Expression and localization of four putative fatty aldehyde dehydrogenases in *Yarrowia lipolytica*

by

Walter Joseph Müller

January 2006

University of the Free State
Universiteit van die Vrystaat

Expression and localization of four putative fatty aldehyde dehydrogenases in *Yarrowia lipolytica*

by

Walter Joseph Müller

B.Sc. Hons. (UFS)

Submitted in fulfillment of the requirements for the degree

MAGISTER SCIENTIAE

In the Faculty of Natural and Agricultural Sciences

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

Bloemfontein

South Africa

January 2006

Supervisor: Dr. J. Albertyn

Co-supervisor: Prof. M.S. Smit

Acknowledgements

The following individuals and institutions are hereby greatly acknowledged.

Dr. Jacobus Albertyn. Thank you very much for your willingness to help and for always providing me with the molecular tools I needed to perform this study. You allowed me to make mistakes and made sure I learnt from them – that is a very valuable lesson, especially in molecular biology. Thank you for that!

Prof. Martie. Smit. You are always so enthusiastic about research and that sometimes made even the negative results look good. Thank you for always being so curious about my work and for your constructive criticism. It is greatly appreciated!

Drs. Jean-Marc Nicaud and **Catherine Madzak** from INRA in Paris (France) who provided me with useful e-mails, vectors and plasmid maps. Thank you for always responding and for your willingness to assist me whenever I was facing a dead end. Special thanks to Dr. Jean-Marc Nicaud who provided us with the nucleotide sequences of the fatty aldehyde dehydrogenases – without it this study could not have been launched.

Prof. Stephan Mauersberger from the Technical University in Dresden (Germany). You always responded to my e-mails and provided wonderful suggestions and supplements that assisted me greatly – even if they were in German sometimes!

I wish to express my sincerest gratitude to my colleagues at the Department of Microbial, Biochemical and Food Biotechnology for their friendly faces; encouragement and making every day seem a bit more tolerable when the research was getting me down. In particular I wish to thank my colleagues in the Molecular Biology Laboratory: **Lallie**, **Michel**, **Michelle**, **Olga** and **Sanet** for their wisdom, suggestions, time but most of all for their friendship. Thank you!

I wish to thank and acknowledge the **National Research Foundation** (NRF) in South Africa for providing me with a Prestigious bursary (GUN: 2066887) to support this study. I am extremely grateful!

Thank you to my loving parents, **Joe** and **Engela** who always supported my studies and impressed on me at an early age the importance of tertiary education. Thank you very much for all your sacrifices, encouragement and love. You have no idea what it means to me.

Thanks to my twin sister, **Beaula** for always giving me courage and for believing in me and telling me that I work too much and go out to little!

A special word of thanks to my long-time friend **Riaan**. Thank you for the conversations, the listening when I was ranting and raving and for your understanding when I was feeling demotivated and frustrated. I appreciate your friendship tremendously!

Thank you to the **Thompson family** in East London. You are my second family away from home and from my studies. I love you all dearly!

Lastly, I wish thank the **Heavenly Father**. Thank you for providing me with all of these wonderful opportunities. Thank you for giving me the determination and strength when I was feeling down but mostly for Your love of which I am not worthy. Glory to You always!

This dissertation is dedicated to my parents (Joe & Engela), my twin sister (Beaula) and my baby brother (Stefan).

“I have no special talents – I am just passionately curious!”

-Albert Einstein-

“Obstacles are those frightful things you see when you take your eyes off your goal.”

-Henry Ford-

List of Essential Abbreviations

FALDH = Fatty Aldehyde Dehydrogenase (gene)

FALDH = Fatty Aldehyde Dehydrogenase (protein)

GFP = Green Fluorescent Protein originating from *Aequorea victoria* isozyme (x = denotes the number of the isozyme)

LS = Localization Sequence

PlacT = Promoter *lacZ* Terminator expression cassette

PT-cassette = Promoter-Terminator cassette

p*Yl*FALDHx = Putative promoter of each *Yarrowia lipolytica* FALDH

*Rn*FALDH = *Rattus norvegicus* Fatty Aldehyde Dehydrogenase

SEP = Sticky-end Polymerase Chain Reaction

TA proteins = Tail-anchored proteins

Vh-ALDH = *Vibrio harveyi* Fatty Aldedyde Dehydrogenase

*Yl*FALDH = *Yarrowia lipolytica* Fatty Aldehyde Dehydrogenase

Table of contents

Chapter 1

Literature survey

1.1 Introduction

Saturated hydrocarbons such as aliphatic, branched and cyclic alkanes are highly reduced forms of carbon that are produced by geochemical processes from decaying plant and algal material. Alkanes constitute about 20-50 % of crude oil, depending on the source of the oil. In addition, alkanes (predominantly long-chain compounds) are produced throughout the biosphere by plants, algae and bacteria as a waste product, structural element, defense mechanism or as a chemoattractant (van Beilen *et al*., 2003). Alkanes which are in excess, whether from biological reactions or pollution need to be degraded. Pollution of terrestrial environments by oil discharges are mainly controlled by physico-chemical and natural degradation measures. However, the bulk of these polluting oils are destroyed in the environment by microbes including bacteria, yeasts, molds and algae, since they possess the capability to utilize petroleum hydrocarbons as carbon source (Ekundayo & Obuekwe, 2000).

Yeasts in particular are well known to utilize alkane, alkene, thiophene and poly-aromatic hydrocarbons and have been extensively studied for the production of single cell protein and emulsifying agents from these substrates (Kim *et al*., 1999). However, relatively little information about the hydrocarbon-degradation potential of yeasts is still available (Margesin *et al*., 2003). Approximately 20 % of the nearly 500 yeast species, mainly belonging to the genera *Candida*, *Pichia* and *Yarrowia* are able to grow alternatively either on carbohydrates or on middle- or long-chain *n*-alkanes as their sole source of carbon and energy (Barth & Gaillardin, 1996).

n-Alkanes (C_nH_{2n+2}) received considerable attention in the mid-sixties as an inexpensive carbon source for the production of especially citrate and isocitrate. High yields of α-keto-glutarate are also obtained when *n*-alkanes are employed as sole carbon source in conjunction with a thiamine limitation (Barth and Gaillardin, 1997). In addition, growth

on alkanes and fatty acids result in long-chain carboxylic acids. These long-chain carboxylic acids are versatile raw materials for the oleochemical industry and are used in the production of fragrances, polyamides, adhesives and macrolide antibiotics (Vanhanen *et al*., 2000). Other *n*-alkane technologies that have been investigated are the production of wax esters from fatty alcohols and the synthesis of poly-3-hydroxy- alkanoate which could have application for the production of biologically-derived plastics (Watkinson & Morgan, 1990; Barth & Gaillardin, 1997; Ishige *et a*l., 2002).

n-Alkanes that are aerobically degraded yield corresponding fatty acids as final metabolites. These fatty acids are most often degraded by the β-oxidation system, which is exclusively localized in the peroxisomes (Ratledge, 1984; Endrizzi *et al*., 1996). The key enzymes prior to β-oxidation include: monooxygenases (belonging to the cytochrome P450 family), fatty alchohol oxidase (FAO) and fatty aldehyde dehydrogenases (FALDH). Several cytochrome P450 (CYP P450) isozymes involved in alkane assimilation have recently been cloned from yeasts like *Candida maltosa*, *Debarymyces hansenii* as well as *Yarrowia lipolytica* (Ohkuma *et al*., 1998; Yadav and Loper, 1999; Iida *et al*., 2000). Fatty alcohol dehydrogenation has been ascribed to a NAD(P)-dependent alcohol dehydrogenase as well as a membrane-bound, flavin-dependent FAO. Genes coding for the latter have been isolated from several *Candida* spp. (Vanhanen *et al*., 2000; Eirich *et al*., 2004).

FALDH activity has been detected in *C. maltosa* and other yeasts (Mauersberger *et al*., 1996) but to date; the *FALDH* gene has not been cloned and characterized. Genes coding for enzymes with confirmed FALDH activity have been cloned and sequenced from mammals (human, rat and mouse) (de Laurenzi *et al*., 1996; Miyauchi *et al*., 1991; Vasiliou *et al*., 1996) as well as from three bacteria (an *Acinetobacter* sp., *Pseudomonas putida* and *Vibrio harveyi*) (Ishige *et al*., 2000; Kok *et al*., 1989 ;Vedadi *et al*., 1995). A review of these confirmed FALDHs was compiled to aid in the characterization of four putative *FALDH* encoding genes indentified in the recently sequenced genome of the *n*-alkane utilizing yeast *Y. lipolytica* (Fickers *et al.*, 2005).

1.2 Aldehyde dehydrogenases: A general introduction

Aldehyde dehydrogenases (ALDHs), (E.C. 1.2.1.3) comprise a superfamily of ubiquitous enzymes which catalyze the oxidation of a wide spectrum of endogenous and exogenous aldehydes (R-CHO) to their corresponding carboxylic acids (R-COOH). All ALDHs are $NAD(P)^+$-dependent and catalize the oxidation of a wide variety of aliphatic and aromatic aldehydes but will vary in their specificity for these substrates (Ratledge *et al*., 1984; Perozich *et al*., 2001; Sophos *et al*., 2001; Vasiliou *et al*., 2000). ALDHs occur throughout all phyla and a genome analyses done in 2003 revealed 555 distinct genes of which 32 were found in archae, 351 in eubacteria and 172 in eukaryota (Table 1.1) (Sophos & Vasiliou., 2003). These results are indicative of the cardinal role these enzymes play in biological functions. The latter is justified since most aldehydes are toxic at low levels due to their chemical reactivity. Consequently, these toxic metabolites have to be carefully regulated and detoxified by ALDHs.

Table 1.1 Summary of the ALDH genes

Superkingdom	Taxon	Number of genes	Total
Archae	Crenarchaeota	12	
	Euryarchaeota	20	32
Eubacteria	Aquificales		
	Cyanobacteria	2	
	Firmicutes	5	
	Fusobacteria	113	
	Proteobacteria	2	
	Spirochaetales	216	
	Thermotogales	1	
	Thermus/Deinococcus group	1	
		11	351
Eukaryota	Diplomonadida	1	
	Euglenozoa	2	
	Entamoebidae	2	
	Fungi	32	
	Metazoa	90	
	Viridiplantae	45	172

Detoxification of aldehydes is accomplished by oxidation of the carbonyl functional group to acids, which are then further degraded. However, a number of ALDH-mediated oxidations form products that are known to possess significant biological, therapeutic and/or toxic activities. These include: retinoic acid, an important element for vertebrate development, γ-aminobutyric acid (GABA), an important neurotransmitter and trichloroacetic acid, a potential toxicant (Vasiliou *et al*., 2000).

1.3 The ALDH gene superfamily

In phylogenetic terms a gene superfamily is defined as a cluster of evolutionary related sequences and consists of homologous gene families, which are clusters of genes from different genomes that include both orthologs and paralogs. Orthologs are genes in different species that evolved from a common ancestor by separation, whereas paralog genes are products of gene duplication events within the same genome (Vasiliou *et al*., 2000). As with other superfamilies, members of the ALDH superfamily are classified to 'gene families' and 'subfamilies' based on the percentage identity of each protein compared to the others. An ALDH protein is designated to a family if it has < 40% similarity to that of any other family and proteins that display ≥ 60% sequence similarity are considered to belong to the same subfamily (Sophos & Vasiliou, 2003).

A standardized ALDH gene nomenclature system similar to the system used for other superfamilies and accepted by both the Human and Mouse Genome Projects have also been instituted for the ALDH superfamily (Sophos & Vasiliou, 2003). In this system the root symbol **"*ALDH*"** denoting "**al**dehyde **de**h**h**ydrogenase" is followed by an Arabic number representing the family, and - when needed - a letter designating the subfamily and an Arabic number denoting the individual gene within the subfamily. The 172 eukaryotic ALDH genes identified up to 2003 are currently classified into 20 families. Eukaryotic ALDH gene families 1,4,5,10 and 18 also contain fungal ALDHs and gene families 14 – 16 are present exclusively in the fungal taxon (Sophos & Vasiliou, 2003).
In 1999 Perozich and co-workers performed a sequence alignment with, at that time, 145 ALDH sequences to determine relationships between the ALDH families. From these

alignments 13 families were assigned (Fig. 1.1). This tree is easier to interpret than later phylogetic trees that only used the standardized numbering of the ALDH families (Sophos & Vasiliou, 2003), because the families were named according to confirmed substrate specificity or the organisms in which the genes were found. Non-specific ALDHs similar to the mammalian ALDH1, ALDH2 and ALDH3 families were termed Class 1, Class 2 and Class 3. The phylogenetic tree constructed by Perozich et al. (1999) comprised of two trunks - a'Class 3' trunk (Class 3 ALDHs down to betaine-ALDHs) and a 'Class 1/2' trunk (Class 1 ALDHs up to Group X).

When considering protein structure Class 1 and 2 ALDHs are broadly distinguished from Class 3 ALDHs by having an extra 56 residues at the NH_2-terminus and by their tetrameric, instead of dimeric quaternary structure. Class 3 ALDHs prefer NAD as coenzyme but can use NADP effectively *in vitro*. They function as dimers of identical ~50 kDa monomers and share about 30% sequence identity with either of the tetrameric Class 1 or 2 enzymes (Liu *et al*., 1997b; Hempel *et al*., 1999).

Mammalian class 3 ALDHs can be divided into the *ALDH3A* and *ALDH3B* subfamilies. Subfamily *ALDH3B* comprises of two structurally related genes, *ALDH3B1* and *ALDH3B2* for which there is currently no functional data available for either enzyme (Vasiliou *et al*., 2000). The *ALDH3A* subfamily is divided into the *ALDH3A1* and *ALDH3A2*, which are primarily involved in the oxidation of medium and long-chain aliphatic and aromatic aldehydes. ALDH3A2 (EC 1.2.1.48), which is also known as FALDH is a microsomal enzyme which catalyzes the oxidation of fatty aldehydes to fatty acids (ALDH3A2 will at times be referred to as human or mammalian FALDH)

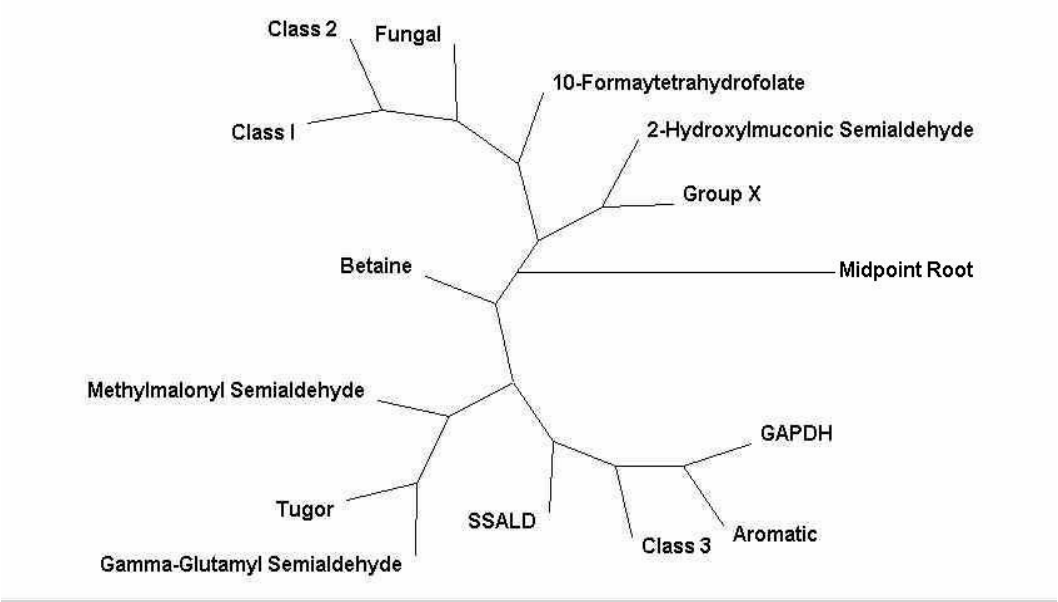


Fig. 1.1 A phylogenetic tree illustrating evolutionary relationships among ALDH subfamilies (Perozich *et al*., 1999).

