94°C for 2 minutes. This cycle was followed by 40 cycles with denaturation at 94°C for 30 seconds; annealing at 59°C for 1 minute; primer extension at 68°C for 1 minute and finally primer extension at 68°C for 7 minutes.

The PCR products were electrophoresed on a 1% (w/v) agarose gel containing 0.5 μg/ml ethidium bromide. The appropriate size bands were sliced from the gel and purified using the GFX™ PCR DNA and Gel band Purification kit according to supplier's instructions (Amersham Biosciences). The purified DNA samples were ligated into a pGEM-T Easy

vector according to the protocol provided by the supplier (Promega). The ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells. The positive clones were grown in LB medium containing 100 μg/ml ampicillin. Plasmid was isolated from each clone by an alkaline lysis miniprep method (Sambrook *et. al*., 1989). The presence of appropriate sized insert was determined by cleavage of each isolated plasmid with *Eco*RI*,* which cleaves on either side of the ligation site of the pGEM-T Easy vector.

Clones with correct size inserts were digested with *Avr*II. The released insert was ligated into JMP62 and JMP64 vectors (Fig. 1) that were digested with *Avr*II (Biolabs) and dephosphorylated with alkaline phosphatase (Promega). The ligation mixture was incubated at 14°C for 16 hours, and the ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells. The positive clones were grown in LB medium containing 50 μg/ml kanamycin and plasmid was isolated. The orientation of the cloned cDNA was determined by digesting the clones with *Eco*RI that cuts both the vector and the cloned cDNA.

4.2.5 RT-PCR amplification of mRNA isolated from *Rhodotorula* sp. CBS 8446 grown on decane

The primers were designed to amplify the full length coding sequence of the *CYP557A1* gene from *Rhodotorula* sp. CBS 8446. The primers used were: Mol*Bam*HI5'cgggatccATGCTCGCGCTCG3' and Mol*Avr*II5'ggcctaggTCATCGTCG GGATATTGTTACC 3'. The sequences in the lower case letters represent the *Bam*HI and *Avr*II recognition sites respectively.

cDNA was synthesized using a 5'/3'RACE kit (Roche). cDNA was synthesized in a 20 μl reaction containing in 1x cDNA synthesis buffer (50 mM Tris-HCl, 8 mM $MgCl_2$, 30 mM KCl, 1 mM dithiotreitol, pH 8.5), 34 ng of poly(A)$^+$ RNA in 12 μl of water, 12.5 μM of oligo(dT)-anchor primer, 20 mM of each dNTP and 20 U AMV (Roche). The reaction mixture was incubated at 55°C for 60 minutes and further incubated at 65°C for 10 minutes. PCR was performed using 1 μl of the synthesized cDNA template. The PCR reaction mixture of 50 μl contained in 1x polymerase buffer containing 1.75 mM $MgCl_2$, 10 mM of

each dNTP, 12.5 µM of each Mol*Bam*HI and Mol*Avr*II primers and 2.6 U *Taq* DNA expand long template (Roche). The following cycle profiles were used: an initial DNA denaturation at 94°C for 2 minutes, followed by 10 cycles with denaturation at 94°C for 15 seconds; annealing at 62°C for 30 seconds; primer extension at 68°C for 40 seconds. These cycles were followed by 20 cycles with denaturation at 94°C for 15 seconds; annealing at 62°C for 30 seconds; primer extension at 68°C for 40 seconds with 20 seconds increase per cycle, then finally primer extension at 68°C for 7 minutes.

The PCR products were electrophoresed on a 1% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide. The appropriate sized bands were sliced from the gel, purified and ligated into a pGEM-T Easy vector. The ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells and the positive clones were grown in LB medium containing 100 µg/ml ampicillin. Plasmid DNA was isolated from each clone and the presence of appropriate sized insert was determined by cleavage of each isolated plasmid with *Eco*RI. Clones with correct size inserts were double digested with *Bam*HI (Roche) and *Avr*II (Biolabs). The released insert was ligated into JMP62 and JMP64 vectors (Fig. 1) that were double digested with *Bam*HI and *Avr*II. The ligation mixture was incubated at 14°C for 16 hours, and the ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells. The positive clones were grown in LB medium containing 50 µg/ml kanamycin and plasmid was isolated from positive clones.

4.2.6 Preparation of competent *Yarrowia lipolytica* cells

Competent cells of *Yarrowia lipolytica* strain E150 and CPR containing strain E150-CPR were prepared as described by Barth and Gaillardin (1996). Cells from fresh plates were inoculated into 10 ml of YEPD (1% yeast extract, 1% peptone, 1% glucose) containing 50 mM citric acid–Na citrate (pH 4) in 100 ml flasks, incubated at 25°C and grown for 8 hours. The concentration of the cells was determined by counting in a haemocytometer. The cells were inoculated into 25 ml of YEPD in 250 ml flasks at the concentration of 1 x 10^6 cells/ml and incubated at 25°C on a shaker. The cells were grown until the concentration was between 9 x 10^7 and 1.5 x 10^8 cells/ml. The cells were centrifuged at 5000 x g for 5 minutes

and washed twice with 10 ml TE (50 mM Tris, 5 mM EDTA, pH 8). Cells were resuspended in 25 ml of 0.1 M lithium acetate buffer, pH 6 and incubated at 28°C with gentle shaking. The cells were harvested by centrifugation and resuspended in 0.1 M lithium acetate buffer at a concentration of 5 x 10^8 cells/ml.

4.2.7 Transformation of *Yarrowia lipolytica* with vectors containing cDNA

Vector DNA (JMP62 or JMP64) (5 μg) containing the *CYP53B1* or *CYP557A1* cDNA inserts were digested with *Not*I (Roche) restriction enzyme which cuts within the zeta regions of the vectors. The vectors containing the *CYP53B*1 gene were transformed into strain E150, while the vectors containing the *CYP557A*1 gene were transformed into strain E150-CPR. The transformation was performed in a 10 ml falcon tube containing 100 μl of competent cells, 25 μg of carrier DNA (Separations) and 5 μg digested vector was added and mixed gently with a pipette. The mixture was incubated for 15 minutes in 28°C waterbath and 700 μl of 40% PEG 4000 (in 0.1 M lithium acetate, pH 6) was added and incubated at 28°C with shaking for 1 hour. The cells were heat shocked by incubation at 39°C for 10 minutes and 1200 μl of 0.1 M lithium acetate was added. For *URA*3 selection, cells were plated out on YNBcasa plates [0.17 % YNB without amino acids and $(NH_4)_2SO_4$, 0.4% NH_4Cl, 1% glucose, 0.2% casamino acids, 50 mM phosphate buffer and 2 % agar] (Wang *et al*., 1998; 1999). The recombinant cells with JMP64 vector that showed the highest activity for CYP53B1 were transformed further with JMP21-CPR vector. For *LEU2* selection, cells were plated out on YNBD plates (0.17 % YNB without amino acids and $(NH_4)_2SO_4$, 0.4% NH_4Cl, 1% glucose, 50 mM phosphate buffer and 2 % agar) (Wang *et al*., 1998; 1999) and were incubated at 28 °C until colonies appeared.

4.2.8 Southern Hybridization

Five micrograms of total genomic DNA, isolated from strains transformed with JMP64-*CYP53B1*, was double digested overnight with *Pst*I and *Xba*I restriction enzymes (Roche) at 37ºC. The digested DNA was electrophoresed on 1% agarose gel at 80 V for 4 hours. The DNA was depurinated by washing the gel in 0.25 M HCl for 7 minutes. Denaturation was performed by washing the gel twice for 15 minutes in denaturation solution (0.5 M NaOH;

1.5 M NaCl). The gel was neutralized in 0.5 M Tris pH 7.0, 1.5 M NaCl for 30 minutes. The DNA was blotted onto 0.22 micron nylon membrane (Separations) by capillary transfer. The membrane was washed in 10 x SSC (1.5 M NaCl, 0.15 M Na-citrate, pH 7.0) for five minutes, and the DNA was cross-linked to the membrane by exposure to UV light using a GS Gene Linker (Bio-Rad).

A 1.5 kb PCR product from *CYP53B1* cDNA was digoxigenin labeled using a DIG DNA labeling kit according to the supplier's protocol (Roche) and hybridization was performed at 65ºC for 24 hours. High stringency post hybridization washes were performed as follows: 2 x 5 minutes in 2 x SSC, 0.1% SDS at room temperature, followed by washing of the membrane 2 x 15 minutes in 0.1 %SSC, 0.1% SDS at 65ºC. Detection was performed as described by the supplier's protocol (Roche).

4.2.9 Biotransformation of benzoic acid

Transformants with vectors carrying CYP53B1 cDNA were streaked four times on YEPD plates. The recombinant cells were pre-cultured in 10 ml YEPD medium in 100 ml flasks and incubated at 28°C with shaking for 24 hours. A 10 ml YEPS medium (1% yeast extract, 2% peptone and 2% stearic acid) in 100 ml flask was inoculated with 500 µl of the pre-cultures and incubated at 28°C for 24 hours before benzoic acid 0.2 % w/v) was added. Samples (200 µl) were taken after 0, 5, 24, 48, 72, 96 and 120 hours. The samples were acidified to pH 3 with 3% formic acid (100 µl) and extracted with ethyl acetate (200 µl). The extraction mixtures were vortexed for 5 minutes and centrifuged at maximum speed for 10 minutes. The organic layer was transferred to a new tube and concentrated in a speedVac drier. The dried samples were resuspended in 50 µl of ethyl acetate and 10 µl aliquots were spotted on Alugram Sil G/UV_{245} TLC plates (Machery–Nagel). The plates were developed in a mobile phase which consisted of the top layer of a mixture of di-*n*-butyl ether (Merck), formic acid (Merck) and H_2O (90:7:3). Plates were air dried and viewed under UV light (short wavelength).

4.2.10 Biotransformation of hexylbenzene

Clones from transformations with vectors carrying *CYP557A1* cDNA were pre-cultured in 10 ml YEPD medium in 100 ml flasks and incubated at 28°C with shaking for 24 hours. YEPD (10 ml in 100 ml flasks) was inoculated with 500 µl of the pre-culture and incubated at 28°C for 24 hours before oleic acid (Fluka) (1 % v/v) was added. After another 24 hours hexylbenzene (Fluka) (1 % v/v) was added. Sampling, extraction and TLC was done as for benzoic acid biotransformations.