1.4 The mechanism of aldehyde oxidation

ALDHs are characterized by 10 conserved motifs that reside at or near the active site of the protein (*Fig*. 1.*2*). A large portion of the β-sheet structure is highly conserved. Nearly all conserved motifs contain a turn or loop with a well-conserved small amino acid residue such as glycine (G), proline (P), aspartic acid (D) or asparagine (N). The well-conserved large hydrophobic amino acid side chains in these 10 motifs often point away from the rest of the motif and appear to anchor these elements to the core of the protein. Table 1.2 provides a summary of the 10 conserved amino acid sequence motifs from rat Class 3 ALDH and Figure 1.2 provides the 3 dimensional structure of this protein with the numbered corresponding motifs.

Aldehydes are highly reactive electrophilic compounds, which interact with thiol and amino groups (Vasiliou *et al*., 2000). Insights into to the ALDH mechanism of converting aldehydes to their carboxylic acids have been provided by information on conserved amino acids and by the quaternary structure of the rat Class 3 ALDH (ALDH3A2), for which X-ray crystallographic data is available, in conjunction with other information

derived by site-directed mutagenesis and NMR (Wymore *et al*., 2004). Unless otherwise stated numbering of amino acid residues refer to the rat ALDH.

The quaternary structure of the rat ALDH (a Class 3 ALDH), revealed that each subunit of the dimeric protein contains a NAD-binding domain, a catalytic domain and an oligomerization domain (Liu *et al*., 1997a; Wymore *et al*., 2004).

Table 1.2 Ten most conserved sequence motifs in ALDHs (Adapted from: Perozich *et al*., 1999). Numbering of motives is as indicated on Fig. 1.2.

Motif number	Length	Motif [b]
1	5	[Past]-[WFy]-[**N**e]-[FYgalv]-[**P**tl]
2	14	[Apnci]-[Liamv]-[Avslcimg]-[ACtlmvgf]-**G**-[Ncdi]-[Tavcspg]-[Vaimfcltgy]-[Vil]-[Lvmiwafhcy]-[**K**h]-[Ptvghms]-[ASdhp]-[Epsadqgilt]
3	10	[Grkpwhsay]-[FLeivqnarmhk]-[Pg]-[Plakdievsrf]-[Gnde]-[Vliat]-[VLifyac]-[Nglqshat]-[VIlyaqgfst]-[IVlms]
4	10	[IVlgfy]-[SAtmnlfhq]-[Fyla]-[Tvil]-**G**-[Sgen]-[Tsvrindepaqk]-[EAprqgktvnldh]-[VTiasgm]-[Gafi]
5	16	[Lamfgs]-[Enlqf]-[Ltmcagi]-[**G**s]-[Ga]-[Knlmqshiv]-[SNadc]-[Pahftswv]-[cnlfmgivahst]-[Ivlyfa]-[Viamt]-[Fdlmhcanyv]-[Daeskprnt]-[Dsntaev]-[Acvistey]-[Dnlera]
6	8	[Fyvlma-[Fgylrmdaqetwsvikp]-[Nhstyfaci]-[QAsnhtcmg]-**G**-[Qe]-[crvitksand]-[**C**r]
7	9	[Gdtskac]-[Yfnarthclswv]-[FYlwvis]-[IVlfym]-[Qeapkgrmynhlswyv]-[**P**a]-[Tachlmy]-[V**I**l]-[**F**Livwn]
8	7	[Ektdrqgs]-**E**-[Ivtlnfsp]-**F**-[Ga]-[**P**s]-[Vilef]
9	15	[Nrst]-[Dnaseqtkregi]-[TSrvnalcqgik]-[Epdtgqikvrfshyncl]-[Yfkqvm]-[Gpa]-[Lnmv]-[Astgvqcf]-[Agsltfc]-[AGysct]-[VIlfams]-[Fhwyivlem]-[TSag]-[KRnsqteahdp]-[DNsileakt]
10	12	[Pasw]-[Fwyahv]-[Gtqs]-**G**-[Fvyesnimtawrq]-[Kgrn]-[mqarelnskghdpt[-[Stm]-[Gfls]-[Ifntlmygshrvq]-[Gdnhrsy]-[Rdpsagkte]

[a]Motifs are given as ProSite patterns. Each position is separated by a hyphen. Capitalized letters represent residues that are predominant at each bracketed position. Residues highlighted in bold are conserved in at least 95% of known ALDHs

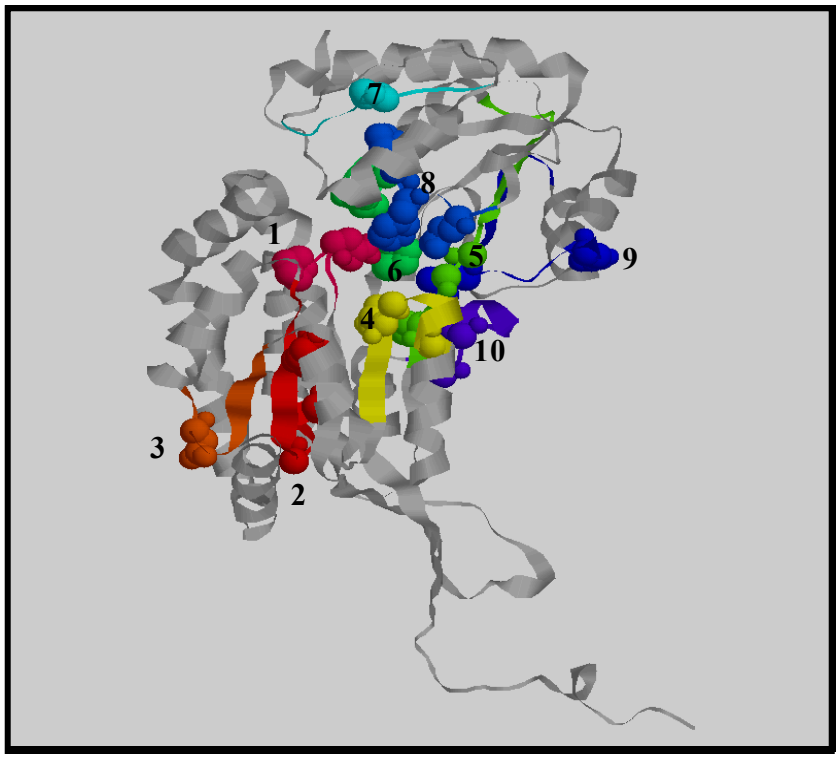


Fig. 1.2 The three dimensional structure of the rat Class 3 aldehyde dehydrogenase monomer determined at a resolution of 2.6 Å. The most highly conserved residues in each motif are shown with space filling side chains. Numbering of the conserved motives correspond to the numbers in table 2. (Perozich *et al*., 1999).

1.4.1 Conserved residues in the catalytic channel

Cysteine (Cys-243 in Class 3, Cys-302 in Class 1 and 2) is the most conserved residue in the ALDH family and has generally been accepted as the catalytic thiol (Liu *et al*., 1997b). The residue closest to the Cys-243 is the Asn-114 (Asn-169 in Class 2 structure) and it is strictly conserved. The Glu-333 helps to create the thiolate of the catalytic Cys-243 and it is strictly conserved and 4.5Å away from Cys-243 (Hempel *et al*., 1999).

The side chain of Cys-243 extends into the center of the catalytic channel with the thiol positioned 6.8 Å from carbon number four of the nicotinamide ring of NAD. The surface of the channel between Cys-243 and the NAD-binding site contains a number of highly conserved residues including Asn-114, Leu-119, Thr-186, Gly-187, Glu-209 Glu-333 and Phe-335. Liu and co-workers (1997b) stated that this concentration of highly conserved residues suggests similar catalytic environments for the Class 1 and 2 enzymes.

1.4.2 General mechanism for aldehyde oxidation

Figure 1.3 depicts a proposed mechanism of aldehyde oxidation based on the tertiary structure of the rat Class 3 ALDH. The mechanism was derived by modeling benzaldehyde into the catalytic site of the protein (Hempel *et al*., 1999).

(I) The substrate carbonyl O projects directly between the side-chain O and N of Asn-114. The thiolate of the catalytic Cys-243 can be generated in concert with Glu-333 through an intervening water molecule in order for it to attack the substrate carbonyl carbon.

(II) After the thiolate attack, a transition state tetrahedral intermediate forms by means of transient stabilizing H-bonding between the substrate carbonyl O and the NH side chain of the conserved Asn-114.

(III) Hydride transfer is accomplished by proton abstraction from a water molecule set between the Cys-243 sulfur and Glu-333.

(IV) After hydride transfer the end result is a second tetrahedral intermediate (thiolester).

(V) When the energy-rich thiolester intermediate collapses, the product acid is formed which may reprotonate the thiolate.

(VI) On dissociation of the products, the enzyme can bind the reactants again.

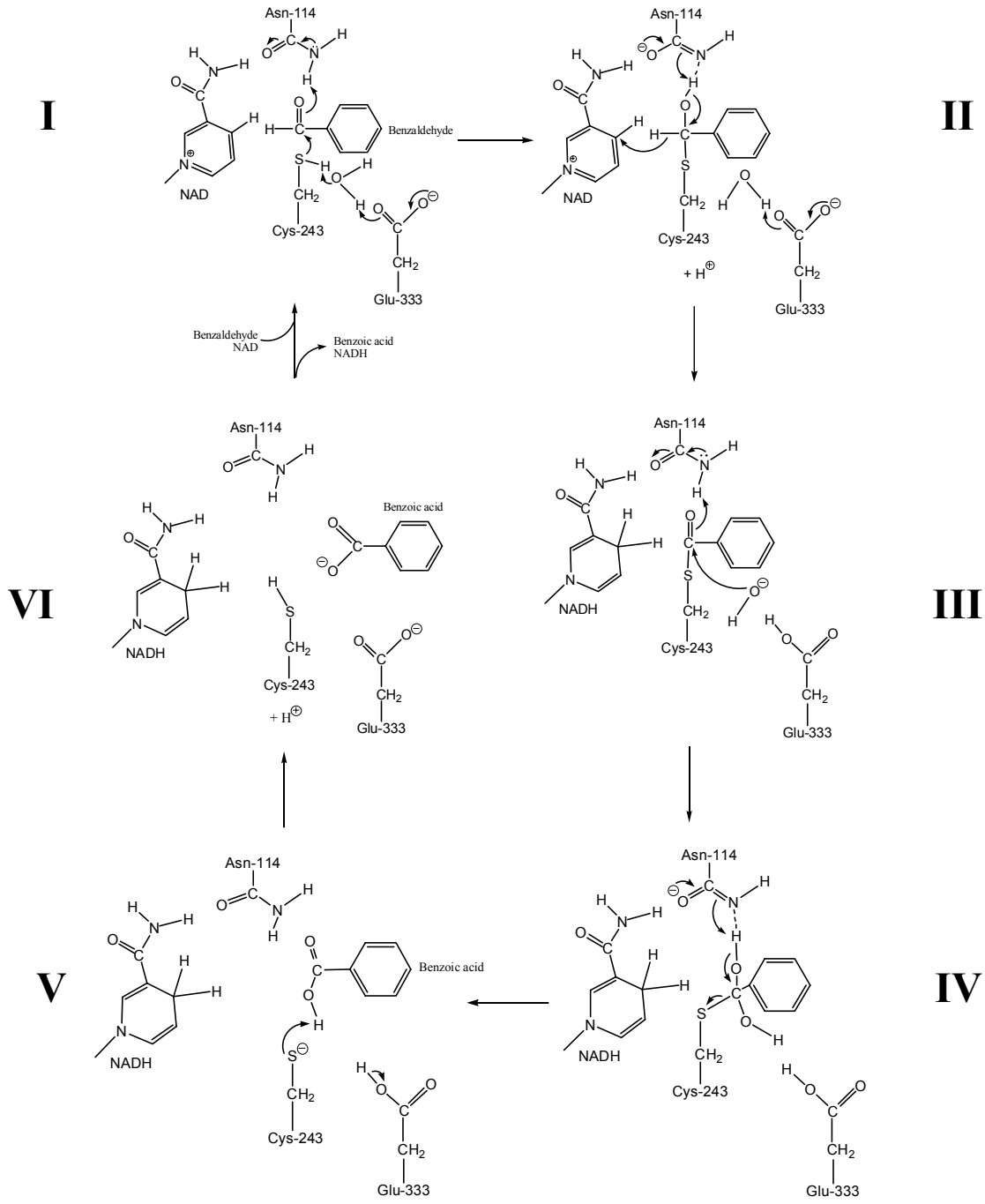


Fig. 1.3 Proposed mechanism of aldehyde oxidation based on Class 3 ALDH tertiary structure (Hempel *et al*., 1999).

1.5 Characterized enzymes with confirmed FALDH activity

1.5.1 Mammalian FALDHs

The existence of FALDH activity in mammalian liver has been known for more than 30 years and the enzyme was often referred to as a 'high K_m' isozyme (Rizzo *et al*., 2001). Substrate specificities suggest that human FALDH, designated ALDH3A2, plays a crucial role in detoxification of aldehyde metabolites of lipid peroxidation, ω-oxidation of 20-CHO-leukotriene B_4 and in the oxidation of aldehydes resulting from alcohol metabolism. This enzyme preferentially catalyzes the oxidation of medium and long-chain saturated and unsaturated aliphatic aldehydes ($C_2 - C_{24}$ in length) to fatty acids with K_m values between 6 and 50μM at pH 9.8. The enzyme also displays high activity (K_m 6μM) towards a 20-carbon branched chain aldehyde, dihydrophytal (Rizzo *et al*., 2001).

FALDH has been purified from human liver, rabbit intestine and rat liver and has a subunit molecular weight of *ca*. 54 kDa, similar to most ALDHs. Human microsomal FALDH has a pH optimum of 9.8 in glycine buffer and is thermostable at 47ºC after 5 minutes (Kelson *et al*., 1997). The native FALDH is probably a homodimer, like its other Class 3 ALDH counterparts but experimental proof for this is lacking since purified FALDH forms large polymeric aggregates and this precludes accurate determination of its native size (Kelson *et al*., 1997; Perozich *et al*., 1999; Rizzo *et al*., 2001).

FALDH is synthesized on free polysomes (an mRNA strand complexed with several ribosomes) and then post-translationally inserted into the ER of the liver although it must be stressed that FALDHs do not undergo post-translational modifications (Schlegel, 1995; Rizzo *et al*., 2001). No human FALDH activity is present in peroxisomes. Rodent FALDH has been localized to the microsomal fraction of liver and intestine based on differential centrifugation methods (Nakayasu *et al*., 1978). Unlike human FALDH, enzyme activity with kinetic properties resembling microsomal ALDH has been detected in rat liver peroxisomes (Lin *et al*., 2000). The human FALDH consists of 485 amino acids but the rodent protein has 484 amino acids, the difference being due to an extra

amino acid at the carboxy-terminus in humans. Human and rat proteins share 84% amino acid identity, whereas the rat and mouse enzymes are 95% identical. FALDHs possess a hydrophobic carboxy-terminal tail consisting of *ca*. 35 amino acids which is a distinctive feature of FALDHs. Site-directed mutagenesis studies of the rat FALDH carboxy-terminus has indicated that this hydrophobic domain is essential for anchoring the protein to the microsomal membrane.

1.5.1.1 Genomic organization and expression of mammalian *FALDH*

The importance of FALDH in human biology is delineated by the fact that genetic deficiency of FALDH in humans is associated with Sjögren-Larson syndrome (SLS), an autosomal recessive, neurocutaneous disorder characterized by mental retardation, spastic di- or tetraplegia and congenital ichthyosis (Lin *et al*., 2000; Vasiliou *et al*., 2000; Rizzo *et al*., 2001).

In both human and mouse the FALDH gene (*ALDH3A2*) is closely linked to *ALDH3A1*. Complete sequence analysis of the human locus indicates that the *ALDH3A2* and ALDH3A1 genes are parlalogs which are 60 kb apart (Fig. 1.4). The close proximity along with similarities in genomic organization and coding sequence suggest that these two genes arose from a duplication event (Vasiliou *et al*., 2000; Rizzo *et al*., 2001).

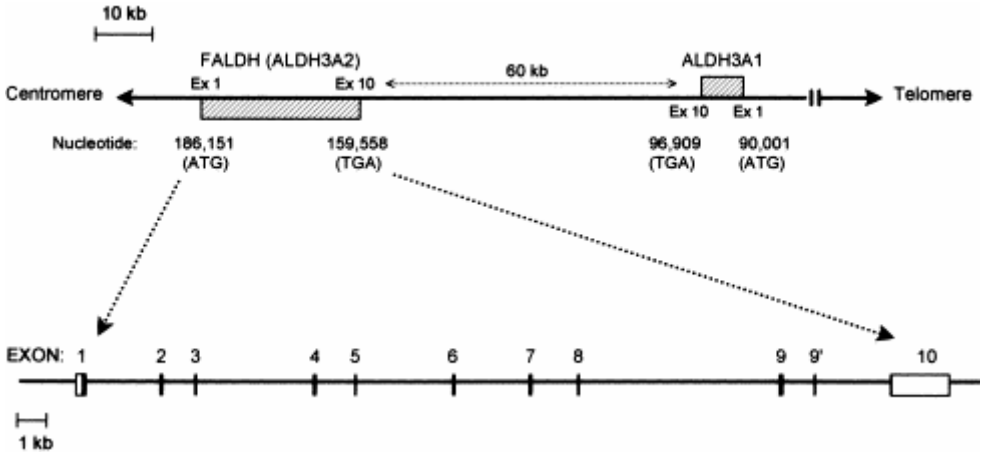


Fig. 1.4 Organization of the human *FALDH* on chromosome 17p11.2. The locus has been sequenced and assembled into a contig of about 191 kb (GenBank accession AC005772) (Rizzo *et al*., 2001).

The ALDH3A1 is coded on the sense strand, whereas ALDH3A2 is encoded on the antisense strand in the opposite orientation. Like the *ALDH3A1* gene, the coding sequence of the *ALDH3A2* gene in interrupted by nine introns. Exon 9 of human *ALDH3A2* is larger than exon 9 of *ALDH3A1* (236 and 131 bp respectively) owing to the presence of a unique sequence encoding a hydrophobic carboxy-terminal domain (see later).

The human and mouse FALDH genes consists of 11 exons and ten introns. A comparison of exon/intron structures of all four Class 3 ALDHs suggests that merging and partitioning of exons have been common events in the evolution of this gene family (Rogers *et al*., 1997; Rizzo *et al*., 2001). The mouse gene spans *ca*. 25 kb and is smaller in comparison to its larger 31 kb human relative due to decreased intron size. However, the intron-exon boundaries are identical in both species. The transcription initiation site in mice is located at nucleotide -121 relative to the translation initiating codon, whereas that of humans is located -258 in relation to the translation initiating codon. In humans FALDH is widely expressed as three transcripts of 2, 3.8 and 4.0 kb, while the major transcript in mice is a 3.0 kb transcript composed of exon 1 – 10. Alternative splicing of

exon 9′, located between exon 9 and 10, has been identified in both humans and mice (Lin *et al*., 2000; Rizzo *et al*., 2001). RT-PCR on mouse liver RNA revealed two distinct exon amplicons. The shorter but more abundant fragment consisted of exon 9 to 10, whereas the less abundant but longer fragment had an additional 113 bp inserted between exon 9 and 10. Translation of this alternatively spliced exon 9′ results in a variant protein called FALDH*v* containg a carboxy-terminal domain comprised of 26 and 27 amino acids in mice and humans respectively of which 11 are hydrophobic (Rizzo *et al*., 2001). All exon and flanking intron sequences are available in GenBank (Accession numbers AF289804 – AF289813).

Both human and mouse promoters lack TATA motifs but possess multiple CpG islands (Lin *et al*., 2000) upstream of the translation initiation codon. In addition, the mouse promoter has a 13 TG repeat at nucleotide -509 to -534 which is absent from the human gene. Human FALDH promoters have a Sp1-binding site 51 bp upstream of the transcription initiation site. The mouse promoter also contains an upstream Sp1-binding site (CG-box) but also has several putative transcription factor-binding sites *e.g*. AP1, N-myc and NF1 (Kelson *et al*., 1997; Rogers *et al*., 1997; Lin *et al*., 2000; Rizzo *et al*., 2001). Northern analyses reveals FALDH is expressed widely in humans and mouse tissues, with the highest expression levels located in the intestines, kidneys and liver.

1.5.1.2 COOH-terminals aid in FALDH localization

Nascent proteins have specific biological functions in different domains of a cell. Before proteins undergo posttranslational modifications *e.g*. glycosilation and phosphorylation, they need to be exported to a specific cellular destination, whether it be intra- or extracellular to fulfill their biological functions. Mature proteins are ‘guided’ by an amino acid sequence that contains information as to where the protein is destined to be exported. Much attention has been given to resolving the mechanism for sorting and targeting of mature proteins to their final destinations and this remains one of the fundamental problems in cell biology (Masaki *et al*., 1994).

Newly synthesized proteins follow two distinct routes, depending on the presence or absence of a localization sequence (LS). These LS are typically 15 – 20 amino acids in length and are predominantly composed of hydrophobic amino acids (Madigan *et al*., 2000). The signal recognition particle (SRP) in the cytosol recognizes and binds to the LS of the nascent protein. The resulting SRP-ribosomal-nascent peptide complexes are targeted to the ER membrane through interactions with the docking protein complex (Rapieko & Gillmore, 1992). After translocation across the ER membrane, the secretory and plasma membrane proteins proceed through the Golgi complex and then to the cell surface. Resident proteins in the central vacuolar system are localized to their final destinations with the aid of either specific targeting or retention signals (Kornfeld & Mellman, 1989; Nilsson *et al*., 1989).

Proteins lacking a LS at their amino termini are synthesized on free polysomes and are posttranslationally directed to intracellular organelles like the mitochondria, peroxisomes, nucleus and the ER. Proteins that are not imported into organelles remain in the cytosol (Masaki *et al*., 1994). A common trait in most COOH- terminal proteins is a hydrophobic region that serves as a membrane anchor (Masaki *et al*., 1994; Horie *et al*., 2002; Masaki *et al*., 2003; Demozay *et al*., 2004).

Borgese *et al*., 2003 define COOH-tail anchored or tail anchored (TA) proteins as a group of integral membrane proteins with a cytosolic NH_2-terminal domain that is anchored to the phopholipid bilayer by a single segment of hydrophobic amino acids close to the COOH-terminus. TA proteins lack a NH_2-terminal LS and their membrane-interacting region is so close to the COOH-terminus that it emerges from the ribosome only upon termination of translation. This hydrophobic COOH-terminus region is therefore unlikely to interact with SRP, which binds signal peptides or signal anchors only as long as they are part of a nascent polypeptide chain (Borgese *et al*., 2003).

TA proteins are classified into two categories according to their intracellular localization: proteins localized along the exocytic pathway and proteins localized to the mitochondrial outer membrane (MOM). Recent studies have revealed that tail-anchored proteins are

first inserted into the ER membrane and subsequently sorted to their respective final destinations by vesicle-mediated trafficking. However, relatively little is known about ER targeting and integration of tail-anchored proteins (Masaki *et al*., 2003). Table 1.3 reveals localization and function of TA proteins.

Table 1.3 Localization and functions of TA proteins

Example of TA protein	**Localization**	**Function**
Cytochrome b_5	ER	**Enzymatic**
MOM isoform of cytochrome b_5	MOM	
Heme oxygenase I and II	ER	
UBC6	ER	
		Protein localization
Sec61γ, Sec 1β	ER	Translocation
TOM5, TOM6	MOM	
Pex1 5p	Peroxisomes	
OMP25	MOM	Adaptors
		Vesicular traffic
Target SNAREs	Target membranes for	SNARE proteins
Vesicular SNAREs	vesicular fusion	
Giatin	Transport vesicles	Tethering proteins
	Golgi complex	
Bcl-2	MOM and ER	**Regulation of apoptosis**
Bcl-X_L	MOM	
Bax	Cytosol and MOM	
Us9 protein of a herpes virus	Trans-Golgi network	**Constituent of viral envelope**

Abbreviations: Bax = Bcl2-associated X protein; Bcl-2 = B-cell CLL/lymphoma 2 protein; Bcl-X_L = an anti-apoptotic member of the Bcl-2 family ; MOM = mitochondrial

outer membrane; OMP25 = outermembrane immunogenic protein precursor; Pex15p = phosphorylated tail-anchored type II integral peroxixomal membrane protein; Sec61β and Sec 61γ = subunits of mammalian integral inner membrane proteins found in secretory (Sec) pathway; SNARE = soluble *N*-ethylmaledimide-sensitive factor attachment protein receptor; TOM = translocase of the outer mitochondrial membrane; UBC6 = ubiquitin conjugating enzyme 6. (Adapted from: Borgese *et al*., 2003).

FALDHs are microsomal enzymes that are synthesized on free-polysomes and then post-translationally inserted into the ER. To date, FALDHs are not known to undergo post-translational modifications (Rizzo *et al*., 2001). Mammalian FALDHs from human, mouse and rat display a characteristic hydrophobic tail (Fig. 1.5) at the COOH-terminal (Masaki *et al*., 1994; Yoshida *et al*., 1998; Lin *et al*., 2000; Rizzo *et al*., 2001). The hydrophobic tail consists of *ca*. 35 amino acids and site directed mutagenesis studies conducted on rat FALDH has indicated that this hydrophobic tail is essential for anchoring the protein to the microsomal membrane (Masaki *et al*., 1994; Vasiliou *et al*., 2000; Rizzo *et al*., 2001).

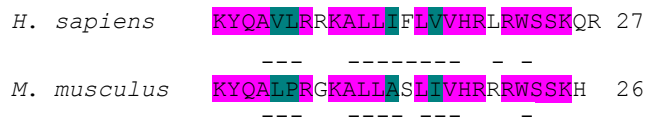


Fig. 1.5 Alignment of human and mouse COOH-terminal sequences respectively. Underlined amino acids represent hydrophobic residues. Sequence alignment was performed using DN*Assist* 2.2. (Adapted from: Rizzo *et al*., 2001).

R. norvegicus FALDH is a membrane-bound enzyme that is a useful model protein for studying posttranslational localization to its final destination. Masaki and co-workers (1994) successfully illustrated that when cDNA was expressed in COS-1 cells, the protein exclusively localized in the well-developed ER. The authors were interested to see whether the COOH-terminal portion could direct another heterologously expressed protein (β-galactosidase) to the ER membrane. For this purpose the expression vectors pCDALDH (containing the full FALDH), pCH110 (containing β-galactosidase) and

pCH110/ALDH (containing β-galactosidase and the last 35 amino acids of the FALDH) were constructed.

Immunoblotting using mouse antibodies to *E. coli* β-galactosidase revealed that wild-type β-galactosidase (expressed from pCH110) was localized, as expected, mainly in the cytosolic fraction. Contrasting to this, the β-galactosidase/ALDH chimera was concentrated in the mitochondria/microsome fraction. In addition, the chimera in the mitochondria/microsome fraction was resistant to alkali extraction, indicating a tight anchoring of this protein to intracellular membranes (Masaki *et al*., 1994). A very similar experiment was performed by Masaki *et al*., 2003 by taking advantage of Green Fluorescent Protein (GFP) as a localization indicator.

1.5.2 *Vibrio harveyi* FALDH protein

A prime example of a well studied prokaryotic FALDH protein is the fatty aldehyde dehydrogenase (Vh-ALDH) from the bioluminescent bacterium, *Vibrio harveyi*, of which the crystal structure has been solved (Ahvazi *et al*., 2000; Zhang *et al*., 2001). Elucidation of the crystal structure of Vh-ALDH revealed interesting differences in the nucleotide binding and catalytic sites between this particular ALDH and the same sites on six other ALDHs whose structures have been determined (Zhang *et al*., 2001).

The Vh-ALDH protein has a preference for long chain aldehydes as reflected by a large decrease in the K_m for aldehydes on increasing the chain length from acetaldehyde to tatradecanal (Bognar & Meighen, 1978; Vetadi *et al*., 1996; Ahvazi *et al*., 2000). Tetradecanal is one of the substrates in the luciferase system and thus some researchers have proposed that Vh-ALDH plays a role in the luminescence of the bacterium but this has not yet been proved conclusively (Vetadi *et al*., 1996).

Although the sequence similarity of the Vh-ALDH protein with other ALDHs is low (18 – 22.5%), the *Vibrio harveyi* enzyme is in some aspects most closely related to the mammalian FALDHs (Ahvazi *et al*., 2000). Both these enzymes possess a dimeric

structure as well as higher V_{max}/K_m values for aldehydes with increasing chain length, which suggest a hydrophobic pocket at the active site.