4.2.11 Biotransformation of oleic acid

Twenty clones, (ten from a transformation with JMP62-*CYP557A1* and ten from a transformation with JMP64-*CYP557A1*) were pre-cultured in 10 ml YEPD medium in 100 ml flasks at 28°C with shaking for 24 hours. YEPD (10 ml in 100 ml flasks) was inoculated with 500 µl of the pre-cultures and incubated at 28°C for 24 hours before oleic acid (1 % v/v) was added. Two samples (200 µl) were drawn every 2 hours over a period of 24 hours from each flask. The samples were acidified to pH 3 with 3% formic acid (100 µl) and extracted with ethyl acetate (200 µl) (with and without 0.1% myristic acid as internal standard for GC analysis). The extraction mixtures were vortexed for 5 minutes and centrifuged at maximum speed for 10 minutes. The organic layers were removed and transferred to new tubes. Aliquots (15 µl) from samples without myristic acid as internal standard was spotted on TLC plates. The TLCs were developed in a mobile phase containing formic acid, ethyl ether and hexane (4:20:80). The plates were heated in an oven at 100 ºC for 15 minutes and stained with bromocresol green.

The extracts containing myristic acid as internal standard were dried in a speedVac, resuspended in tert-butyl methyl ether (Fluka) (50 µl) and methylated with trimethylsulfonium hydroxide (TMSH) (50 µl). The derivatized samples were analysed on a Hewlett Packard HP 6890 series GC system equipped with a CP wax 52CB polar column measuring 30 x 44 mm x 1 µm. The initial oven temperature was 120 °C for 5 minutes then

increased at a rate of 10 °C/min to a final temperature of 260°C for 7 minutes. Nitrogen was used as a carrier gas at a rate of 38.4 ml/min and split ratio of 5:1. The flame ionisation detector (FID) temperature was at 350 °C.

The same experiment was repeated with clones 4 and 5 with single copy integration and clones 8 and 10 with multiple copies integrations. The clones were cultured in duplicates and 2% (v/v) of oleic acid was used.

4.3 Results

4.3.1 Functional expression of the benzoate-*para*-hydroxylase from *Rhodotorula minuta* in *Y. lipolytica*

4.3.1.1 RT-PCR amplification of *CYP53B1* cDNA

RT-PCR was performed to amplify a cDNA fragment of 1.5 kb from mRNA isolated from *R. minuta* CBS 2177 grown in CD medium supplemented with L-phenylalanine as only carbon source (Fig. 2). No RT-PCR product was obtained from RNA isolated from cells grown in CD medium supplemented with glucose as a carbon source. These results were in accordance with the results from previous studies which had shown that expression of the *CYP53B1* gene is induced by L-phenylalanine and repressed by glucose (Fukuda *et al*., 1993; Fujii *et al*., 1997; 1989; 1987).

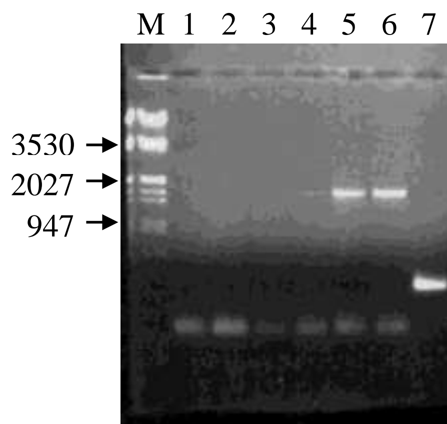


Figure 2. RT-PCR amplification of mRNA from *Rhodotorula minuta* grown in CD medium supplemented with glucose and L-phenylalanine as carbon sources**.** Ethidium bromide stained 1% (w/v) agarose gel showing the RT-PCR amplification of mRNA. Lanes 1-3 = RT- PCR with RNA isolated from cells grown in CD medium with glucose as carbon source, 1.5- 2.5 mM $MgSO_4$. Lanes 4-6 = RT- PCR with RNA isolated from cells grown in CD medium with L-phenylalanine as carbon source, 1.5- 2.5 mM $MgSO_4$. Lane 7 = RNA positive control supplied with a kit. M = λ III marker.

The amplified cDNA was cloned into pGEMT-Easy vector. The cloned cDNA was sequenced in both directions by using T7 and Sp6 universal primers to confirm that the amplified product was *CYP53B1* cDNA. Comparison with the sequence in the NCBI database showed 100 % identity with CYP53B1 in the database.

DNA from clones that contained the cDNA insert in pGEMT-Easy vector was digested with *Avr*II to release the insert. The cDNA insert was ligated into expression vectors JMP62 and JMP64. The ligation mixture was transformed into $CaCl_2$ treated *E. coli* TOP 10 cells. Plasmids were isolated from 10 clones and the orientation of the cloned cDNA was determined. Plasmid DNA isolated from clones with correct orientation was digested with the *Not*I restriction enzyme and transformed into *Y. lipolytica*.

4.3.1.2 Biotransformation studies of benzoic acid

Transformants appeared on minimal medium without uracil 3 days after transformation with JMP62 vector and 10 days after transformation with JMP64 vector. The transformants were streaked on a minimal media four times to stabilise, since transformants with multiple integrations are initially unstable after growing on rich media (Pignede *et al*., 2000a).

In order to test for benzoate *para*-hydroxylase activity in *Y. lipolytica* expressing *CYP53B1*, seven clones with single integration and 10 clones with multiple integrations were grown in YEP medium supplemented with stearic acid, a strong inducer of the *POX*2 promoter, that controls the expression of the cloned cDNA. After growing the cells for 24 hours benzoic acid was added as a substrate for benzoate *para*-hydroxylase. The product was extracted from the cells and spotted on the TLC plates together with *para*-hydroxybenzoic acid standard. The clones that were transformed with void vectors did not form any product. The transformants obtained after transformation with JMP62, which was expected to give single copy integration, showed faint bands for the product after 48 hours (Fig. 3A). No increase in product was observed in samples taken after 72 h and 96 h.

The transformants obtained after transformation with JMP64, which was expected to give multicopy integration, produced within 24 h sufficient product to detect on the TLC plates (Fig. 3B). The concentration of the product increased with time for transformants with multiple integrations. Clones 3, 7 and 8 showed the highest product formation (Fig. 3B). After 72 hours the product formation levelled off. Since clones with multiple integrations gave better results than clones with single integration, only the clones with multiple integrations were used in further studies.

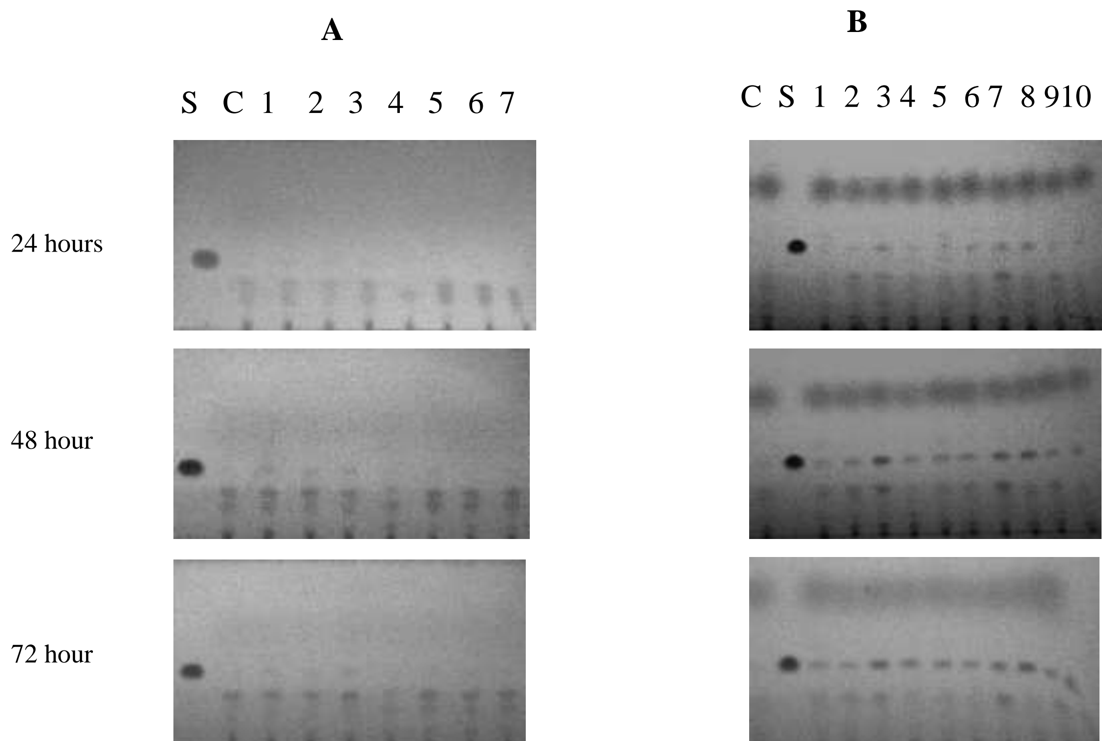


Figure 3. Biotransformation of benzoic acid. Biotransformation was performed with (**A**) single integration and (**B**) multiple integration transformants. Lanes S refer to standard *para*-hydroxybenzoic acid. Lanes C refer to clones transformed with void vectors. Lanes 1-10 refer to different transformants. Samples were taken after 24, 48 and 72 hours after addition of benzoic acid. The samples were extracted with ethyl acetate and dried. The dried samples were resuspended in ethyl acetate and spotted on TLC plates.

4.3.1.3 Confirmation of integration of *CYP53B1* into the *Yarrowia lipolytica* genome

In order to compare the level and nature of *CYP53B1* integration into genome of *Yarrowia lipolytica* southern hybridization was performed. Genomic DNA was isolated from transformants transformed with JMP64 and double digested with *Xba*I/*Pst*I that do not cut within the *CYP53B1* coding sequence. The digested DNA was transferred to a nylon membrane.

A 1.5 kb coding sequence of *CYP53B1* was labeled and used as a probe for detection of the integration of this gene in the genome of *Y. lipolytica*. A major band with a size of

approximately 3.5 kb was observed for all the clones except for clone 4 (Fig. 4). This common major band is an indication that the integration was in tandem. Such a result was expected, since the zeta sequences are tandemly arranged in the genome (Juretzek *et al.,* 2001). The intensity of this major band correlated with results obtained from the biotransformation studies of benzoic acid (Fig. 3B). The band observed for clones 3, 7 and 8 had the highest intensity and these clones also gave the highest concentration of *para*-hydroxy benzoic acid in the bioconversion experiments. In many clones several other bands were also observed (Fig. 4). This is an indication that *CYP53B1* was also integrated at other sites in the genome.