Class 3 ALDHs are notable as the only ALDH family with a well established ability to use either NAD^+ or $NADP^+$, although given the physiological concentrations of each coenzyme, they are generally considered likely to operate only with NAD^+ *in vivo* (Perozich *et al*., 2000). Class 3 ALDHs have been shown to exhibit a relatively weak interaction with the $NADP^+$ cofactor (K_m values > 60μM). The Vh-ALDH is unique from other ALDHs due to its high affinity for $NADP^+$ (K_m 1.4μM) (Byers & Meighen, 1984; Ahvazi *et al*., 2000). This is the lowest K_m towards $NADP^+$ ever reported for any ALDH, with the previous lowest K_m towards $NADP^+$ (24.5μM) reported for the nonphosphorylating glyceraldehyde 3-phosphate dehydrogenase from *Streptococcus mutans* (Marchal & Branlant, 1999; Ahvazi *et al*., 2000). The binding of $NADP^+$ appears to arise from an interaction of the 2′ phosphate of the adenosine moiety of $NADP^+$ with a threonine (Thr-175) and an arginine (Arg-210), which forms a hydrogen bond with the negatively charged 2′-phosphate of $NADP^+$ (Ahvazi *et al*., 2000; Zhang *et al*., 2001).

The active site of Vh-ALDH contains the usual conserved amino acids namely (see Fig. 3) a cysteine (Cys-289), two glutamates (Glu-253 and Glu-377) and an asparagine (Asn-147). However, Vh-ALDH has a single polar residue in the active site that distinguishes it from other ALDHs; a histidine (His-450) is in close contact with the Cys-289 nucleophile. The close proximity of His-450 with Cys-289 could serve directly to increase the reactivity of the catalytic Cys-289 (Zhang *et al*., 2001).

1.5.3 *Acinetobacter* sp. *FALDH* and wax ester production

Under nitrogen limiting conditions, *Acinetobacter* sp. strain M-1 accumulates wax esters while utilizing *n*-alkanes as sole carbon source. Wax esters are enveloped by a single membrane and serve as a cell reserve. Various types of wax esters are widely used in the fine chemical industry to synthesize *e.g*. cosmetics, candles, printing inks, lubricants and coating stuffs. (Ishige *et al*., 2000, 2002). The starting point for wax ester synthesis is assumed to be from fatty acyl coenzyme A (acyl-CoA), which is derived from *n*-alkane metabolism (Fig. 1.6).

Singer *et al*. had already demonstrated in 1985 that *Acinetobacter* sp. strain HO1-N produced four FALDH isozymes (Ald-a to Ald-d) when cultured on different fatty alcohols and fatty aldehydes. Two distinct FALDH activities could be demonstrated in this strain: (i) a membrane-bound, NADP-dependent FALDH and (ii) a constitutive, NAD-dependent membrane-localized FALDH. The NADP-dependent isozyme activity was most strongly induced by dodecyl aldehyde (15-fold) when compared to hexadecane (9-fold) and hexadecanol (5-fold). The NAD-linked FALDH activity was constitutively expressed in this strain and an increase in activity was observed in fatty aldehyde-exposed cells, which suggests a dissimilatory role for this enzyme in fatty aldehyde metabolism. It thus appeared that *Acinetobacter* sp. strain HO1-N possesses constitutive and inducible enzyme systems for the oxidation of fatty aldehydes, since intact cells oxidized fatty aldehydes constitutively, whereas fatty aldehyde was oxidized *in vitro* by both constitutive (NAD-dependent FALDH) and inducible (NADP-dependent FALDH) enzymes. Further more it was noted that *n*-hexadecane did not induce FALDH but rather the products of its oxidation (fatty alcohol or fatty aldehyde) acted as the inducer molecule (Singer *et al*., 1985).

Later, a long-chain aldehyde dehydrogenase , Ald1, was found in the soluble fraction of *Acinetobacter* sp. strain M-1. The purified enzyme utilized only NAD^+ as cofactor. The gene coding for this FALDH was cloned and sequenced and designated as ‘ald1’. Through Northern analyses it was concluded that *n*-alkanes induced the Ald1 gene and

that *n*-hexadecane (C16) was the substrate of preference (Ishige *et al*., 2000). After gene disruption studies of *Ald1* it was ultimately concluded that the gene indeed plays a pivotal role in both the growth on *n*-alkanes and wax ester formation in *Acinetobacter* sp. strain M-1 (Ishige *et al*., 2000).

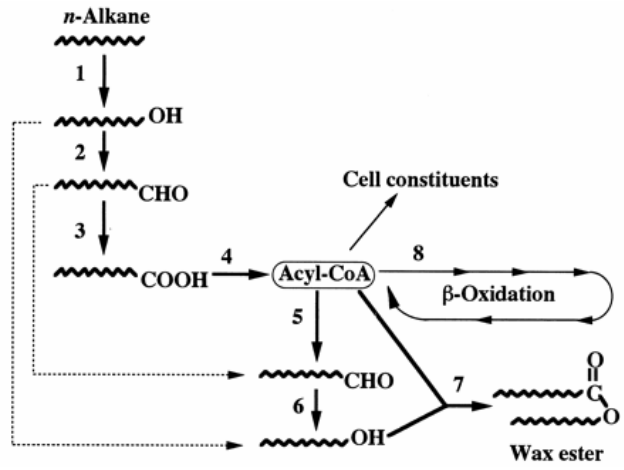


Fig. 1.6 The proposed carbon flow from *n*-alkanes to wax esters in *Acinetobacter* spp. Dotted lines indicates alternative routes in wax ester synthesis. The NAD(P)-dependent ALDH (3) is indicated (Ishige *et al*., 2002).

1.5.4 *Pseudomonas* spp. ALDHs in *n*-alkane metabolism

Pseudomonas spp. have often been isolated from petroleum-contaminated soils and are well known to utilize a variety of aliphatic hydrocarbon substrates. Prime examples of such species are *Pseudomonas butanovora*, *Pseudomonas fluorescens* and *Pseudomonas putida* (Barathi and Vasudevan, 2001; van Beilen *et al*., 2003). The ability of *Pseudomonas putida* GPo1 to utilize *n*-alkanes as a sole carbon and energy source is conferred by a catabolic OCT-plasmid which harbours two operons: *alkBFGHJKL* and *alkST*. The operons are located end to end on the OCT-plasmid and are separated by 9.7 kb of DNA (van Beilen *et al*., 2001). These operons encode the enzymes necessary to convert *n*-alkanes into fatty acids. The first cluster, the *alkBFGHJKL* operon, contains all but one of the structural genes for conversion of *n*-alkanes to the corresponding alkanoic

acids and coupling of these compounds to coenzyme A. The second cluster contains the remaining structural gene, *alkT* and the gene encoding the regulatory AlkS protein (Fig. 1.7). Expression of the first gene cluster is driven by the *alkB* promoter which is induced by the octane or dicyclopropylketone (DCPK) activated AlkS protein (Kok *et al*., 1989; van Beilen *et al*., 1992; Panke *et al*., 1999).

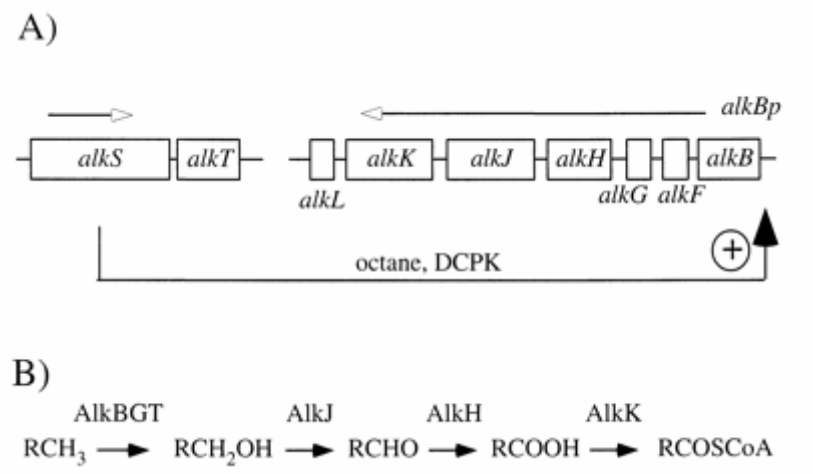


Fig. 1.7 **(A)** Organization and regulation of the *alk* genes on the OCT-plasmid in *P. putida*. The directions of transcription are indicated by arrows with open arrowheads. **(B)** Alkane degradation by enzymes encoded by *alk* genes (Adapted from: Panke *et al*., 1999).

The third largest open reading frame, corresponding to the *alkH* gene encodes a cytoplasmic, NAD-dependent ALDH that is located from nucleotide 3058 – 4506. The gene has a coding capacity for a 52.7 kDa polypeptide and the protein sequence of the transcript shows considerable homology with previously characterized ALDHs from mammalian and fungal origin (Table 1.4). The *alkH* gene was previously tentatively identified as a rubredoxin reductase but comparison of the amino acid composition of the putative translation product to that of rubredoxin reductase unambiguously showed that *alkH* could not be a rubredoxin reductase (Kok *et al*., 1989). The *alkH* gene complements, albeit weakly, a chromosomal ALDH mutation in *P. putida* which closely resembles a broad substrate specific, tumor inducible enzyme from rat liver. Although data strongly suggests that the fourth ORF of the *alkBFGHJKL* operon encodes an

ALDH, poor expression of *alkH* has precluded its unambiguous identification as such *in vitro* (Kok *et al*., 1989).

Table 1.4 Amino acid identity of confirmed eukaryotic and prokaryotic FALDHs[a]

Reference sequence (nucleotide)	Subject sequence (nucleotide)	Subject sequence length (bp)	% identity	Amino acid length	% identity	GenBank number of reference sequence
H. sapiens	*H. sapiens*	1791	100	485	100	L47162
	M. musculus	2209	5	484	83	
	R. norvegicus	2977	73	484	84	
	Acinetobacter sp.	1820	3	503	22	
	P. putida	1452	37	483	35	
	V. harveyi	1649	2	510	14	
M. musculus	*M musculus*	2209	100	484	100	AF289813
	R. norvegicus	2977	51	484	94	
	Acinetobacter sp.	1820	3	503	23	
	P. putida	1452	2	483	36	
	V. harveyi	1649	1	510	14	
R. norvegicus	*R. norvegicus*	2977	100	484	100	M73714
	Acinetobacter sp.	1820	11	503	22	
	P. putida	1452	36	483	35	
	V. harveyi	1649	3	510	14	
Acinetobacter sp.	*Acinetobacter* sp.	1820	100	503	100	AB042203
	P. putida	1452	4	483	22	
	V. harveyi	1649	2	510	13	
P. putida	*P. putida*	1452	100	483	100	AJ245436
	V. harveyi	1649	1	510	12	
V. harveyi	*V. harveyi*	1649	100	510	100	U39638

[a]All alignments were performed with ClustalW.

1.6 Comparison and overview on the confirmed FALDHs

It is evident from Table 1.4 that the bacterial FALDHs from *V. harveyi* and the *Acinetobacter* sp. have very low amino acid identity with the mammalian FALDHs. This is also evident from the phylogetic tree (Fig. 1.8) showing the relationship between confirmed FALDHs and other ALDHs. From this tree it is evident that the FALDH from *V. harveyi* is more closely related to the glyceraldehyde-3-phosphate dehydrogenases than to class 3 ALDHs, while the FALDH from the *Acinetobacter* sp. possibly belongs in the Class 1/2 trunk of the ALDH tree (Fig. 1.1). The FALDH from *P. putida* on the other hand is more closely related to the mammalian FALDHs and is a class 3 enzyme. This comparison shows that amino acid identity analysis does not always allow prediction of substrate specificity.

Important characteristics of the different FALDHs are given in Table 1.5. Clearly, the most research has been done on eukaryotic FALDHs on a genomic as well proteonomic level. On the prokaryotic front, it seems that *Vibrio harveyi* has enjoyed much scientific scrutiny – especially on protein level. Other prokaryotic FALDHs have been studied on either genomic or protein level but never both. Unfortunately, no comprehensive research has been done on FALDH from yeasts, although FALDH activity has been reported in *C. maltosa* and *Y. lipolytica* (Barth & Gaillardin, 1996; Fickers *et al*., 2005).

1.7 Four putative *FALDH* encoding genes in the genome of *Y. lipolytica*

Four putative *FALDH* encoding genes, labeled *FALDH1, FALDH2, FALDH3* and *FALDH4*, have been identified in the recently sequenced genome of *Y. lipolytica* (Fickers *et al*., 2005; Matatiele, 2005). The phylogenetic tree in Figure 1.8 illustrates that these four putative *FALDH*s from *Y. lipolytica* are more closely related to Class 3 ALDHs than they are to other members of the Class 3 trunk (Fig. 1.1). Northern blot analysis using unique fragments of the *FALDH3* and *FALDH4* genes as probes had indicated that *FALDH4* was induced during growth on *n*-alkanes but not during growth on glucose or glycerol (Matatiele, 2005). However, when quadruple *FALDH* deletion mutants were evaluated for growth on *n*-alkanes, there was no significant difference between the wild-

type strain with intact *FALDH*s and the strain with all four *FALDH* genes disrupted (Matatiele, 2005). From the latter observation one could conclude that the quadruple deletions do not have an effect on *n*-alkane assimilation and that other enzymes may play a role in the pathway. This may very well be the case, since at present, 28 protein sequences from *Y. lipolytica* have been designated as ALDHs in the NCBI database and may work in conjunction with these four putative *FALDH*s in the metabolism of *n*-alkane intermediates. Alternatively the cytochrome P450 monooxygenases that hydroxylate the alkanes might overoxidise the initial alcohols to the corresponding fatty acids (Scheller *et al*., 1998). Even if functional FALDHs are not essential for growth of *Y. lipolytica* on *n*-alkanes, the questions remain as to why there would be four of these genes present, what would be the conditions under which these genes are induced, what are the natural substrates for these enzymes and where are these proteins localized?

Table 1.5 Properties of various confirmed FALDHs

Description and references	***H. sapiens***	***M. musculus***	***R. norvegicus***	***V. harveyi***	***Acinetobacter sp.* Strain M-1**	***P. putida***
Cloned	Yes	Yes	Yes	Yes	Yes	Yes
Coding sequence (amino acid)	485	484	484	510	503	483
Promoter characteristics	Lacks TATA box. Has CpG islands.	Lacks TATA box. Has CpG islands. Contains 13 TG repeats	Lacks functional TATA box. Is expressed differentially.	*	*	Drives first cluster of an operon. Induced by octane or DCPK.
Hydrophobic carboxyl-terminal tail	Yes	Yes	Yes	*	*	*
Sub-cellular localization	Microsomes	Microsomes	Microsomes	*	Membrane-bound	Cytoplasm

Table 1.5 continued…

Description and references	***H. sapiens***	***M. musculus***	***R. norvegicus***	***V. harveyi***	***Acinetobacter* Strain M-1**	***P. putida***
Quaternary structure	Homodimer	Homodimer	Homodimer	Homodimer	Homotetramer	*
Binds both NAD^+/$NADP^+$	Yes	Yes	Yes	Yes	Yes	*
Preferred cofactor	NAD^+	NAD^+	NAD^+	$NADP^+$	NAD^+	*
K_m value for cofactor (μM)	280	*	11.3	1.4	*	*
References	Kelson *et al*., 1997 Yosihida *et al*., 1998 Rizzo *et al*., 2001	Lin *et al*., 2000 Rizzo *et al*., 2001	Miyauchi *et al*., 1991 Xie *et al*., 1996 Kelson *et al*., 1997 Liu *et al*., 1997 Perozich *et al*., 2000	Vedadi *et al*., 1995 Ahvazi *et al*., 2000	Ishige *et al*., 2000	Kok *et al*., 1989 van Beilen *et al*., 2001

[a] Singer *et al*., 1985 did report localization of two distinct FALDHs in *Acinetobacter* sp. strain HO1-N: NAD^+-linked FALDH activity was primarily in the soluble fraction while the $NADP^+$-linked FALDH activity was membrane bound.