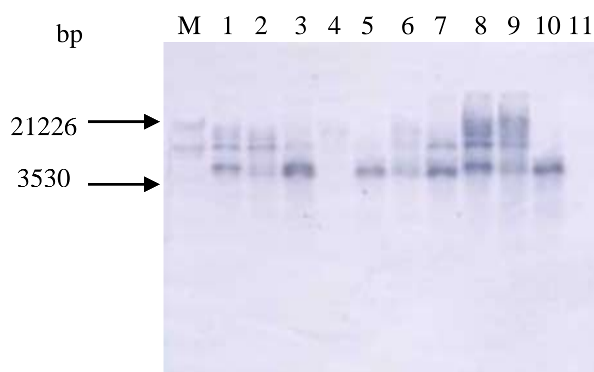


Figure 4. Southern blot analysis of multicopy transformants**.** DNA was double digested with *Xba*I and *Pst*I. Digested DNA was run on the gel and transferred to a membrane. CYP53B1 cDNA was used as a probe. Lanes 1-10 refer to clones 1-10. Lane 11 refers to E150 transformed with void vector. M = λ DNA digested with *Hind*III/*Eco*RI. Hybridization was performed at 65 °C and all washes were of high stringency.

4.3.1.4 Expression of a cytochrome NADPH reductase (*CPR*) gene in *Y. lipolytica*

Several researchers had previously shown that the activity of cloned cytochrome P450 monooxygenases might be limited by the level of cytochrome NADPH reductase activity (van den Brink *et al*., 1996; Fisher *et al*., 1992; Dong and Porter, 1996; Blake *et al*., 1996). Clones 3, 7 and 8 were therefore subsequently transformed with JMP21-CPR. The JMP21-CPR vector contains the *Y. lipolytica* cytochrome P450 NADPH reductase gene (*CPR*) under control of the strong *ICL* (isocitrate lyase) promoter (Barth and Scheuber, 1993).

Transformation was only successful with clone 7. The clones that were transformed with the JMP21-CPR vector showed a significant improvement in the conversion of benzoic acid to *para*–hydroxy benzoic acid (Fig. 5). The product was detected within five hours after the addition of benzoic acid. Clone 7-3 showed the highest activity for the product formation that peaked at 96 hours (Fig. 5).

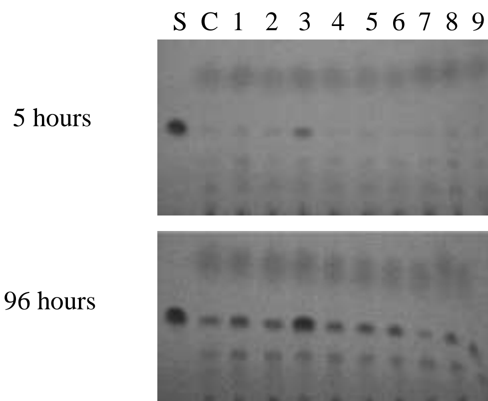


Figure 5. Biotransformation of benzoic acid**.** The TLC plates showing bioconversion of benzoic acid to *para*-hydroxy benzoic acid by multicopy transformants transformed with the *CPR* gene. Lanes S refer to *para*-hydroxy benzoic acid standard. Lanes C refer to a transformant with a void vector JMP21. Lanes 1-9 refer to transformants with CPR gene. Samples were taken after 5 and 96 hours after addition of benzoic acid.

These results showed that cytochrome CPR activity was a limiting factor in the conversion of benzoic acid to *para*-hydroxy benzoic acid. The results reported here are only preliminary screening results to find clones with the highest benzoate-*para*-hydroxylase activities. Studies to quantify hydroxylase activity and determine factors limiting biotransformation of benzoic acid by *Y. lipolytica* strains expressing *CYP53B1* are currently in progress in our research group. Since these preliminary results confirmed that expression of a *Rhodotorula* cytochrome P450 in *Y. lipolytica* together with the *Y. lipolytica* CPR yielded strains with the desired hydroxylase activity, the same expression system was used to express the newly isolated ω-hydroxylase gene (*CYP557A1*) from *Rhodotorula* sp. CBS 8446.

4.3.2 Functional expression of the fatty acid omega-hydroxylase from *Rhodotorula* sp. CBS 8446 in *Y. lipolytica*

4.3.2.1 RT-PCR amplification of CYP557A1 cDNA and transformation into *Y. lipolytica*

mRNA was isolated from cells of *Rhodotorula* sp. CBS 8446 grown on decane and cDNA synthesis was performed. Primers were designed to amplify the full-length cDNA of the *CYP557A*1 gene. The primers amplified the expected cDNA fragment of approximately 1.6 kb (Fig. 6). No PCR product was obtained from cells grown on glucose (data not shown). The amplified cDNA fragment was cloned into pGEM-T Easy vector. The cloned insert was released from pGEM-T Easy vector with *Avr*II and *Bam*HI, ligated into the JMP62 and JMP64 vectors and then transformed into *E. coli*. The plasmid was isolated from positive clones containing the gene and digested with *Not*I restriction enzyme to expose the zeta sequences for integration into the *Y. lipolytica* genome. The digested DNA was transformed into *Y. lipolytica* strain E150-CPR. Strain E150-CPR already contained an additional copy of the *Y. lipolytica CPR* gene under the control of the *ICL* promoter. The cells were also transformed with empty vectors. Transformants which appeared on medium without uracil 3 days after transformation with JMP62 and 10 days after transformation with JMP64, were streaked on a rich medium (YEPD) and subcultured several times to obtain plasmid stability. These strains were tested for alkyl benzene and fatty acids biotransformation.

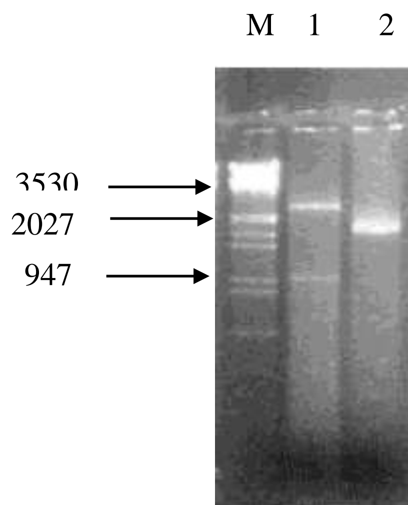


Figure 6. Isolation of full-length CYP557A1 cDNA by RT-PCR using mRNA isolated from decane grown cells**.** Ethidium bromide stained 1 % (w/v) agarose gel showing the PCR on genomic DNA (lane1) and amplified RT-PCR product (lane 2) using Mol*Bam*HI and Mol*Avr*II primers. M = λ DNA digested with *Hind*III/*Eco*RI.

4.3.2.2 Biotransformation of alkylbenzenes

It is difficult to measure alkane or omega hydroxylase activity in whole cells of alkane utilizing yeasts with intact β-oxidation, because no product accumulates. When alkanes are used as substrates, it is also difficult to follow disappearance of the substrate because the substrate is not soluble in water. Fatty acids such as stearic acid and lauric acid are also solids, which only dissolve in water at high pH. Disappearance of fatty acids of chain lengths C14 to C18 might also be ascribed to incorporation into lipids.

Alkane utilizing yeasts transform alkylbenzenes to phenyl acetic acid (Mauersberger *et al*., 1996), a product which is easy to detect with TLC and GC. It has been demonstrated in our laboratory that *Y. lipolytica* E150-CPR containing multiple copies of the *Y. lipolytica ALK1* or *ALK2* genes, accumulates phenyl acetic acid at an increased rate (Dr. E. Setati, personal communication).

Transformants picked up after transformation with JMP62 (10 clones) and JMP64 (10 clones) carrying *CYP557A1* were tested for biotransformation of hexylbenzene. In these experiments cultures were grown in YEPD medium for 24 hour and then supplemented with oleic acid, a strong inducer of the *POX*2 promoter, which is regulating the cloned cDNA. After growing the cells for a further 24 hours hexylbenzene was added as a substrate. The product was extracted from the cells and spotted on the TLC plates together with phenylacetic acid standard (Fig. 7). The product was detected after 5 hours for both the single and multiple transformants. It appeared as if after 5 hours, less product was formed by the clones containing void vectors, however after 24 hours the product from clones with void vectors and clones expressing the omega hydroxylase gene was at the same level (data not shown). The results obtained with TLC also did not correlate very well with GC analyses (data not shown). After several experiments it was concluded that there was not a significant increase in the rate of phenyl acetic acid formation by clones with single and multiple integrations of the *CYP557A1* gene. Thus, these results could not be used to show that the omega hydroxylase gene was expressed in the *Y. lipolytica* transformants.

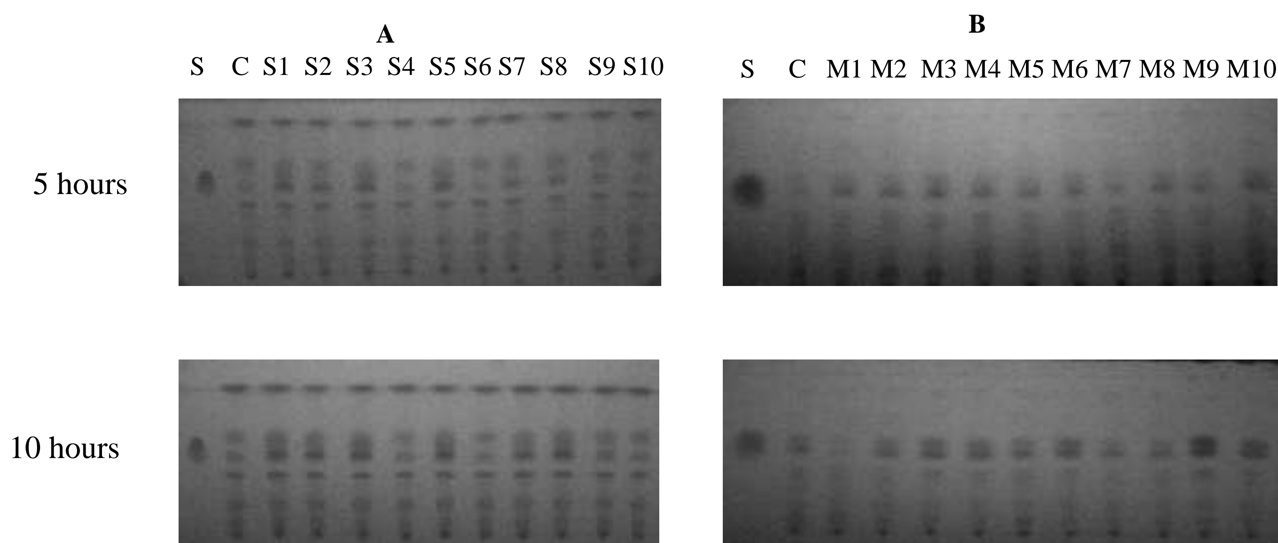


Figure 7. Biotransformation of hexylbenzene with (**A**) single integration and (**B**) multiple integrations transformants. On the TLC plates lanes S refer to phenyl acetic acid standard. Lanes C refer to strain E150 with CPR and void JMP64. Lanes S1-S10 and M1-M10 refer to clones with single and multiple integrations respectively. Samples were taken at 5 hours and 10 hours after addition of hexylbenzene. The TLC plates were air dried and exposed to UV light for visualization.