* Before final submission of this dissertation, no definitive information could be obtained to unequivocally confirm the omitted information.

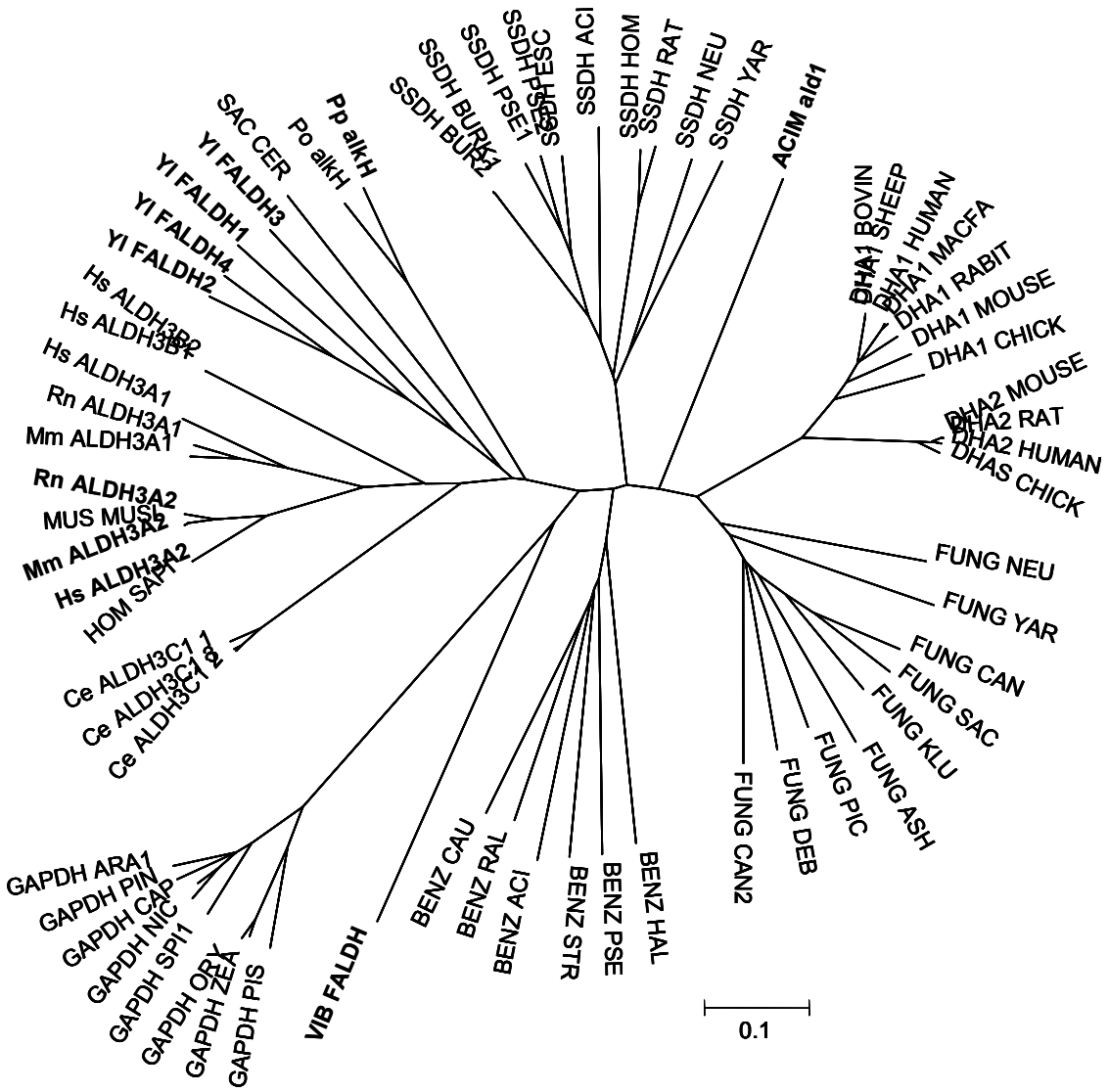


Fig. 1.8 Phylogenetic relationships of the four putative FALDH protein sequences in relation to other ALDHs. The sequences were aligned in ClustalX and the Neighbour-Joining (N-J) tree constructed with TreeExplorer by using default settings. Confirmed *FALDH*s as well as the *Y. lipolytica FALDH* isozymes are in bold. Abbreviations for ALDHs: BENZ = Benzaldehyde dehydrogenase; FUNG = Fungal dehydrogenase; GAPDH = Glyceraldehyde-3-phosphate dehydrogenase and SSDH = Succinic semialdehyde dehydrogenase (also see Table 1.6) (Adapted from: Matatiele, 2005, Ph.D. thesis).

Table 1.6 Abbreviations used on the phylogenic tree with corresponding accession numbers

Tree abbreviations	Accession number	Organism
BENZ_ACI	Q6FCB6	*Acinetobacter* sp. (ADP1)
BENZ_CAU	Q9A5Q0	*Caulobacter crescentus*
BENZ_HAL	Q9HMJ6	*Halobacterium* sp. (NRC-1)
BENZ_PSE	P43503	*Pseudomonas putida*
BENZ_RAL	Q8XT86	*Ralstonia solanacearum*
BENZ_STR	Q9L124	*Streptomyces coelicolor*
BENZ_XAN	Q8PD14	*Xanthomonas campestris (*pv. *campestris)*
Ce_ALDH3C1_1	Q60WY8	*Caenorhabditis briggsae*
Ce_ALDH3C1_2	O16518	*Caenorhabditis elegans*
Ce_ALDH3C1_3	Q86S57	*Caenorhabditis elegans*
Ce_ALDH3C2	Q8MXJ7	*Caenorhabditis elegans*
DHA1_BOVIN	P48644	*Bos taurus*
DHA1_CHICK	P27463	*Gallus gallus*
DHA1_HUMAN	P00352	*Homo sapiens*
DHA1_MACFA	Q8HYE4	*Macaca fascicularis*
DHA1_MOUSE	P24549	*Mus musculus*
DHA1_RABIT	Q8MI17	*Oryctolagus cuniculus*
DHA1_SHEEP	P51977	*Ovis aries*
DHA2_HUMAN	O94788	*Homo sapiens*
DHA2_MOUSE	Q62148	*Mus musculus*
DHA2_RAT	Q63639	*Rattus norvegicus*
DHAS_CHICK	O93344	*Gallus gallus*
Dr_ALDH3D1_1	NP_775328.2	*Danio rerio*
FUNG_ASH	Q758W1	*Ashbya gossypii*
FUNG_CAN	Q6FVP8	*Candida glabrata* CBS138

Tree abbreviations	Accession number	Organism
FUNG_CAN2	Q6FPK0	*Candida glabrata* CBS138
FUNG_DEB	Q6BJB3	*Debaromyces hansenii* CBS767
FUNG_KLU	Q6CLU0	*Kluyveromyces lactis* NRRL Y-1140
FUNG_NEU	Q8X0L4	*Neurospora crassa*
FUNG_PIC	Q12648	*Pichia angusta*
FUNG_SAC	P46367	*Saccharomyces cerevisiae*
FUNG_YAR	Q6CD79	*Yarrowia lipolytica* CLIB99
GAPDH_ARA1	P25857	*Arabidopsis thaliana*
GAPDH_CAP	QVWP2	*Capsicum annuum*
GAPDH_NIC	P09044	*Nicotiana tabacum*
GAPDH_ORY	Q7X8A1	*Oryza sativa (*japonica cultivar-group*)*
GAPDH_PIN	P12859	*Pisum sativum*
GAPDH_PIS	Q41019	*Pinus sylvestris*
GAPDH_SPI1	P12860	*Spinacia oleracea*
GAPDH_ZEA	Q6LBU9	*Zea mays*
HOM-SAPL	P51648	*Homo sapiens*
Hs_ALDH3A1	Q6PKA6	*Homo sapiens*
Hs_ALDH3A2	P51648	*Homo sapiens*
Hs_ALDH3B1	Q8N515	*Homo sapiens*
Hs_ALDH3B2	P43353	*Homo sapiens*
Mm_ALDH3A1	P47739	*Mus musculus*
Mm_ALDH3A2	P47740	*Mus musculus*
Mm_ALDH3B1	X8VHW0	*Mus musculus*
Mm_ALDH3A2	Q99L64	*Mus musculus*
MUS_MUSL	AAK01551.1	*Mus musculus*
Po_alkH	P12693	*Pseudomonas oleovarans*
Pp alkH	CAB51050.1	*Pseudomonas putida*
Rn_ALDH3A1	P11883	*Rattus norvegicus*

Tree abbreviations	Accession number	Organism
Rn_ALDH3A2	P30839	*Rattus norvegicus*
SSDH_ACI	Q6F9G0	*Acinetobacter sp.* ADP1
SSDH_BUR2	Q62B48	*Burkholderia mallei* ATCC 23344
SSDH_BURK1	Q63NL9	*Burkholderia pseudomallei*
SSDH_ESC	Q8X950	*Escherichia coli* O157:H7
SSDH_HOM	Q8N3W6	*Homo sapiens*
SSDH_NEU	Q7SFB1	*Neurospora crassa*
SSDH_PSE1	Q916M5	*Pseudomonas aeruginosa*
SSDH_PSE2	Q88AT6	*Pseudomonas syringae* (pv. tomato)
SSDH_RAT	P51650	*Rattus norvegicus*
SSDH_YAR	Q6C0B4	*Yarrowia lipolytica* CLIB99
SAC_CER	NP_013828.1	*Saccharomyces cerevisiae*
YlFALDH1	Q6CG32	*Yarrowia lipolytica* CLIB99
YlFALDH2	Q6C0L0	*Yarrowia lipolytica* CLIB99
YlFALDH3	Q6CGN3	*Yarrowia lipolytica* CLIB99
YlFALDH4	Q6C5T1	*Yarrowia lipolytica* CLIB99

1.8 Concluding remarks

ALDH enzymes are widely distributed throughout the evolutionary scale. This seems logical considering the detoxification role ALDHs play in living systems. ALDHs are present as isozymes and this feature seems to be prevalent in prokaryotes. This is justified due to the diverse environments and wide substrate specificities of these organisms. Non-specific ALDHs are grouped into three classes based on their substrate specificity, subcellular distribution and primary sequence similarity.

Class 1 and 2 ALDHs have enjoyed extensive scientific scrutiny while Class 3 ALDHs conversely has not. A trademark of Class 1 and 2 ALDHs is their dependence on $NADP^+$ while Class 3 ALDHs, in contrast, function with either NAD^+ or $NADP^+$. The mammalian FALDHs are part of Class 3 and are referred to as ALDH3A2. These enzymes preferentially oxidize long-chain aldehydes to their corresponding carboxylic acids. The mammalian FALDHs have been extensively studied on a genomic as well as protein level. The only prokaryotic FALDHs which group with the Class 3 ALDHs are the FALDHs from *Pseudomonas putida* and *Pseudomonas oleovorans*. These enzymes have not been studied in depth. The FALDH from *V. harveyi,* on the other hand, has enjoyed much attention. This enzyme, of which the three dimensional structure has been determined, is more closely related to the glyceraldehyde-3-phosphate dehydrogenases than to the mammalian FALDHs. A fourth type of FALDH which has also not been studied in depth is an FALDH from an *Acinetobacter* sp. The deduced amino acid sequence of the FALDH is more similar to that of Class 1 and 2 ALDHs than to Class 3 ALDHs.

No extensive research has been undertaken with regard to FALDHs in *n*-alkane assimilating yeasts. This is truly surprising since 20% of nearly all 500 yeast species are capable of hydrocarbon assimilation. Yeast genera that have been synonymous with hydrocarbon degradation are *Pichia*, *Candida* and *Yarrowia*. Four putative FALDH encoding genes have been identified in the recently sequenced genome of *Y. lipolytica*. It must still be established whether these genes code for functional FALDHs.

Chapter 2
Introduction to present study

2.1 A condensed background of *Yarrowia lipolytica*

Yarrowia lipolytica first presented itself to the scientific world when it was isolated in 1945 from a jar of fiber tailings in a corn processing plant. The species was first classified as a *Candida*, since no sexual state was described. However, the perfect form of *Candida lipolytica* was identified in the late 1960s by Wickerham at the Northern Regional Research Laboratory of the USDA at Peoria (Barth and Gaillardin, 1996). *Y. lipolytica* is an obligate respiratory organism and catabolism of all carbon sources transits through the tricarboxylic acid cycle. This differs markedly from fermentative yeasts such as *Saccharomyces cerevisiae*, where catabolism of glucose and other sugars proceeds mainly by fermentation (Flores & Gancedo, 2005). In addition, this yeast does not ferment sugars and is unable to assimilate nitrate (Kurtzman & Fell, 1998; Casaregola *et al*., 2000).

This ascomycetous yeast is the only species in its genus and was found to be genetically distant from well known members of *Candida* (*i.e*. *Candida maltosa* and *Candida tropicalis*) and several teleomorphic ascomycetous yeast genera (*i.e*. *Debaryomyces* and *Pichia*) based on the following properties: (i) a high GC content (49.6 - 51.7%); (ii) unusual structure of rDNA genes; (iii) a lack of RNA polymerase I consensus sequences present in other yeasts; (iv) higher eukaryotic-like size of small nuclear RNA and of 7S RNA (Barth & Gaillardin, 1997). This genetic divergence is a positive aspect, since the observation of structural conservation between *Y. lipolytica* genes and other yeast genes is much more likely to reflect functional constraints than when it happens between closely related yeasts (Barth & Gaillardin, 1997). Taxonomists Kurtzman and Fell, (1998) have assigned *Y. lipolytica* to the family *Dipodascacae,* while Kurtzman and Robnett (1998) placed it, together with *Arxula* and some *Candida* species, in the *Metschinikowia*/*Stephanoascus* clade based on 26S-rDNA analyses (previous two

references cited from Fickers *et al*., 2005). *Y. lipolytica* is dimorphic, heterothallic and is not considered pathogenic, unlike other *n*-alkane degrading counterparts *e.g*. *Candida albicans* and *Candida tropicalis*. This has led the American Food and Drug Administration (FDA) to classify *Y. lipolytica* as Generally Regarded as Safe (GRAS) for citric acid production (Fickers *et al*., 2005).

Strains are often isolated from hydrophobic substrates such as cheese, sausage, margarine, olive oil as well as sewage and oil plants (Barth & Gaillardin 1996; Casaregola *et al*., 2000; Juretzek *et al*., 2001; Madzak *et al*., 2004). It is thus no surprise that this yeast is capable of utilizing fatty acids and *n*-alkanes as a carbon source. *Y. lipolytica* very efficiently utilizes long- and short-chain triglycerides and their corresponding *n*-alkanes ranging from decane (C_{10}) to octadecane (C_{18}) and even longer chain *n*-alkanes are utilized (Mauersberger *et al*., 2001).

Y. lipolytica is one of the most extensively studied non-conventional yeasts and in 2003 the complete genome sequence of *Y. lipolytica* E150 (CLIB99) was released by the Génolevures Consortium (Dujon *et al*., 2004; Fickers *et al*., 2005). It initially aroused industrial interest due to its capacity for high molecular weight protein secretion, production of high amounts of organic acids and its ability grow on *n*-paraffins (Juretzek *et al*., 2001; Fickers *et al*., 2003; Madzak *et al*., 2004). The latter property accentuated *Y. lipolytica* in the 1960's as a source of yeast protein from hydrocarbons - a popular idea in an era of low cost petroleum. Production of the industrially important acidulant: citric acid by *Y. lipolytica*, signaled a dramatic change in the fermentation industry, since the increased yields were *ca*. two fold compared to that of *Aspergillus niger*, the industry mainstay since the early 1900's (Kurtzman & Fell, 1998). *Y. lipolytica* is currently also a workhorse in the biotransformation of hydroxy fatty acids into lactones. One of these lactones, γ-decalactone, which has a distinct peach flavour, is obtained by the bioconversion of alkyl ricinoleate (a castor oil derivative) and a process using *Y. lipolytica* has been patented by BASF (DE4126997, 1993) (Barth & Gaillardin, 1997; Nicaud *et al*., 1998).

2.2 *Y. lipolytica* is genetically accessible

Yeast genetics and *Saccharomyces cerevisiae* have become scientific household names and the same holds true for the bacterial counterpart *Escherichia coli*. This is due to the development of efficient molecular tools for these two organisms, making them genetically accessible and exploitable. The potential of *Y. lipolytica* has long been realized but is only quite recently that its power could be harnessed and focused on specific outcomes. Nearly all natural isolates of *Y. lipolytica* have haploid or near haploid genomes and a sexual life cycle. These properties make *Y. lipolytica* much more advantageous for genetic and molecular analyses of genes involved in *n*-alkane assimilation than the anamorphic *Candida* yeasts (Barth & Gaillardin, 1997; Iida *et al*., 1998).

2.2.1 Genetic tools for *Y. lipolytica*

To prevent this chapter from becoming too long only the following genetic tools used in *Y. lipolytica* will be discussed: (a) integrative vectors, (b) replicative vectors, (c) expression and secretion vectors, (d) the long terminal repeats and *zeta* elements and (e) the Cre *lox*P/R recombinase system.

(a) Integrative vectors

Integrative vectors require a homologous region within chromosomal DNA. High transformation efficiency is achieved in *Y. lipolytica* when using the lithium acetate procedure and a lower level when using electroporation. (Barth & Gaillardin, 1996). Integration occurs almost exclusively by homologous recombination and integration is greatly stimulated by linearizing the vector in the homology region. It is recommended that homology should extend a few hundred base pairs on both sides of the linearized fragment. A combination of the linearized vector and the flanking regions result in more than 80% integration of a single copy at the chosen site (Barth & Gaillardin, 1996; Madzak *et al*., 2004). For maintenance in yeast, the autonomously replicating sequences

(ARS) are also included in the vector design. There are however, a few minor drawbacks regarding ARS. Firstly, ARS-based vectors are limited to 1-3 copies/cell which is also correlated with related gene expression. Secondly, ARS-based vectors require that a selective pressure be maintained which may pose industrial management hurdles (Madzak *et al*., 2004).

Multicopy vectors integrate into the recipient genome either by homology with multiple target sites or non-homology and the amplification is selected for using a defective selection marker (Le Dall *et al*., 1994; Madzak *et al*., 2004). Three targets of integration have been tested: tandemly repeated sequences of rDNA, single-copy genes like *XPR2* and dispersed repeated sequences like the Ylt1 transposon. Stability of the integrated plasmids was high, except when *XPR2* transcription was turned on in high copy number transformants (Barth & Gaillardin, 1996).

(b) Replicative vectors

Replicative vectors carry both an origin of replication (ORI) and a centromeric region and are thus relatively stable under nonselective conditions. Transformation can be achieved by the lithium acetate procedure or by electroporation. Both procedures yield 10^3-10^4 transformants/µg DNA. Linearization of the vector in conjunction with the lithium acetate method elevates the transformation frequency but in contrast to electroporation, the transformation is nearly abolished. For a more comprehensive review see Barth & Gaillardin, 1996.

(c) Expression and secretion vectors

These vectors are probably the most cardinal of all since they bridge the gap between the laboratory and the industry. All expression vectors are shuttle vectors (hybrids between yeast-derived and bacterial sequences), which typically contain (i) a selection marker, (ii) an expression cassette and (iii) elements for maintenance in yeast cells. Table 2.1

illustrates the most important components of the expression/secretion vectors for *Y. lipolytica*.

Table 2.1 Components of the expression/secretion vectors for *Y. lipolytica*

Component	Characteristics
Marker genes	
LEU, *URA3*, *LYS5*, *ADE1* (*Y. lipolytica*)	Auxotrophy complementation
ura3d4 (*Y. lipolytica*)	Defective marker yielding multi-copy integrants
Phleo^{R}, *hph* (*E. coli*)	Phleomycin and hygromycin B resistance
SUC2 (*S. cerevisiae*)	Sucrose utilization
Promoters (source)	
p*LEU2* (β-isopropylmalate dehydrogenase)	Inducible by leucine precursor
p*XPR2* (alkaline extracellular protease)	Inducible by peptones
p*POX2*, *pPOT1* (acyl-CoA oxidase 2 and 3-oxo-acyl-CoA thiolase respectively)	Inducible by fatty acids and derivatives, and *n*-alkanes
p*ICL1* (isocitrate lyase)	Inducible by fatty acids and derivatives, *n*-alkanes, ethanol and acetate
p*POX1*, p*POX5* (acyl-CoA oxidases 1 and 5)	Weakly inducible by *n*-alkanes
p*G3P* (glycerol-3-phosphate dehydrogenase)	Inducible by glycerol
p*MTP* (bidirectional: metallothioneins 1 and 2)	Inducible by metallic salts
hp4d (hybrid promoter derived from p*XPR*2)	Growth-phase dependent (quassi-constitutive)
p*TEF*, p*RPS7* (translation elongation factor-1α, ribosomal protein S7 respectively)	Constitutive
Secretion signals	
Native	Frequently efficient in *Y. lipolytica*
XPR2 prepro	13 aa pre/10 XA or XP dipeptides/122 aa pro/KR cleavage site
XPR2 pre + dipeptide	13 aa pre/5-10 XA or XP dipeptides
XPR2 pre	13 aa pre/LA cleavage site
LIP2 prepro	13 aa pre/4 XA or XP dipeptides/10 a pro/KR cleavage site

Table 2.1 continued…

Components	Characteristics
Terminators	
*XPR*2t, *LIP*2t,l *PHO5*t	430, 150 and 320 base pair fragments respectively
Minimal *XPR*2t	PCR-synthesized 100 base pair fragment with added restriction sites
Elements for maintenance in yeast cells	
ARS18, ARS68	Autonomously replicative vectors can maintain only 1-3 copies/cell
Homology to genome	Homologous integration (in *LEU2*, *URA3*, *XPR2* terminator, rDNA, or, when present, in zeta or pBR322 docking platform)
Zeta (Ylt1 LTR)	Non-homologous integration in Ylt1-devoid strains

Abbreviations: aa, amino acids; X, any amino acid; A, alanine; P, proline; K, lysine; R, arginine; L, leucine
(Adapted from: Madzak *et al*., 2004)

(d) Transposable elements

Transposable elements are ubiquitous components of eukaryotic genomes and they are designated into two types: retrotransposons with identical sequences at both ends (Long Terminal Repeats or LTR) and retrotransposons lacking LTR. In addition, LTR retrotransposons are grouped into two different classes: the Ty1/*copia* and the Ty3/*gypsy* classes (Senam & Barth, 2003). Both types of retrotransposons have been detected in strains of *Y. lipolytica* (Senam & Barth, 2003). Ty1 elements carry strong promoters for RNA polymerase and constitute 5 – 10% of total yeast mRNA (Watson *et al*., 1996).

The first retrotransposon described in *Y. lipolytica* was Ylt1 which belongs under the high-copy number Ty3-*gypsy* fungal retrotransposon group. Ylt1 is not present in all wild-type strains of *Y. lipolytica* and it is thought to have originated from the wild American isolate YB423 (Madzak *et al*., 2004). The Ylt1 retrotransposon consists of an internal region termed as *eta* and is flanked by two LTRs termed as *zeta*, which consists of 714bp (Schmid-Berger *et al*., 1994). The complete retrotransposon is 9.4kb in size. The sequence of the *zeta* element is well conserved and is bound by nucleotides 5′-TG … CA-3′, which is part of a short inverted repeat – a feature conserved in the LTR's of retroviruses and retrotransposons (Schmid-Berger *et al*., 1994). The Ylt1 contains two overlapping large open reading frames (*YltA* and *YltB*) and they display similarities to retroviral elements (Barth & Gaillardin, 1997; Madzak *et al*., 2004).

Strains harboring the Ylt1 have *ca.* 35 copies per genome in a dispersed fashion. In addition, zeta LTR's were also additionally found as solo elements, with at least 30 – 60 copies per genome depending on the strain (Madzak *et al*., 2004). This of course implies that there are at least 100 transposable targets in the genome of a Ylt1-carrying strain. Such repetitive elements are often used as target sites for integration of plasmids carrying genes to be amplified. In *Y. lipolytica* the rDNA locus is also commonly used (Pignède *et a*l., 2000).

Zeta elements have quite recently been utilized as a target site for integration by homologous recombination in the genome of *Y. lipolytica*. Pignéde *et al*., 2000 adapted their single and multicopy vectors to use *zeta* elements (from the *Y. lipolytica* retrotransposon Ylt1) as targeting sites and to obtain integrative cassettes free of bacterial DNA. Characteristic of these cassettes were that they were *Not*I-*Not*I cassettes containing a defective *URA3* allele, a polylinker sequence and the *zeta* elements for targeting multiple sites in the genome of the recipient. These plasmids were modified by Juretzek *et al*., 2001 by including the inducible *ICL1* or *XPR2* promoters for gene expression. They observed a strong increase in the plasmid copy numbers after rDNA or *zeta*-directed integration was employed. The authors concluded that multiple integration of expression vectors into the *zeta* and rDNA loci represent useful tools to increase the copy number of expression cassettes and that these new vectors are suitable for regulated expression of homologous and heterologous genes in *Y. lipolytica*.

Transposon mutagenesis is a valuable genetic tool for research in especially bacterial genetics and since the development of "shuttle mutagenesis"; a wide variety of species can be targeted for transposition. This technology has also been tailored for yeasts, enabling analyses of protein production, localization and function. A multipurpose transposon system has been developed in which a mini-transposon (mTn), containing an expression reporter construct, is able to transpose at multiple independent sites in a cloned gene.

Such a multipurpose transposon system has been employed in *S. cerevisiae* where the transposon is fused to a coding region for β-galactosidase or green fluorescent protein (GFP) from *Aequorea victoria* (Ross-Macdonald *et al*., 1997). Gene expression can therefore be monitored by chemical or fluorescence assays. A tagging-mutagenesis system for *Y. lipolytica* was developed by modifying the mTn-3xHA/GFP transposon of Ross-Macdonald *et al*., 1997. The aim of this system was to identify genes that are non-essential for growth on glucose *e.g.* genes involved in dimorphism, biodegradation of alkanes, oil, fatty acids or peroxisome biogenesis. Other *Y. lipolytica* studies conducted in conjunction with transposons include *e.g.* the tagging of morphogenetic genes by

mutating wild-type strains with a Tn3 derivative and determining the cell wall composition and structure of *Y. lipolytica* transposon mutants (Richard *et al*., 2001; Ruiz-Herrera *et al*., 2003). In an attempt to tag mutations in genes involved in hydrophobic substrate utilization, Mauersberger *et al*., 2001 generated insertions of a *zeta*-*URA3* mutagenesis cassette (MTC) into the genome of *zeta*-free and *ura3* deletion-containing strains of *Y. lipolytica*. The conclusion reached from the study, was that MTC integration is a powerful tool for generating and analyzing tagged mutants in *Y. lipolytica*.

Zeta-technology has its drawbacks. The most prominent being that the transformation of a Ylt1-free strain with a *zeta*-based vector is much lower than with a classical integrative vector. In addition, the use of a defective marker decreases the transformation frequency from 10^6 transformants/µg DNA to *ca*. 10^2 transformants/µg DNA when using a *zeta*-based vector as to a classical integrative vector. Despite this lower transformation efficiency, non-homologous transformation of Ylt1-free *Y. lipolytica* strains with zeta-based vectors yield integrated copies that are more dispersed in the genome which leads to better stability of high-copy-number integrants (Madzak *et al*., 2004).

Quite recently Neuvéglise *et al*., 2005 identified a new type of DNA transposon, *Mutyl*, in the sequenced genome of *Y. lipolytica*. For more comprehensive details please refer to Neuvéglise *et al*., 2005.

(e) Using the *Cre-lox* recombination system in *Y. lipolytica* to study gene function

The *Y. lipolytica*-Cre-*lox* system (Fickers *et al*., 2003) is a fusion between the SEP (Sticky-End Polymerase Chain Reaction) method (Maftahi *et al*. 1996) and the *Cre-lox* recombination system (Sauer, 1987). Phage P1 encodes an efficient site-directed recombination system consisting of a short asymmetric DNA sequence called *loxP* and a 38 kDa protein called Cre. The *loxP* site is a 34bp sequence composed of two 13bp inverted repeats separated by an asymmetric 8bp core sequence (GCATACAT). The Cre-protein enters the nucleus and recognizes chromosomal *lox*-sites, which flank a certain DNA sequence. After recognition Cre performs synapsis and recombination events.