4.3.2.3 Effect of transformation with *CYP557A1* on oleic acid consumption

Our failure to demonstrate an increased rate of alkylbenzene transformation by transformants containing the *CYP557A1* gene, might be ascribed to the fact that *CYP557A1* codes for a fatty acid omega hydroxylase without alkane hydroxylase activity. It was therefore possible that these strains might show an increased rate of oleic acid consumption. In order to investigate this possibility, transformants were grown in YEPD medium until they reached stationary phase (24 hours) followed by the addition of oleic acid. Since these strains still have active β-oxidation, accumulation of a product was not expected. Initially 10 clones from both the single and multiple integration experiments were tested and disappearance of oleic acid was only monitored on TLC plates (Fig. 8). In these experiments oleic acid was added at an initial concentration of 1% (v/v). After 2.5 hours there were already significant differences between the different clones transformed with the multiple integration vector JMP64, and oleic acid could not be observed in samples from clones M8 and M10 (Fig. 8B).

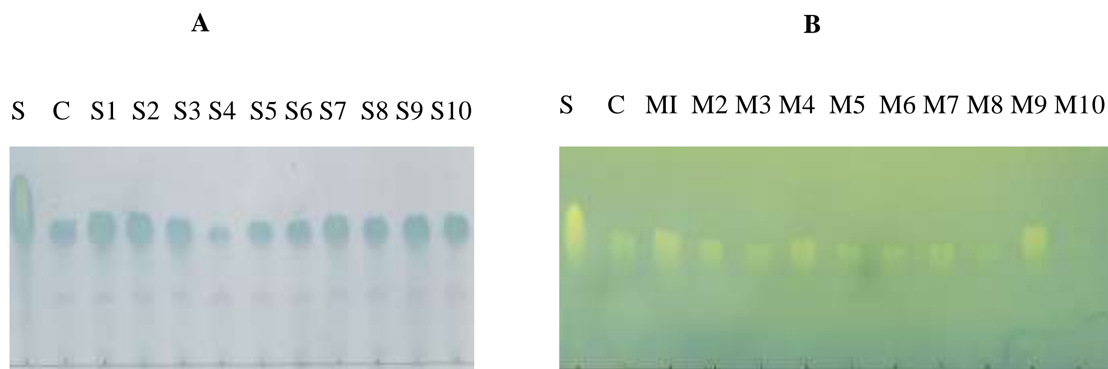


Figure 8. The TLC plates showing consumption of oleic acid by transformants with (**A**) single integration and (**B**) multiple integration. On the TLC plates lanes S refer to oleic acid standard. Lane C refers to strain E150 with CPR and void JMP64. Lanes S1-S10 and M1-M10 refer to clones with single and multiple integrations respectively. Samples were taken at 5 hours for single integration and at 2.5 hours for multiple integrations after addition of oleic acid. The TLC plates were stained with bromocresol green.

After 5 hours oleic acid had disappeared from all the multiple integration clones except M1 and M9 (data not shown). It was also still present in samples from the control strain E150-

CPR and in the clones transformed with the single integration vector JMP62. Of the clones transformed with JMP62 only S4 and S5 had less oleic acid left than the control strain E150-CPR (Fig. 8A).

The 2.5 hours samples from clones M8, M10 and E150-CPR and the 5 hours samples from clones S4, S5 and E150-CPR were subjected to GC analysis and the results from the GC analysis correlated with the TLC analysis (data not shown). In a subsequent experiment with strains E150-CPR, S4, S5, M8 and M10 the oleic acid concentration was increased to 2% (v/v). The disappearance of oleic acid from the growth medium was monitored by GC analysis in duplicate samples taken 2, 4, 8 and 12 hours after substrate addition. Oleic acid consumption was again fastest for clones M8 and M10 transformed with the multiple integration vector JMP64 and slowest for the control strain E150-CPR (Fig. 9). However, in this experiment the oleic acid never disappeared completely and 12 hours after substrate addition the oleic acid concentration increased again in samples taken from clone M10. It is thus evident that oleic acid consumption is not only due to hydroxylation followed by β-oxidation, but probably also to incorporation into the lipid fractions from where it can be released again. Therefore the results from these last experiments, although apparently confirming fatty acid omega-hydroxylase activity of the cloned *CYP557A1* gene product, should be treated/interpreted with caution. It is also strange that oleic acid consumption is different for the different strains so soon after addition of oleic acid, which should first induce expression of the cloned *CYP557A1* gene. Dry weight determinations done on samples taken after 24 hours after addition of oleic acid, confirmed that the differences in oleic acid consumption was not due to differences in biomass produced by the different strains. All the strains had produced 23.5 g dry biomass.l^{-1}.

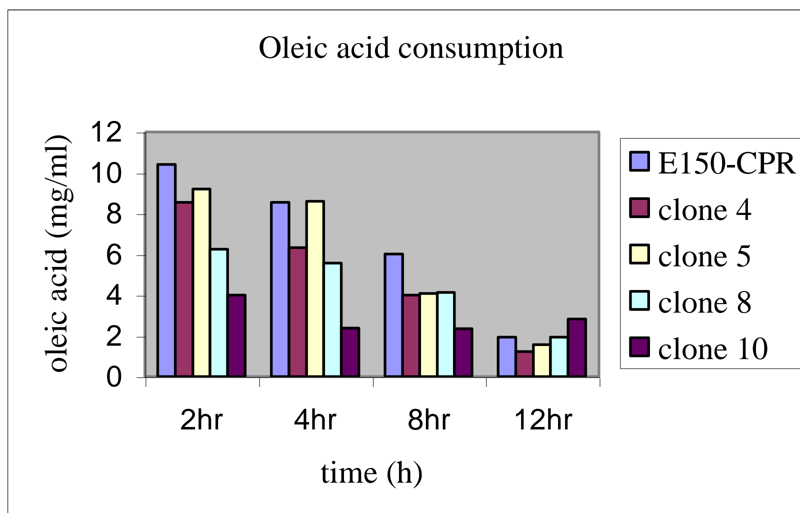


Figure 9. Monitoring the disappearance of oleic acid. The disappearance of oleic acid was monitored by using GC. Strain E150-CPR contains *CPR* gene and void JMP64 vector. Clones 4 and 5 contain single copy integration, while clones 8 and 10 contain multiple integrations. Samples were taken at 2, 4, 8 and 12 hours after addition of oleic acid. Samples were extracted with ethyl acetate containing 0.1% myristic acid as internal standard. Dried samples were resuspended in tert-butyl methyl ether and methylated with TMSH and then subjected to GC. The values are means of two independent experiments.

4.4 Discussion

The heterologous expression of P450s is very important for biotechnological research such as degradation of pollutants in bioremediation, degradation of xenobiotics, and production of hydroxylated fine chemicals in biocatalysis. Most heterologous expression studies have concentrated on expression of mammalian P450s in *E. coli* (Dong and Porter, 1996; Blake *et al*., 1996; Fisher *et al.,* 1992; Barnes *et al*., 1991; Iwata *et al*., 1998) and *S. cerevisiae* (Sakaki *et al*., 1994; Murakami *et al*., 1990). In these studies the transfer of electrons from NADPH to cytochrome P450 was improved by co-expression of CPR. The results obtained from these studies had contributed for instance to the development of therapeutic drugs. However, the heterologous expression of fungal P450s is limited except for studies that were performed on the CYP52 family.

The cytochrome P450s belonging to the CYP52 family from alkane-assimilating yeasts represent one family of fungal P450, which has been studied extensively. Many CYP52

isozymes were isolated from *C. maltosa* and *C. tropicalis* (Segezzi *et al*., 1991; Zimmer *et al*., 1996). To determine the substrate specificities and regio-selectivities of these enzymes, they were expressed in *S. cerevisiae* (Zimmer *et al*., 1996; 1995; Schunck *et al.,* 1991) and in insect cells (Eschenfeldt *et al*., 2003). Enzyme assays performed with microsomal fractions containing the expressed CYP52s demonstrated that they have different specificities and regioselectivities towards alkanes and fatty acids with an overlap in activities.

Another fungal P450 which has been studied extensively is the benzoate *para*-hydroxylase (CYP53A1) from *Aspergillus niger* (van Gorcom *et al*., 1990; van de Brink *et al*., 1996; 2000; Boschloo *et al*., 1990; Faber *et al.,* 2001). In these studies multiple copies of the indigenous *CPR* as well as *CYP53A1* were introduced into *A. niger* to increase the cytochrome P450 activity. However this expression cannot be regarded as heterologous expression since the genes were not expressed in a foreign host.

In all the studies on heterologous expression of fungal P450s, the bioconversions were performed with enzyme assays on microsomal fractions. The cells were harvested and microsomal fractions prepared, which were used for reconstitution studies. The purpose of this study was to develop an expression system, which will allow us to use whole cells for bioconversions instead of microsomal fractions. Since the expression tools are available for the yeast *Y. lipolytica* it was used as a host for heterologous expression of *CYP53B1* and *CYP557A1*.

The *CYP53B1* gene had previously been isolated from *Rhodotorula minuta* and its product was found to be bifunctional for isobutene production and *para*-hydroxylation of benzoic acid (Fujii *et al*., 1997). Through expression of the *CYP53B1* in *Y. lipolytica* it was demonstrated that a gene from a basidiomycetous fungus, which is different from an ascomycetous fungus can be expressed successfully in *Y. lipolytica*, an ascomycete. Nthangeni *et al*., (2004) recently demonstrated that a mammalian gene, *CYP1A1* can be functionally expressed in *Y. lipolytica* with the same expression vectors. Thus, the *Y. lipolytica* expression system is suitable for expression of both mammalian and fungal P450s.

Multiple integration of the P450 gene together with co-expression of *Y. lipolytica* CPR was necessary, both in the case of CYP1A1 and in the case of CYP53B1, to obtain high hydroxylation activity.