(Sauer, 1987; Nagy, 2000). Sauer (1987) illustrated that the prokaryotic recombinase can enter an eukaryotic nucleus and, moreover, that the Cre recombinase could perform precise recombination events on the chromosomes of *S. cerevisiae*. The author fused the *cre* gene to the *GAL1* promoter of *S. cerevisiae* as well as the *LEU2* gene, flanked by *lox* sites. The *LEU2* gene was integrated in both orientations into two different chromosomes. Excisive recombination at the *lox* sites was accurately achieved (as measured by loss of the *LEU2* gene) by the expressed Cre protein that was dependent on galactose for its induction (Sauer, 1987).The SEP method was employed for systematic gene disruption and was used to disrupt three open reading frames identified during the sequencing of COS14-1 from chromosome XIV of *S. cerevisiae* (Maftahi *et al*., 1996). Table 2.2 provides a summary of a typical disruption procedure.

In our study the SEP method was used to construct Promoter-*lac*Z-Terminator cassettes (PLT-cassettes) to study expression of the four putative *FALDH*s of *Y. lipolytica*. *FALDH* deletion mutants (used in our study) were constructed in our laboratory by Dr. P.R. Matatiele with the aid of the Cre-*lox* recombination system (Fickers *et al*., 2003). Two of the PLT-cassettes from the SEP experiment were provided to Dr. Matatiele for completion of the *FALDH* deletion mutants. In exchange, we were provided with a set of triple and quadruple *FALDH* deletion mutants.

Table 2.2 Summary of the disruption procedure when using SEP

Step	Procedure
1.	PCR1: separately amplify the promoter (P) and terminator (T) regions that contain rare restriction enzyme sites (*e.g*. *Asc*I or I-*Sce* I)
2.	PCR2: Pool P and T (to act as template for the polymerase), anneal and amplify to yield a PT-cassette. Purify the PT-cassette and polish with T4 DNA polymerase
3.	Vector preparation: cut the required cloning vector with a blunt-end enzyme
4.	Disrupting vector construction: ligate the PT-cassette (from step 2) into the cloning vector (from step 3) in the presence of the blunt-end restriction enzyme and transform *E. coli*
5.	Prepare the disrupting vector DNA *i.e.* small scale plasmid isolation
6.	Linearize the disrupting vector with *Asc*I or I-*Sce* I
7.	Transform yeast (with *e.g*. the lithium acetate procedure)

(Adapted from: Maftahi *et al*., 1996).

In a study by Luo *et al*., (2000, 2002) the SEP method was adapted to for gene studies in *Y. lipolytica*. The authors were exploring the role and function of acyl-CoA oxidase (Aox) in the degradation of fatty acids and in the production of lactones by bioconversion. By constructing mono-disrupted strains, they concluded that of the five Aox isozymes, *AOX2* codes for an Aox which is more active towards long chain fatty acids like Meristyl-CoA (C_{14}), while *AOX3* encodes a Aox which displays a higher affinity for short chain fatty acids *e.g.* Hexanoyl-CoA (C_6). Strains disrupted for *AOX1*, *AOX4* and *AOX5* showed higher activity of Aox on every fatty acid than the wildtype (Luo *et al*., 2000, 2002).

In *Y. lipolytica* few methods for gene disruption have been reported and current methods are time consuming and laborious. The latter problems were circumvented by Fickers *et al*., 2003 by employing versatile homologous auxotrophic markers like yl*URA3*, yl*LEU2* and the heterologous antibiotic resistance marker *hph* (coding for hygromycin-B resistance). The chief advantages of the *hph* marker, was that gene disruption by the

cassette permitted integration at the correct locus (a 45% frequency on average) due to its lack of homology with the yeast genes. Secondly, direct selection of Hyg^R (hygromycin resistant) transformants was very efficient and transformation frequencies similar to that obtained with classic auxotrophic markers were obtained. Lastly, it was possible to introduce gene disruptions in both wild-type and industrial strains, which often lack the standard auxotrophic markers present in laboratory counterparts. The Cre-*lox* technology was applied in conjunction with the SEP method to generate rapid gene disruptions.

Since there are a limited number of marker genes in *Y. lipolytica*, this procedure presented an efficient way to rescue and reuse the marker. Upon expression of the Cre-recombinase, the marker was excised at a frequency of 98%, by recombination between the two *lox* sites. Consequently, this technology provides a powerful tool for the functional analysis of gene families or creating large mutant collections (Fickers *et al*., 2003).

2.2.2 Enzymes involved in *n*-alkane assimilation

Since *n*-alkanes are lipophilic organic compounds, it is essential that either the substrates or the cells should be modified in such a way as to facilitate entry of the substrate into the cell. The first step in *n*-alkane assimilation is likely to be an emulsification at the surface to form micelles or microscopic oil droplets which can be internalized (Barth & Gaillardin, 1996). *Y. lipolytica* has been known to produce a 27 kDa extracellular, *n*-alkane inducible emulsifier, called liposan, which consists of 88% carbohydrate and 12% protein (Kim *et al*., 1999). Besides the production of biosurfactants, cell hydrophobicity, is considered to be one of the most cardinal factors controlling hydrocarbon assimilation since highly hydrophobic cells have a greater chance to adhere to oil droplets *versus* cells with lower hydrophobicity (Barth & Gaillardin., 1996; Kim *et al*., 1999).

A *Y. lipolytica* 180 strain exhibited interesting cell surface characteristics when cultured on hydrocarbons. Electron microscopy studies revealed protrusions on the cell surface of cells cultured on crude oil as well as a thicker periplasmic space and cell wall *versus* cells

cultured on glucose (Kim *et al*., 2000). During growth on hydrocarbons, two types of emulsifiers were produced; one was extracellular and hydrophilic, whilst the other was cell wall-associated. At low concentrations (0.12%) both emulsifying materials enhanced the oil-degrading activity of *Moraxella* sp. K12-7.

After entry into the cell, *n*-alkanes are hydroxylated by a cytochrome P-450 monooxgenase system (Fickers *et al*., 2005). Alkanes are usually activated by terminal oxidation to the corresponding primary alcohol and once the initial oxidation of the alkane has been accomplished, the hydroxylfunctional group then becomes the focal point for subsequent oxidation. The alcohol is thus converted to the corresponding aldehyde and then to the fatty acid. It has long been assumed that in *Y. lipolytica* the latter reactions are catalyzed by long-chain fatty alcohol oxidase (LC-FAO) and fatty aldehyde dehydrogenase (FALDH) respectively, since the corresponding enzyme activities have been detected in cell free extracts of *n*-alkane grown cells (Ratledge, 1984; Barth & Gaillardin., 1996; van Beilen *et al*., 2003). Table 2.3 provides a summary of characteristics of the enzymes catalyzing the first two steps of *n*-alkane assimilation.

However, BLAST searches of the *Y. lipolytica* genome sequence database using the deduced amino acid sequence from the FAO gene of *C. maltosa* as query did not deliver any putative FAO encoding genes (Matatiele, 2005). However, when the amino acid sequence of human FALDH (Rizzo *et al*., 2001) was used as the blast query, four putative *FALDH* genes named *FALDH 1* to *FALDH4* were located. The deduced proteins displayed 34 - 39% amino acid identity with the confirmed human FALDH. Deletion of these four genes with the Cre-*lox* system had no effect on *n*-alkane utilization (Matatiele, 2005).

Table 2.3 Characteristics of monooxygenase and LC-FAO in *Y. lipolytica*

Description	Monooxygenase system	LC-FAO	Reference
EC number	1.6.2.4	1.1.13.13	White & White, 1997
Reaction catalyzed	$R\text{-}CH_3 + O_2 \rightarrow R\text{-}CH_2\text{-}OH + H_2O$	$R\text{-}CH_2\text{-}OH + O_2 \rightarrow R\text{-}CHO + H_2O_2$	Mauersberger *et al*. 1992 Sumita *et al*., 2002
Essential coenzymes	NADPH, FAD and FMN-containing reductases	Flavin-dependent but not FMN, FAD or riboflavin	Ratledge, 1984 Dickinson & Wadforth, 1992 Urlacher *et al*., 2004 van Beilen *et al*., 2003
Localization in system	Membranes of the endoplasmic reticulum	Membrane bound	Barth & Gaillardin, 1996 Iida *et al*., 2000
Induced by	C_{10}-C_{16} *n*-alkanes	*n*-alkanes	Mauersberger *et al*., 1992 Iida *et al*., 2000 van Beilen *et al*., 2003
Designated isozymes	*YlALK1 – YlALK12*	None	Yamagami *et al*., 2004
Cloned and sequenced	Yes (*YlALK1 – YlALK8*)	No[a]	Eirich *et al*., 2004 Fickers *et al*., 2005

[a]Although Fatty Alcohol Oxidase (FAO) has been isolated from *Y. lipolytica* (Kemp *et al*., 1990) no gene has been isolated to date. Eirich *et al*., 2004 have cloned and characterized three FAO genes from *Candida tropicalis*.

Abbreviations used: FAD = Flavin-adenine dinucleotide; FMN = Flavin mononucleotide; NADH = Nicotinamide adenine nucleotide

2.3 Aim of this study

Matatiele (2005) performed Northern blot analysis to investigate induction of the above mentioned *FALDH3* and *FALDH4* genes. It appeared as if *FALDH4* was induced during growth on *n*-alkanes, but the results were not compelling enough. It thus became the aim of this project to study the four putative *FALDH*s from *Y. lipolytica* with regard to their expression and subcellular localization in order to contribute to existing information on these putative *FALDH*s.
The following genetically engineered *Y. lipolytica* strains were constructed to achieve these goals:

- Strains expressing β-galactosidase as reporter gene under control of the native promoters of the four *FALDH*s.
- Strains expressing GFP fused to the native, putative localization sequences from each *FALDH*.

Chapter 3

Materials and Methods

Table 3.1 Fungal and bacterial strains used in this study

Strains	Genotype / characteristics	Source / Reference
Saccharomyces cerevisiae CEN PK 42	*MATα, ura3-52, leu2-3/112 trp1-289, his3Δ1, MAL2-8^c SUC3*	UFS culture collection
Y. lipolytica E150	*MATB, his1, ura3-302, leu2-270, xpr2-322, XPR^p::SUC2*	Barth & Gaillardin, 1996
Y. lipolytica Po1d[a]	MATA, *ura3-302, leu2-270, xpr2-322, XPR^p ::SUC2*	Juretzek *et al*., 2000
Y. lipolytica Po1g	*MATA, leu2-270, ura3-302::URA3, xpr2-322, axp-2*	Madzak *et al*., 2000
Y. lipolytica W29	*MATA*	UFS culture collection
E. coli ER2925	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT2 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New England Biolabs
E. coli XL-10-Gold	*Tet^r D(mcrA)183 D(mcrCB-hsdSMR-mrr)173endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacI^{q}ZDM15 Tn10 (Tet^r) Amy Cam^r]*	Stratagene

[a]This is a recombinant Po1d strain containing the pINA354b-*POX2*p plasmid

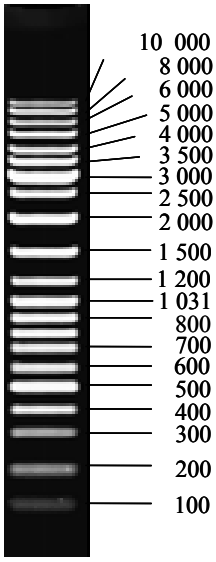
3.1 Enzymes, chemicals, kits and other consumables

All DNA modifying enzymes with their respective buffers, PCR reagents and DNA purification kits were obtained from Amersham, Applied Biosystems, Fermentas, New England Biolabs or Roche Molecular Biochemicals unless otherwise stated.

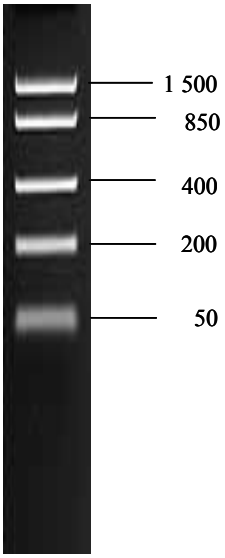
Chemicals used in this study were of an analytical or molecular biology grade and were obtained from either Merck or Sigma-Aldrich, unless otherwise stated.

All oligonucleotides were obtained from Inqaba Biotechnical Industries (South Africa) and oligonucleotide design and analyses were performed using an algorithm from IDT (Integrated DNA Technologies) with default settings. Primers for sequencing were obtained from IDT.

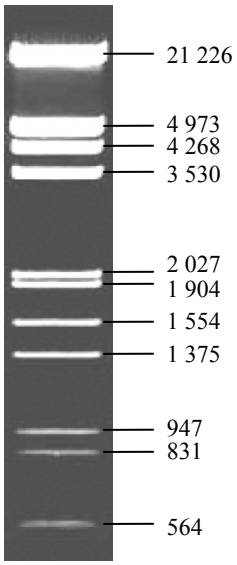
All molecular weight markers were obtained from Fermentas unless otherwise stated in the text.



O'GeneRuler™ 1kb ladder
(1% agarose)



Low Range ladder
(2% agarose)



Lamda (λ) III ladder
(1% agarose)

3.2 Media and growth conditions

Amplication of plasmids in *Escherichia coli* was performed in Luria-Bertani (LB) broth (Sambrook *et al*., 1989) at 37˚C with agitation. Selective pressure was maintained by supplementing the LB broth with either 100 µg/mL ampicillin or 50 µg/mL kanamycin. Solid media cultivations were performed by supplementing broth with 15 g/L bacteriological agar and selective pressure was achieved with 60 µg/mL ampicillin or 50 µg/mL kanamycin.

All *Y. lipolytica* strains were cultured in YPD broth (unless otherwise stated) containing yeast extract (10 g/L), peptone (10 g/L) and glucose (20 g/L). Strains were maintained on YPD solid media containing 15 g/L bacteriological agar or by preserving the culture in 50% glycerol at -72˚C.

3.3 Recombinant DNA techniques

3.3.1 Quantification of nucleic acids

Nucleic acid quantifications were determined with the Eppendorf BioPhotometer (Eppendorf, AG).

3.3.2 Amplification of genes or inserts

Polymerase chain reactions (PCR) were performed with Supertherm (Southern Cross Biotechnology, South Africa) or the Expand Long Template system (Roche Molecular Biochemicals) according to the manufacture's specifications. The reaction mixtures contained the following components: 5 µl of the appropriate 10x buffer (buffer no.3 in the case of Expand Long Template), 1 µM forward and reverse primers (melting temperatures were always 55˚C unless otherwise stated), 200 µM dNTP's, 0.5 µg template DNA and 3.75 U polymerase mix. The reactions were filled to a final volume of 50 µL with PCR grade water.

Thermal cycling was performed using either an Eppendorf Mastercycler Personal or an Applied Biosystems Thermal Cycler 2720. Standard cycling programs (Table 3.2 and 3.3) were used throughout the study for Supertherm and Expand Long Template systems unless otherwise stated.

Table 3.2 Standard PCR reaction for the Supertherm system

Step	Temperature	Time	Cycle number
Initial denaturation	94˚C	2 min	1
Denaturation	94˚C	15 sec	
Annealing	55˚C	30 sec	10
Elongation	72˚C	45 sec[a]	
Denaturation	94˚C	15 sec	15
Annealing	55˚C	30 sec	
Elongation	72˚C	45 sec + 5 sec for each successive cycle	
Final elongation	72˚C	7 min	1

[a] Elongation time was based on 45 sec/1kb.