Van Gorcom *et al*., (1990) observed in *A. niger* that if the copy number of the *CYP53A1* increased above five the benzoate-*para*-hydroxylase (BPH) activities also decreased and cells tended to grow slowly on benzoic acid. It was demonstrated in *Y. lipolytica* that the transformants need 10-13 copies to be stable, if the copy number is lower than 10 the growth could be retarded (Juretzek *et al*., 2001; Le Dall *et al*., 1994). In this study high copy numbers apparently did not cause retardation in growth of the transformants. The availability of electrons to the cytochrome P450 protein was the limiting factor for product formation. The indigenous CPR was not enough to transfer electrons sufficiently to the expressed P450. When an additional copy of the CPR gene was introduced into clones already expressing the *CYP53B1*, an increase in product formation was observed (Fig. 5). These results showed that CPR was the limiting factor for the conversion of benzoic acid to *para*-hydroxy benzoic acid. In a related study it was observed in *A. niger* that the introduction of multiple copies of the *CYP53A1* gene in the genome increased the cytochrome P450 level, but only slightly increased benzoate-*para*-hydroxylase activity. However, the addition of multiple copies of *CPR* increased the benzoate-*para*-hydroxylase activity by about two folds (van den Brink *et al*., 1996). It was also observed in *Y. lipolytica* expressing multiple copies of *CYP1A1* that CPR was the limiting factor for bioconversion (Nthangeni *et al*., 2004). Thus, high cytochrome P450 activity can only be achieved if both components of the P450 system are co-expressed. The CPR should form a complex with the cytochrome P450 to facilitate a transfer of electrons from NADPH to the cytochrome P450 (Sutter *et al*., 1990; Backes and Kelly, 2003). Another factor that can be a limiting factor for bioconversion is cytochrome P450: CPR ratio. Some studies had shown that the ratio could be a limiting factor depending on the P450 being expressed (Dong and Porter, 1996; Backes and Kelly, 2003).

Y. lipolytica as a host for benzoate *para*-hydroxylase made screening for activity relatively easy, since it does not metabolize benzoic acid or *para*-hydroxy benzoic acid. In addition the

substrate (benzoic acid) is relatively cheap and the product, *para*-hydroxy benzoic acid, is easy to detect with TLC. The results obtained in this study should be treated as preliminary results to demonstrate that the *Y. lipolytica* expression system is functional. Other members of our group will use the strains created in this study to do quantitative biotransformations and to optimize induction and biotransformation conditions. This system can also be used to investigate other factors that might limit the activity of heterologously expressed P450s.

Heterologous expression of an omega-hydroxylase in *Y. lipolytica* is more complicated, since *Y. lipolytica* has indigenous omega hydroxylases that can also act as alkane hydroxylases (Iida *et al*., 1998; 2000) and *Y. lipolytica* normally does not accumulate any intermediates from alkane or fatty acid degradation. However, it accumulates phenyl acetic acid from alkylbenzenes, which should be hydroxylated by the same P450s. It was demonstrated in our group that the expression of multiple copies of *Y. lipolytica* CYP52s *YlALK1* and *YlALK2* in *Y. lipolytica* strain E150 resulted in an increased rate of production of phenyl acetic acid from alkylbenzenes (Dr. E. Setati, personal communication). Thus alkylbenzenes as substrates can be useful to detect alkane hydroxylase activity in strains with active beta-oxidation. It is therefore interesting that when *CYP557A1* was expressed in *Y. lipolytica* lower rates of alkylbenzene hydroxylation was observed. It is possible that *CYP557A1* codes for an omega-hydroxylase that does not accept alkylbenzenes or alkanes as substrates, since it is more similar to CYP86s and CYP94s (Benveniste *et al*., 1998; Tijet *et al*., 1998; Le Bouquin *et al*., 1999), which have no alkane hydroxylase activity, than to the CYP52s.

It is very difficult to test omega hydroxylase activity in whole cells of *Y. lipolytica* with an intact β-oxidation pathway. *Y. lipolytica* E150-CPR which was transformed with the *CYP557A1* gene still has an active β-oxidation pathway. Therefore no accumulation of a product was expected from fatty acids. In this case it was only possible to measure the rate at which the fatty acid substrate disappeared. Transformants that contained multiple copies of *CYP557A1* gene showed the highest oxidation of oleic acid. The results obtained with multicopy transformants showed the same pattern with results obtained from the induction studies of omega-hydroxylase genes of *C. tropicalis* (Craft *et al*., 2003). The *CYP52A13* and

CYP52A14 showed the highest induction within one hour after addition of oleic acid and induction levelled off with time (Craft *et al*., 2003). Our results also correlated with results obtained from induction studies of *POX2* and *ICL* promoters driving the *Lac Z* gene where the highest induction by oleic aid was observed between three and five hours (Juretzek *et al*., 2000).

Our present data showed that the consumption of oleic acid could not be attributed to the expression of CPR only, since transformants with both *CYP557A1* and *CPR* showed faster oxidation of oleic acid than the transformant with CPR only. On the other hand, the results obtained in this study should be treated with caution because oleic acid can also be incorporated in lipids. Thus, further experiments are required to confirm omega-hydroxylase activity of CYP557A1. These results can be confirmed by transforming the *CYP557A1* into strains with the β-oxidation pathway blocked, or by using substrates that are not incorporated into lipids.

It was demonstrated in this study that *Y. lipolytica* can be used as a host for the expression of P450s from *Rhodotorula* spp. and probably other fungi. Biotransformations were done with whole cells instead of microsomal fractions. This system can work well for compounds such as benzoic acid, which is not degraded by *Y. lipolytica*. However, further genetic engineering of *Y. lipolytica* is necessary to eliminate indigenous enzymes involved in degradation of substrates such as *n*-alkanes and fatty acids, if *Y. lipolytica* is to be used as a host for the expression of P450s involved in the hydroxylation of such substrates.

4.5 References

Backes, WL. and Kelley, RW. (2003). Organization of multiple cytochrome P450s with NADPH-cytochrome P450 reductase in membranes. *Pharmacol Ther*. **98**: 221-233.

Barnes, HJ., Arlotto, MP. and Waterman, MR. (1991). Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. *Proc Natl Acad Sci U S A*. **88**: 5597-5601.

Barth, G. and Gaillardin, C. (1996). *Yarrowia lipolytica.* In Wolf K (ed.), Nonconventional Yeasts in Biotechnology, Springer - Verlag Berlin pp. 313-388.

Barth, G. and Gaillardin, C. (1997). Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol Rev*. **19**: 219-237.

Barth, G. and Scheuber, T. (1993). Cloning of the isocitrate lyase gene (*ICL*1) from *Yarrowia lipolytica* and characterization of the deduced protein. *Mol Gen Genet*. **241**: 422-430.

Benveniste, I., Tijet, N., Adas, F., Philipps, G., Salaun, JP. and Durst, F. (1998). *CYP86A1* from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid omega-hydroxylase. *Biochem Biophys Res Commun*. **243**: 688-693.

Blake, JA., Pritchard, M., Ding, S., Smith, GC., Burchell, B., Wolf, CR. and Friedberg T. (1996). Coexpression of a human P450 (CYP3A4) and P450 reductase generates a highly functional monooxygenase system in *Escherichia coli*. *FEBS Lett*. **397**: 210-214.

Boshloo, JG., Paffen, A., Koot, T., van den Tweel, WJ. van Gorcom, RF., Cordewener, JH. and Bos, CJ. (1990). Genetic analysis of benzoate metabolism in *Aspergillus niger*. *Appl. Microbiol. Biotechnol*. **34**: 225-228.

Craft, DL., Madduri, KM., Eshoo, M. and Wilson, CR. (2003). Identification and characterization of the CYP52 family of *Candida tropicalis* ATCC 20336, important for the conversion of fatty acids and alkanes to alpha, omega-dicarboxylic acids. *Appl Environ Microbiol*. **69**: 5983-5991.

Davidow, LS., O'Donnell, MM., Kaczmarek, FS., Pereira, DA., DeZeeuw, JR. and Franke, AE. (1987). Cloning and sequencing of the alkaline extracellular protease gene of *Yarrowia lipolytica*. *J Bacteriol*. **169**: 4621-4629.

Dong, J. and Porter, T. D. (1996). Coexpression of mammalian cytochrome P450 and reductase in *Escherichia coli. Arch. Biochem. Biophys*. **327**: 254-259.

Durham, DR. (1984). Initial reactions involved in the dissimilation of mandelate by *Rhodotorula graminis*. *J Bacteriol*. **160**: 778-780.

Durham, DR., McNamee, CG. and Stewart, DB. (1984). Dissimilation of aromatic compounds in *Rhodotorula graminis*: biochemical characterization of pleiotropically negative mutants. *J Bacteriol*. **160**: 771-777.

Eschenfeldt, WH., Zhang, Y., Samaha, H., Stols, L., Eirich, LD., Wilson, CR. and Donnelly, MI. (2003). Transformation of fatty acids catalyzed by cytochrome P450 monooxygenase enzymes of *Candida tropicalis*. *Appl Environ Microbiol*. **69**: 5992-5999.

Faber, BW., van Gorcom, RF. and Duine, JA. (2001). Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (*CYP53A1*), from *Aspergillus niger*. *Arch Biochem Biophys*. **394**: 245-254.

Fisher, CW., Shet, MS., Caudle, DL., Martin-Wixtrom, CA. and Estabrook, RW. (1992). High-level expression in *Escherichia coli* of enzymatically active fusion proteins

containing the domains of mammalian cytochromes P450 and NADPH-P450 reductase flavoprotein. *Proc Natl Acad Sci U S A*. **89**: 10817-10821.

Fujii, T., Nakamura, K., Shibuya, K., Tanase, S., Gotoh, O., Ogawa, T. and Fukuda, H. (1997). Structural characterization of the gene and corresponding cDNA for the cytochrome P450rm from *Rhodotorula minuta* which catalyzes formation of isobutene and 4-hydroxylation of benzoate. *Mol Gen Genet*. **256**: 115-120.

Fujii, T., Ogawa, T. and Fukuda, H. (1987). Isobutene production by *Rhodotorula minuta Appl. Microbiol. Biotechnol.* **25**: 430-433.

Fujii, T., Ohga, Y., Ogawa, T. and Fukuda, H. (1989). The role of L-phenylalanine in the production of isobutene by *Rhodotorula minuta*. *J. Ferment. Bioeng*. **67**: 115-118.

Fukuda, H., Fujii, T., Daimon, H., Iwata, M., Ogawa, T., Tanase, S. and Morino, Y. (1993). Purification and characterization of cytochrome P450 from an isobutene-forming microorganism, *Rhodotorula minuta*. *Biosci Biotechnol Biochem*. **57**: 1599-15601.

Iida, T, Ohta, A. and Takagi, M. (1998). Cloning and characterization of an *n*-alkane-inducible cytochrome P450 gene essential for *n*-decane assimilation by *Yarrowia lipolytica*. *Yeast*. **14**: 1387-1397.

Iida, T., Sumita, T., Ohta, A. and Takagi, M. (2000). The cytochrome P450ALK multigene family of an *n*-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning and characterization of genes coding for new CYP52 family members. *Yeast*. **16**: 1077-1087.

Iwata, H., Fujita, K., Kushida, H., Suzuki, A., Konno, Y., Nakamura, K., Fujino, A. and Kamataki, T. (1998). High catalytic activity of human cytochrome P450 co-expressed with human NADPH-cytochrome P450 reductase in *Escherichia coli*. *Biochem Pharmacol*. **55**: 1315-1325.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., Nicaud, J. (2001). Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast*. **18**: 97-113.

Juretzek, T., Wang, H., Nicaud, JM., Mauersberger, S., and Barth, G. (2000). Comparison of promoters suitable for regulated overexpression of heterologous genes in the alkane-utilizing yeast *Yarrowia lipolytica*. *Biotechnol. Bioprocess Eng*. **5**: 320-326.

Le Bouquin, R., Pinot, F., Benveniste, I., Salaun, JP. and Durst, F. (1999). Cloning and functional characterization of *CYP94A2*, a medium chain fatty acid hydroxylase from *Vicia sativa*. *Biochem Biophys Res Commun*. **261**: 156-162.

Le Dall, MT., Nicaud, JM. and Gaillardin, C. (1994). Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr Genet***.** **26**: 38-44.

Liu, Y., Kondo, A., Ohkawa, H, and Shiota, N. and Fukuda, H. (1998). Bioconversion using immobilized recombinant flocculent yeast cells carrying a fused enzyme gene in an 'intelligent' bioreactor. *Biochemical Engineering Journal* **2**: 229 – 235.

Madzak, C., Treton, B. and Blanchin-Roland, S. (2000). Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J. Mol. Microbiol Biotechnol.* **2**: 207-216.

Madzak, C., Blanchin-Roland, S., Cordero Otero, RR. and Gaillardin, C. (1999). Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiology*. **145**: 75-87.

Mauersberger, S., Ohkuma, M., Schunck, W-H. and Takagi, M. (1996). *Candida maltosa* In Wolf K (ed.), Nonconventional Yeasts in Biotechnology, Springer - Verlag Berlin pp. 411- 579.

Moleleki, N. (1998). The involvement of cytochrome P-450 monooxygenases in the hydroxylation of monoterpenes by yeasts. (M.Sc. Thesis).

Murakami, H., Yabusaki, Y., Sakaki, T., Shibata, M., and Ohkawa, H. (1990). Expression of cloned yeast NADPH-cytochrome P450 reductase gene in *Saccharomyces cerevisiae. J. Biochem*. **108**: 859 – 865.

Nicaud, JM., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P. and Gaillardin, C. (2002). Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res*. **2**: 371-379.

Nicaud, JM., Fabre, E., Beckerich, JM., Fournier, P. and Gaillardin, C. (1989). Cloning, sequencing and amplification of the extracellular protease (*XPR2*) gene of the yeast *Yarrowia lipolytica*. *J. Biotechnol.* **12**: 285-298.

Nthangeni, MB., Urban, P., Pompon, D., Smit, M. and Nicaud, J-M. (2004). The use of *Yarrowia lipolytica* for the expression of human cytochrome P450 1A1. *Yeast* (in press).

Omura, T. and Sato, R. (1964). The carbon monooxide-binding pigment of liver microsomes. *J. Biol. Chem*. **239**: 2379-2385.

Park, CS., Chang, CC. and Ryu, DD. (2000). Expression and high-level secretion of *Trichoderma reesei* endoglucanase I in *Yarrowia lipolytica*. *Appl Biochem Biotechnol*. **87**: 1-15.

Park, CS., Chang, CC., Kim, JY., Ogrydziak, DM. and Ryu, DD. (1997). Expression, secretion, and processing of rice alpha-amylase in the yeast *Yarrowia lipolytica*. *J Biol Chem*. **272**: 6876-6881.

Pignede, G., Wang, HJ., Fudalej, F., Seman, M., Gaillardin, C. and Nicaud, JM. (2000a). Autocloning and amplification of *LIP2* in *Yarrowia lipolytica*. *Appl Environ Microbiol*. **66**: 3283-3289.

Pignede, G., Wang, HJ, Fudalej, F., Gaillardin, C., Seman, M., Nicaud, JM. (2000b). Characterization of an extracellular lipase encoded by *LIP2* in *Yarrowia lipolytica*. *J Bacteriol.* **182**: 2802-2810.

Porter, TD. and Coon, MJ. (1991). Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem*. **266**: 13469-13472.

Sakaki, T., Kominami, S., Takemori, S., Ohkawa, H., Akiyoshi-Shibata, M. and Yabusaki, Y. (1994). Kinetic studies on a genetically engineered fused enzyme between rat cytochrome P4501A1 and yeast NADPH-P450 reductase. *Biochemistry*. **33**: 4933-4939.

Sambrook, J., Fritsch, EF. and Maniates F. (1989). Molecular cloning: a laboratory manual, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanglard, D. and Loper, CJ. (1989). Characterization of the alkane-inducible cytochrome P450 (P450alk) gene from the yeast *Candida tropicalis*: identification of a new P450 gene family. *Gene.* **76**: 121-136.

Scheller, U., Zimmer, T., Becher, D., Schauer, F. and Schunck, WH. (1998). Oxygenation cascade in conversion of *n*-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A3. *J Biol Chem*. **273**: 32528-32534.

Schunck, WH., Vogel, F., Gross, B., Kargel, E., Mauersberger, S., Kopke, K., Gengnagel, C. and Muller HG. (1991). Comparison of two cytochromes P-450 from *Candida maltosa*: primary structures, substrate specificities and effects of their expression in *Saccharomyces cerevisiae* on the proliferation of the endoplasmic reticulum. *Eur J Cell Biol*. **55**: 336-345.

Seghezzi, W., Sanglard, D. and Fiechter, A. (1991). Characterization of a second alkane-inducible cytochrome P450- encoding gene, *CYP52A2*, from *Candida tropicalis*. *Gene.* **106**: 51-60.

Sutter, TR., Sanglard, D., Loper, JC. and Sangard, D. (1990). Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from *Candida tropicalis*. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins. *J Biol Chem*. **265**: 16428-16436.

Tijet, N., Helvig, C., Pinot, F., Le Bouquin, R., Lesot, A., Durst, F., Salaun, JP. and Benveniste, I. (1998). Functional expression in yeast and characterization of a clofibrate-inducible plant cytochrome P-450 (*CYP94A1*) involved in cutin monomers synthesis. *Biochem J*. **332**: 583-589.

van den Brink, JM., Punt, PJ., van Gorcom, RF. and van den Hondel, CA. (2000). Regulation of expression of the *Aspergillus niger* benzoate *para*-hydroxylase cytochrome P450 system. *Mol Gen Genet*. **263**: 601-609.

van den Brink, JM., van den Hondel, CA. and van Gorcom, RF. (1996). Optimization of the benzoate-inducible benzoate *p*-hydroxylase cytochrome P450 enzyme system in *Aspergillus niger*. *Appl Microbiol Biotechnol*. **46**: 360-364.

van Gorcom, RF., Boschloo, JG., Kuijvenhoven, A., Lange, J., van Vark, AJ., Bos, CJ., van Balken, JA., Pouwels, PH. and van den Hondel, CA. (1990). Isolation and molecular characterization of the benzoate-para-hydroxylase gene (*bphA*) of *Aspergillus niger*: a

member of a new gene family of the cytochrome P450 superfamily. *Mol Gen Genet*. **223**: 192-197.

Wang, HJ., Le Dall, MT., Wach, Y., Laroche, C., Belin, JM., Gaillardin, C. and Nicaud JM. (1999). Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the *n*-alkane-assimilating yeast *Yarrowia lipolytica*. *J Bacteriol*. **181**: 5140-5148.

Wang, H., Le Clainche, A., Le Dall, MT., Wache, Y., Pagot, Y., Belin, JM., Gaillardin, C. and Nicaud, JM. (1998). Cloning and characterization of the peroxisomal acyl CoA oxidase *ACO3* gene from the alkane-utilizing yeast *Yarrowia lipolytica*. *Yeast*. **14**: 1373-1386.

Werck-Reichhart, D., Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biology*. **1**: 3003.1-3003.9.

Wright, JD. and Ratledge C. (1991). Isolation of two *Rhodotorula rubra* strains showing differences in the degradation of aromatic compounds. *Appl. Microbiol. Biotechnol*. **35**: 94-99.

Zimmer, T., Iida, T., Schunck, W-H., Yoshida, Y., Ohta, A. and Takagi, M. (1998). Relation between evolutionary distance and enzymatic properties among the members of the *CYP*52A subfamily of *Candida maltosa*. *Biochem. Biophys. Res. Commun*. **251**: 244-247.

Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. and Schunck, WH. (1996). The CYP52 multigene family of Candida maltosa encodes functionally diverse *n*-alkane-inducible cytochromes P450. *Biochem Biophys Res Commun*. **224**: 784-789.

Zimmer, T., Kaminski, K., Scheller, U., Vogel, F. and Schunck, WH. (1995). *In vivo* reconstitution of highly active *Candida maltosa* cytochrome P450 monooxygenase systems in inducible membranes of *Saccharomyces cerevisiae*. *DNA Cell Biol*. **14**: 619-628.

CHAPTER 5

CONCLUDING REMARKS

Cytochrome P450s are heme-containing monooxygenases that are widely distributed in nature and found in all kingdoms. These monooxygenases play a vital role in cell processes such as carbon assimilations, biosynthesis of hormones and detoxification of drugs and xenobiotics. Understanding the mechanism of actions of the cytochrome P450s can contribute to the development of effective therapeutic drugs, understand plant-pathogen interactions and eliminate pollutants in nature.

Rhodotorula sp. CBS 8446 is a basidiomycetous yeast that can grow on decane and limonene vapors. No alkane or fatty acid hydroxylases have been isolated from basidiomycetous yeasts. The ability of *Rhodotorula* sp. CBS 8446 to utilize these substrates has stimulated our interest to isolate an alkane or fatty acid hydroxylase from it.

Primers based on the sequences of the helix I (HI) and heme-binding (HR2) domains of 15 different cytochrome P450 proteins from the CYP52 family, alkane and fatty acid hydroxylases from alkane utilizing ascomycetous yeasts, were successfully used to isolate a gene fragment from *Rhodotorula* sp. CBS 8446. Sequence analysis of the amplified gene fragment showed that it belonged to a P450 gene and southern hybridization showed that only one copy of the gene was present in the genome.

The full-length gene sequence was obtained through inverse PCR (IPCR) and the sequence analysis showed that the amplified gene sequence is interrupted by introns. The 3' RACE technique (Frohman *et al*., 1988) was applied to isolate the downstream region of the gene. By using gene specific walking primers the full-length cDNA sequence that included the start and stop codons was isolated. Comparison with the protein sequences

of other P450s showed that the deduced protein sequence had less than 40% amino acid identity with any classified P450 and the new P450 from *Rhodotorula* sp. CBS 8446 was therefore assigned to a new P450 family CYP557 and given the number CYP557A1 (http://drnelson.utmem.edu/biblioC.html).

The mapping of genomic DNA and cDNA sequences revealed that the sequence for genomic DNA is interrupted by 13 introns. Four of the 13 introns did not conform to the canonical rule that introns start with GT and end with AG (Mount *et al*., 1982). Deviations from the canonical rule might play a role in the splicing of genes. Alternatively spliced variants were observed in transcripts of the isolated gene, but the transcripts that retained introns failed to give an ORF and the introns which did not splice correctly were not the ones that did not conform to the GT/AG rule. Alternative splicing has been observed in mammalian P450s where it generates the functional diversity of the cytochrome P450 family (Christmas *et al*., 2001). However, the significance of alternative splicing in the *CYP557A1* gene remains to be determined.

Amino acids sequence analysis of CYP557A1 detected all the conserved motives characteristic of P450s including the conserved helix I (HI) and heme-binding (HR2) domains. Hydrophobicity prediction revealed two hydrophobic segments in the N-terminal, which is an indication that the isolated P450 is associated with microsomes. There are also 18 amino acids that are conserved in CYP557A1 and in the confirmed fatty acid hydroxylases belonging to the CYP52, CYP94, CYP86 and CYP102 families, but not in P450s with other activities. Six amino acids which are conserved in the other fatty acid hydroxylases, but not in CYP557A1 might be of interest for future structure-function relationship studies.

Blast searches of the NCBI databases showed that CYP557A1 has the highest amino acid identity with P450s belonging to the CYP86 and CYP94 families from plants. These P450s are exclusively fatty acid hydroxylases and mostly omega hydroxylases, while the CYP52s from ascomycetous yeasts are both fatty acid and *n*-alkane hydroxylases. It is thus possible that CYP557A1 is also only a fatty acid hydroxylase and not an alkane

hydroxylase. Heterologous expression in a suitable host system will make it possible to investigate the substrate specificity and regioselectivity of CYP557A1.

Most studies on the heterologous expression of P450s have concentrated on the expression of mammalian P450s in *E. coli* (Dong and Porter, 1996; Fisher *et al.,* 1992; Iwata *et al*., 1998), or occasionally in *S. cerevisiae* (Sakaki *et al*., 1994; Murakami *et al*., 1990*)*. The only fungal P450s which have been studied in some detail are the CYP52s from *Candida tropicalis* and *Candida maltosa*, which were also expressed in *S. cerevisiae* (Zimmer *et al*., 1995; 1996; Schunck *et al.,* 1991). A wide range of expression tools have been developed for the yeast *Yarrowia lipolytica* and *Y. lipolytica* has also successfully been used for the expression of mammalian P450s. In the current study *Y. lipolytica* was therefore used as a host for heterologous expression of two P450s from *Rhodotorula* spp. namely CYP53B1*,* the benzoate *para*-hydroxylase from *Rhodotorula minuta* and the newly identified CYP557A1 from *Rhodotorula* sp. CBS 8446. The same vectors JMP62, JMP64 and JMP21-CPR, which were used by Nthangeni *et al*. (2004) for the expression of human CYP1A1, were used for the expression of the *Rhodotorula* P450s.

The *CYP53B1* gene isolated from *Rhodotorula minuta* was used as a control to test if a gene from a basidiomycetous fungus could successfully be expressed in *Y. lipolytica,* an ascomycetous yeast. In this case it was easy to detect activity with whole cells of *Y. lipolytica,* because *Y. lipolytica* cannot hydroxylate benzoic acid or metabolize *para*-hydroxy benzoic acid. *Para*-hydroxy benzoic acid is also a product with strong UV absorbance on TLC plates, which made it easy to screen for transformants with benzoate *para*-hydroxylase activity. In order to obtain maximum benzoate *para*-hydroxylase activity it was necessary to insert multiple copies of the *CYP53B1* gene and to also insert an additional copy of the *Y. lipolytica* cytochrome P450 reductase (*CPR*) gene. These results correlated with the results obtained by Nthangeni *et al*. (2004) for the expression of human CYP1A1. Thus, P450 activity can be significantly increased if both components of cytochrome P450 are co-expressed and are in the correct ratio. The CPR

should form a complex with the cytochrome P450 to facilitate a transfer of electrons from NADPH to the cytochrome P450 (Sutter *et al*., 1990; Backes and Kelly, 2003).

Heterologous production of CYP557A1 in *Y. lipolytica* is more complicated since *Y. lipolytica* has its own fatty acid and alkane hydroxylases (Iida *et al*., 1998; 2000). In our laboratory, it was demonstrated that expression of multiple copies of the *Y. lipolytica* P450s *YlALK1* and *YlALK2* in *Y. lipolytica* E150 resulted in an increased rate of production of phenyl acetic acid from alkylbenzenes (Dr. E. Setati, personal communication). Phenyl acetic acid is also a product with UV absorbance which can easily be detected on TLC and which is not further metabolized by *Y. lipolytica*. When multiple copies of *CYP557A1* were cloned into a *Y. lipolytica* E150 strain already carrying an additional copy of the *Y. lipolytica CPR* none of the transformants displayed an increased rate of phenyl acetic acid production from alkylbenzenes. These results were in accordance with protein sequence analysis results which indicated that CYP557A1 showed higher amino acid identity with the P450s from plants, which have only omega-hydroxylase activity and no alkane hydroxylase activity. It is unfortunately very difficult to test omega hydroxylase activity in whole cells of *Y. lipolytica* E150, which still has intact β-oxidation. *Y. lipolytica* utilizes the oleic acid as carbon and energy source and incorporates it into lipids. Therefore no accumulation of any product occurs, but strains with higher fatty acid omega hydroxylase activity can be expected to utilize oleic acid faster. It was therefore encouraging that transformants which contained an additional copy of the CPR and multiple copies of the *CYP557A1* gene showed in several experiments the fastest consumption of oleic acid when compared with the control strain carrying only the additional *CPR* gene and with strains carrying the *CPR* gene and a single copy of the *CYP557A1* gene.

In this study we have demonstrated the use of *Y. lipolytica* for the heterologous expression of other fungal P450s. The expression system did not require the use of microsomal fractions, but used whole cells for detecting P450 activity. Being the first report on heterologous expression of fungal P450 genes in *Y. lipolytica*, this system has

paved a way for the expression of other fungal cytochrome P450s. The fact that *Y. lipolytica* can metabolize hydrophobic substrates such as *n*-alkanes and fatty acids might in some cases limit the application of this system. However, the full genome of *Y. lipolytica* has been sequenced (Sherman *et al*., 2004) and the Cre-*lox* recombination system has been adapted for *Y. lipolytica,* which makes it possible to delete large numbers of genes with subsequent recovery of genetic markers (Fickers *et al*., 2003). It is therefore possible to engineer strains specifically for the heterologous expression of P450s. With its natural ability to take up hydrophobic substrates and to produce P450s *Y. lipolytica* can become very useful for the production of hydroxylated products from hydrophobic substrates.

References

Backes, WL. and Kelley, RW. (2003). Organization of multiple cytochrome P450s with NADPH-cytochrome P450 reductase in membranes. *Pharmacol Ther*. **98**: 221-233.

Christmas, P., Jones, JP., Patten, CJ., Rock, DA., Zheng, Y., Cheng, SM., Weber, BM., Carlesso, N., Scadden, DT., Rettie, AE. and Soberman, RJ. (2001). Alternative splicing determines the function of CYP4F3 by switching substrate specificity. *J Biol Chem*. **276**: 38166-38172.

Dong, J. and Porter, TD. (1996). Coexpression of mammalian cytochrome P450 and reductase in *Escherichia coli. Arch. Biochem. Biophys*. **327**: 254-259.

Fickers, P., Le Dall, MT., Gaillardin, C., Thonart, P. and Nicaud, J-M. (2003). New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *J Microbiol Methods*. **55**: 727-737.

Fisher, CW., Shet, MS., Caudle, DL., Martin-Wixtrom, CA. and Estabrook, RW. (1992). High-level expression in *Escherichia coli* of enzymatically active fusion proteins containing the domains of mammalian cytochromes P450 and NADPH-P450 reductase flavoprotein. *Proc Natl Acad Sci U S A*. **89**: 10817-10821.

Frohman, MA., Dush, MK., Martin, GR. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A*. **85**: 8998-9002

Iida, T, Ohta, A. and Takagi, M. (1998). Cloning and characterization of an *n*-alkane-inducible cytochrome P450 gene essential for *n*-decane assimilation by *Yarrowia lipolytica*. *Yeast*. **14**: 1387-1397.

Iida, T., Sumita, T., Ohta, A. and Takagi, M. (2000). The cytochrome P450ALK multigene family of an *n*-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning

and characterization of genes coding for new CYP52 family members. *Yeast*. **16**: 1077-1087.

Iwata, H., Fujita, K., Kushida, H., Suzuki, A., Konno, Y., Nakamura, K., Fujino, A. and Kamataki, T. (1998). High catalytic activity of human cytochrome P450 co-expressed with human NADPH-cytochrome P450 reductase in *Escherichia coli*. *Biochem Pharmacol*. **55**: 1315-1325.

Mount, SM. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res*. **10**: 459-472.

Murakami, H., Yabusaki, Y., Sakaki, T., Shibata, M., and Ohkawa, H. (1990). Expression of cloned yeast NADPH-cytochrome P450 reductase gene in *Saccharomyces cerevisiae*. *J. Biochem*. **108**: 859–865.

Nthangeni, MB., Urban, P., Pompon, D., Smit, M. and Nicaud, J-M. (2004). The use of *Yarrowia lipolytica* for the expression of human cytochrome P450 1A1. *Yeast* (in press).

Sakaki, T., Kominami, S., Takemori, S., Ohkawa, H., Akiyoshi-Shibata, M. and Yabusaki, Y. (1994). Kinetic studies on a genetically engineered fused enzyme between rat cytochrome P4501A1 and yeast NADPH-P450 reductase. *Biochemistry*. **33**: 4933-4939.

Schunck, WH., Vogel, F., Gross, B., Kargel, E., Mauersberger, S., Kopke, K., Gengnagel, C. and Muller HG. (1991). Comparison of two cytochromes P-450 from *Candida maltosa*: primary structures, substrate specificities and effects of their expression in *Saccharomyces cerevisiae* on the proliferation of the endoplasmic reticulum. *Eur J Cell Biol*. **55**: 336-345.

Sherman, D., Durrens, P., Beyne, E., Nikolski, M. and Souciet, JL. for the Genolevures Consortium. (2004). Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. *Nucleic Acids Res*. **32**: D315-D318.

Sutter, TR., Sanglard, D., Loper, JC. and Sangard, D. (1990). Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from *Candida tropicalis*. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins. *J Biol Chem*. **265**: 16428-16436.

Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. and Schunck, WH. (1996). The CYP52 multigene family of *Candida maltosa* encodes functionally diverse *n*-alkane-inducible cytochromes P450. *Biochem Biophys Res Commun*. **224**: 784-789.

Zimmer, T., Kaminski, K., Scheller, U., Vogel, F. and Schunck, WH. (1995). *In vivo* reconstitution of highly active *Candida maltosa* cytochrome P450 monooxygenase systems in inducible membranes of *Saccharomyces cerevisiae*. *DNA Cell Biol*. **14**: 619-628.

SUMMARY

Cytochrome P450s are heme-containing monooxygenases that are widely distributed in nature and found in all kingdoms. These monooxygenases play a vital role in cell processes such as carbon assimilation, biosynthesis of hormones and detoxification of drugs and xenobiotics. Understanding the mechanism of action of the cytochrome P450s can help to develop effective therapeutic drugs, understand plant-pathogen interactions and eliminate pollutants in nature.

Yeasts belonging to the basidiomycetous genus *Rhodotorula* can grow on non-carbohydrate carbon sources such as alkanes, aromatic compounds and even monoterpenes. Only one cytochrome P450 encoding gene, that of a benzoate-*para*-hydroxylase (*CYP53B1*) has been isolated from a basidiomycetous yeast, *Rhodotorula minuta*. *Yarrowia lipolytica* is an ascomycetous yeast that can metabolise hydrophobic substrates such as *n*-alkanes and fatty acids. A wide range of expression tools have been developed for *Y. lipolytica*. It was the goal of this study to isolate a second cytochrome P450 encoding gene, coding for an alkane or fatty acid hydroxylase, from a limonene utilizing *Rhodotorula* sp. and to express this gene as well as the *CYP53B1* gene from *R. minuta* in *Y. lipolytica*.

PCR (using primers based on the sequences of the helix I and heme binding domains of 15 alkane and fatty acid hydroxylases from the ascomycetous yeasts) and IPCR were used to clone a cytochrome P450 encoding gene from the limonene utilizing strain *Rhodotorula* sp. CBS 8446. The sequence analysis of the full-length gene sequence showed that the amplified gene is interrupted by introns and southern hybridization showed that only one copy of the gene was present in the genome. By using gene specific primers the full-length cDNA sequence that included the start and stop codons was isolated. Comparison with the protein sequences of other P450s showed that the deduced protein sequence had less than 40% amino acid identity with any classified P450 and therefore it was assigned to a new P450 family CYP557 and given the number CYP557A1. BLAST searches of the NCBI databases showed that CYP557A1 has the highest amino acid identity with fatty acid omega hydroxylases belonging to the CYP86 and CYP94 families from plants.

The benzoate-*para*-hydroxylase encoding gene *CYP53B1* from *R. minuta* was first expressed into *Y. lipolytica* using the same expression system previously used for the expression of a human cytochrome P450, CYP1A1. In order to obtain maximum benzoate *para*-hydroxylase activity with whole cells of *Y. lipolytica* multiple copies of the *CYP53B1* gene as well as an additional copy of the *Y. lipolytica* cytochrome P450 reductase (*CPR*) gene had to be inserted into *Y. lipolytica*.

Expression of *CYP557A1* in *Y. lipolytica* was more complicated, because *Y. lipolytica* has its own fatty acid and alkane hydroxylases. Biotransformation of hexylbenzene to phenyl acetic acid can, however, be used to detect increased levels of alkane hydroxylase activity. When multiple copies of *CYP557A1* were cloned into a *Y. lipolytica* strain already carrying an additional copy of the *Y. lipolytica CPR,* none of the transformants displayed an increased rate of phenyl acetic acid production from hexylbenzene. It is possible that CYP557A1 has no alkane hydroxylase activity, but only fatty acid hydroxylase activity. It was very difficult to test fatty acid hydroxylase activity with whole cells of a *Y. lipolytica* strain, which still has an intact β-oxidation system, because it utilizes the test substrate, oleic acid, as carbon and energy source and incorporates it into lipids. However transformants that contained an additional copy of the *CPR* and multiple copies of the *CYP557A1* gene showed the fastest consumption of oleic acid in comparison with the control strain carrying only the additional *CPR* gene and with strains carrying the *CPR* gene and a single copy of the *CYP557A1* gene.

In this study we have demonstrated the use of *Y. lipolytica* for the heterologous expression of other fungal P450s. Whole cells instead of microsomal fractions could in this case be used for the detection of P450 activity. This is the first study on heterologous expression of fungal P450 genes in *Y. lipolytica*, and has paved the way for the expression of other fungal cytochrome P450s in *Y. lipolytica*.

Keywords: Cytochrome P450, CYP52 family, helix I, heme-binding, *Rhodotorula, Yarrowia lipolytica*

OPSOMMING

Sitochroom P450 is heem bevattend mono-oksigenases wat algemeen in die lewende organismes voorkom. Hierdie ensieme speel ‘n belangrike rol in prosesse soos koolstof assimilasie, biosintese van hormone en detoksifisering van xenobiotiese verbindings. As die aktiwiteit van P450’s beter verstaan kan, word kan hierdie ensieme gebruik word in die ontwikkeling van terapeutiese middels, die bestudering van plant patogeen interaksie asook bioremediëring.

Giste wat aan die basidiomisete genus *Rhodotorula* behoort kan groei op nie-koolhidraat koolstofbronne soos alkane, aromatiese verbindings en selfs monoterpene. Slegs een sitochroom P450 koderende geen, die van ‘n bensoaat *para*-hydroksilase (*CYP53B1*), is al uit ‘n basidiomisete gis, *Rhodotorula minuta,* geïsoleer. *Yarrowia lipolytica* is ‘n askomisete gis wat hidrofobiese substrate soos n-alkane en vetsure kan benut. ‘n Wye reeks uitdrukking-sisteme is reeds vir *Y. lipolytica* ontwikkel. Dit was die doel van hierdie studie om ‘n tweede P450 koderende geen (een wat kodeer vir ‘n alkaan of vetsuur hidroksilase) uit ‘n *Rhodotorula* sp, wat limoneen benut, te isoleer en hierdie geen sowel as die *CYP53B1* geen van *R. minuta* in *Y. lipolytica* uit te druk.

.

PKR (wat gebruik gemaak het van priemstukke gebaseer op die basispaaropeenvolging van die heliks I en die heem bindings domein van 15 alkaan- en vetsuur-hidroksilases van askomisete giste) en OPKR (omgekeerde PKR) was gebruik om ‘n sitochroom P450 te isoleer uit ‘n limineen benuttende stam *Rhodotorula* sp. CBS 8446. Basispaaropeenvolging- analise van die totale geen het gewys dat hierdie geen introne bevat en Southern-klad-hibridisasie het gewys dat daar slegs een kopie van die geen teenwoordig is in die genoom. Deur geen spesifieke priemstukke te gebruik is die volledige cDNA insluitend die begin en eind-kodons geïsoleer. Vergelyking met ander P450 proteïene het getoon dat die afgeleide aminosuur opeenvolging van die nuwe P450 minder as 40% ooreenstemming met enige bekende P450 het, dit was dus toegeken aan ‘n nuwe P450 familie, CYP557 en die nommer CYP557A1 gegee. BLAST soektogte van die NCBI databank het gewys dat CY557A1 die beste ooreenkoms het met plant vetsuur-hidroksilases naamlik CYP86 en CYP94 families.

Die bensoaat *para*-hidroksilase koderende geen, *CYP53B1* van *R. minuta,* is eerstens gekloneer in *Y. lipolytica* deur gebruik te maak van dieselfde uitdrukkings sisteem wat voorheen gebruik is vir die uitdrukking van 'n menslike sitochroom P450, CYP1A1. Om die hoogste vlakke van bensoaat-hidroksilase aktiwiteit met heel selle van *Y. lipolytica* te verkry, moes verskeie kopieë (10-13 kopieë) van die *CYP53B1* geen sowel as 'n addisionele kopie van die *Y. lipolytica* sitochroom P450 reduktase geen in *Y. lipolytica* ingevoeg word.

Uitdrukking van *CYP557A1* in *Y. lipolytica* was meer ingewikkeld omdat hierdie gis sy eie vetsuur- en alkaan hidroksilase bevat. Biotransformasie van heksielbenseen na fenielasynsuur kan egter gebruik word om verhoogde alkaan-hidroksilase aktiwiteit waar te neem. Die klonering van verskeie kopieë van *CYP557A1* in 'n *Y. lipolytica* transformant wat alreeds 'n addisionele kopie van die *CPR* bevat het, het geen transformante gelewer wat verhoogde feniel-asynsuur produkise vanaf heksielbenseen vertoon het nie. Dit is moontlik dat CYP557A1 geen alkaan-hidroksilase aktiwiteit het nie, maar slegs vetsuur-hidroksilase aktiwiteit. Dit was moeilik om vetsuur-hidroksilase met heel selle van *Y. lipolytica* wat nog 'n intakte β-oksidasie sisteem het te toets aangesien dit die toets substraat, oleïnsuur, as beide koolstof en energiebron benut en dit in lipiede inkorporeer. Transformante met 'n addisionele kopie van die *CPR* geen en verskeie kopieë van die *CYP557A1* geen, het oleïnsuur egter vinniger opgeneem as die kontrole stam wat slegs 'n addisionele *CPR* geen bevat het en transformante wat die *CPR* geen en 'n enkel kopie van die *CYP557A1* bevat het.

In hierdie studie het is bewys dat *Y. lipolytica* gebruik kan word vir die heteroloë uitdrukking van P450 gene uit fungi. Heel selle in plaas van mikrosomale fraksies kon in hierdie geval gebruik word vir die bepaling van P450 aktiwiteit. Hierdie is die eerste studie van hetereloë uitdrukking van P450 gene vanuit fungi in *Y. lipolytica* en het die weg gebaan vir die uitdrukking van ander soortgelyke P450 gene in *Y. lipolytica*.