Table 3.3 Standard PCR reaction for the Expand Long Template system

Step	Temperature	Time	Cycle number
Initial denaturation	94˚C	2 min	1
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	10
Elongation	68˚C	4 - 8 min [b]	
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	
Elongation	68˚C	4 – 8 min + 20 sec for each successive cycle	15
Final elongation	72˚C	7 min	1

[a] Elongation time was based on 45 sec/1kb.
[b] Elongation time was dependent on the desired amplicon size.

3.3.3 Sequence analyses

Templates for sequencing were purified using the GFX™ PCR DNA and gel band purification kit according to the manufacture's instructions. For double stranded templates 500 ng of DNA was used in each sequencing reaction. Sequencing was performed with an ABI Prism® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit 3.0 or 3.1 (Applied Biosystems, USA) according to the manufacturer's specifications. Universal sequencing primers T7 or SP6 were used at a final concentration of 3.2 pmol in a reaction volume of 20 µL. The sequencing PCR cycle is given in Table 3.4. Nucleotide composition was determined on an ABI Prism® 377 DNA Sequencer (Perkin Elmer Biosystems, USA). The data was analysed using the Chromas Lite 2.0 and DN*Assist* 2.2 software.

Table 3.4 Sequencing PCR when using T7 or SP6 primers

Step	Temperature	Time	Cycle number
Denaturation	96˚C	10 sec	
Annealing	50˚C	5 sec	30
Elongation	60˚C	4 min	
Hold	4˚C	∞	1

3.3.4 Analyses of PCR amplicons or restriction enzyme products

All amplicons were electrophoresed and assessed on 1% (w/v) agarose gel (unless stated otherwise) containing 2.5 mg/mL ethidium bromide. Agarose gels were prepared and electrophoresed in TAE buffer [0.1 M Tris, 0.05 M EDTA (pH 8.0) and 0.1 mM glacial acetic acid] at 7 V/cm for 40 min. Visualizations were done with a ChemiDoc XRS (Bio-Rad Laboratories) under short wavelength UV light. For cloning purposes, DNA was excised from the agarose gel by using a DarkReader™ transilluminator (Fermentas) for visualization of the DNA.

3.3.5 Transformation of *E. coli* and *Y. lipolytica*

Transformations in *E. coli* were performed with XL-10 Gold® cells obtained from Stratagene. Cells were rendered competent by means of a modified version of the $RbCl_2$ method as originally described by Hanahan, 1983 and used as 50 µL aliquots. Cloning of ligation mixtures were done with pGEM®-T Easy (Promega) according to the manufacture's directions. The transformation was performed as described by Sambrook *et al*., 1989 and cells were then plated onto LB plates supplemented with ampicllin (section 3.2), IPTG [isopropylthio-β-D-galactoside (10 mg/L) and X-gal [5-bromo-4-chloro-3-indolyl- β-D-galactoside (40 mg/mL)]. Plates were incubated at 37˚C for 15 hours. Positive transformants were selected and inoculated into 5 mL LB-media supplemented with the appropriate antibiotic and grown for 15 hours with aeration at 37˚C.

Transformations of integrative vectors (single copy as well as random integration) in *Y. lipolytica* were performed with competent cells as described by Barth & Gaillardin, 1996. The transformants were grown on appropriate selective media for 2 – 3 days at 28˚C. When selecting for leucine auxotrophy, transformants were plated onto YNB media containing: [YNB without amino acids and ammonium sulfate (1.7 g/L), NH_4SO_4 (5 g/L) and glucose (10 g/L)]. Selection for uracil auxotrophy was performed on YNB-casamino acid media [YNB without amino acids and ammonium sulfate (1.7 g/L), glucose (10 g/L), sodium glutamate (1 g/L), casamino acids (1 g/L). All solid media were supplemented with bacteriogical agar as described in section 3.2. Positive transformants were maintained for further studies or preserved as described in section 3.2.

3.3.6 DNA mini-preparations

DNA mini-preparations were performed using the lysis by boiling protocol (Sambrook *et al*., 1989). Screening for positive recombinant plasmids was performed by using restriction analysis and/or sequence analysis. Positive clones were gel purified with the GFX™ PCR DNA and gel band purification kit according to the manufacture's instructions and the DNA was eluted with 50 µl elution buffer (10 mM Tris-HCl, pH 8.0). Purified samples were stored at - 20˚C when not is use.

3.3.7 Genomic DNA isolation

Genomic DNA from all *Y. lipolytica* strains was prepared from 5 mL YPD cultures, grown for 24 hours at 37˚C with agitation. Two milliliters of the culture was harvested by centrifuging for 1 min, at 14 000 x *g* in 2 mL tubes. The supernatant was discarded and the pellets were gently resuspended in cell lysis buffer [100 mM Tris-HCl (pH 8.0), 50 mM EDTA, 1% SDS] and kept on ice. One volume of glass beads (*ca*. 200 µL) was added to the suspension and vortexed for 4 min with immediate cooling on ice afterwards. After cooling, 275 µL of 7 M ammonium acetate (pH 4.0) was added, vortexed briefly and incubated at 65˚C for 5 min. To this 500 µL chloroform was added, vortexed briefly and centrifuged (10 min, 4˚C, 14 000 x *g*). The supernatant was aspirated

and the genomic DNA was precipitated with 1 volume of isopropanol for at least 5 min at room temperature. Genomic DNA was collected by centrifugation (10 min, 4˚C, 14 000 x *g*). The resulting pellet was washed with 70% (v/v) ethanol while centrifuging (5 min, 4˚C, 14 000 x *g*). The ethanol was aspirated and resulting pellets were dried under vacuum in a SpeedyVac (Savant, USA). Dried pellets were suspended in 100 µL TE buffer [10 mM Tris (pH 8.0), and 1 mM EDTA] containing 5 mg/mL RNase for 1 hour at 37˚C. Samples were stored at 4˚C when not in use.

3.4 Construction of Promoter-Terminator (PT) and Promoter-*lacZ*-Terminator (PlacT)- cassettes

In order to study the expression of the four putative *FALDH* promoters, PT- and PlacT-cassettes were constructed by means of two methods. The first method used was the Sticky-end Polymerase Chain Reaction (SEP) method (Maftahi *et al*., 1996; Fickers *et al*., 2003). Since the SEP method yielded variable results and was not very reproducible, a, second method was aslo employed namely: Enzyme-free cloning (Tillet & Neilan, 1999). The PT-cassettes of *FALDH* 3 and 4 were already constructed by Matatiele (2005) and thus only the PT-cassettes of *FALDH* 1 and 2 were constructed. The PT 1 and 2 cassettes were provided for the study of Matatiele (2005) in exchange for triple and quadruple *FALDH* deletion mutants. All four PT-cassettes were then used to clone a β-galactosidase gene (*lacZ*) between the promoter and terminator region to yield a PlacT-cassette. These cassettes would then be used to study expression of the promoters of the four *FALDH* isozymes.

Please note that the author will from now on refer to the *FALDH* promoter regions as ‘promoters’. This is strictly speaking not correct, since these *FALDH* promoters are located within an arbitrary piece of nucleotide sequence. These promoters have not been functionally dissected or analyzed prior to this study. The same principle will apply to the terminator regions that will be referred to as ‘terminators’ of the *FALDH* isozymes from *Y. lipolytica*.

3.4.1 Experimental set up: SEP method

Four putative fatty aldehyde dehydrogenase (*FALDH* 1 - 4) sequences were kindly provided by Dr. Jean-Marc Nicaud. These sequences consisted of a promoter (P), an ORF and terminator (T). *Y. lipolytica* E150 strains were obtained from the UFS culture collection and genomic DNA was extracted from this strain (section 3.3.7). The P and T region of each isozyme was amplified from genomic DNA by using primer sets in the first PCR (Table 3.5) as indicated in Figure 3.1 (A). The rare 18 bp I-*Sce*I sequence was introduced on primers R_1 and F_2.

In Figure 3.1 (B), the P and T amplicons were denatured and I-*Sce* I sticky ends (thick black lines) annealed to yield a small double stranded piece of DNA (thick red lines). Once this double stranded product is formed, there are free 3′-OH ends. A double stranded template and free 3′-OH ends are the most important prerequisites for the polymerase to perform elongation and this small double stranded piece of DNA acts as two 'primers' in both directions. Primers 1 and 4 are also part of the reaction in PCR 2 and help to amplify a complete PT-cassette (C). The PT-cassettes were cloned into pGEM®-T Easy and amplified and isolated from competent *E. coli* (section 3.3.5 - 3.3.6). Initial screening was performed with *Eco*RI to identify positive clones.

The β-galactosidase gene was amplified from the yeast shuttle vector YIp356R with primer sets LacZ F and LacZ R (Table 3.5). PT-cassettes were digested with I-*Sce* I and then purified from an agarose gel. Purified, digested PT-cassettes were then dephosphorylated with calf intestine alkaline phosphatase (Roche) as specified by the manufacture. Purified *lac*Z amplicons were also digested with I-*Sce* I at the same time as the dephosphorylation step progressed.
The dephosphorylated and I-*Sce* I digested samples were then purified from solution and ligated in a 1:1 ratio to each other for at least 1 hour at 22 ˚C. The result is a PlacT-cassette (**P**romoter region-***l****ac****Z***-**T**erminator region) as illustrated in Figure 3.1 (D).

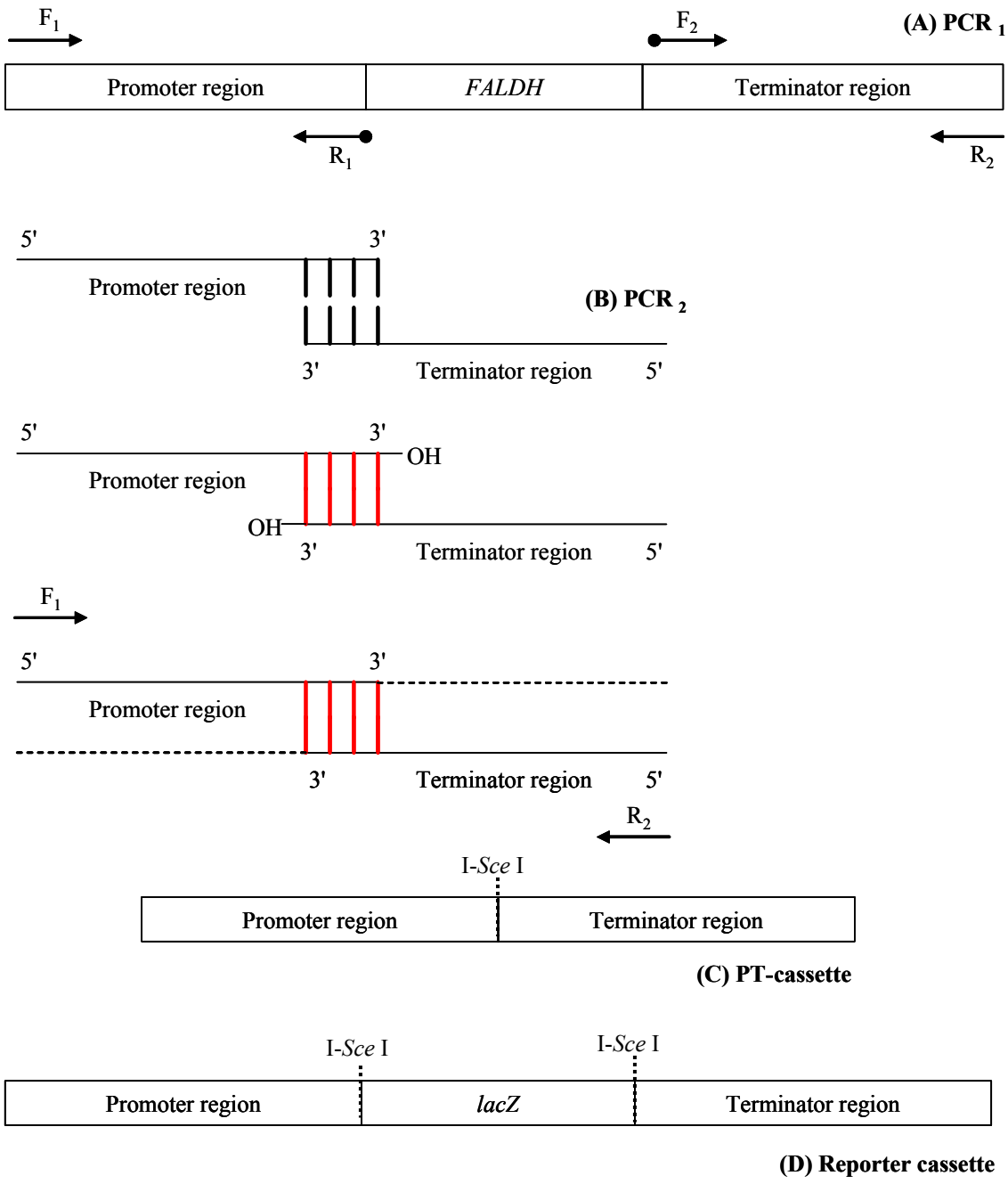


Fig. 3.1 SEP method for constructing a reporter system for *Y. lipolytica*.

Table 3.5 Oligonucleotide primers used for construction of PT and PlacT-cassettes

Primer name	Sequence in 5′ – 3′ orientation	Restriction sites introduced & application
FALDH 1 F1	CGT GGA ATC GGC TGG GAA TGA AAG	SEP
FALDH 1 R1	ATT ACC CTG TTA TCC CTA GTT GTA CAC CCA ACA TTT C	I-*Sce* I
FALDH 1 F2	TAG GGA TAACAG GGT AAT ATG AAT GAT AAA GAG ATA ACA GCG	I-*Sce* I
FALDH 1 R2	CGC GTG TGT GGT AAT TGA AAT TGG GC	SEP
FALDH 2 F1	CTG GGA AGT GCG ACC ATA GAT TGC	SEP
FALDH 2 R1	ATT ACC CTG TTA TCC CTA CGT TGT GTT CAA ATG TTA	I-*Sce* I
FALDH 2 F2	TAG GGA TAA CAG GGT AAT TTT ACT AAC CAC AGT TCG	I-*Sce* I
FALDH 2 R2	GCA CTT GGT GAC AAG CAC ACA GT	SEP
FALDH 3 F1	TTA CCC TGC CTT GAA CGT CGG TGA AAC	SEP
FALDH 3 R1	ATT ACC CTG TTA TCC CTA GGA CAA GTG TCA ATG TTG	I-*Sce* I
FALDH 3 F2	TAG GGA TAA CAG GGT AAT GTG GAG TTT ATG GGT GAG TAA T	I-*Sce* I
FALDH 3 R2	ACC CTT CAA CTA TGA TCA ACC GCT GAA GAG AGA TAT AGC AGT	SEP
FALDH 4 F1	GAT GAC ATG ATG AAG TTG GCC CCT TCA AAT	SEP
FALDH 4 R1	ATT ACC CTG TTA TCC CTA GTT TTT ATT GGT GGT GTG TTT	I-*Sce* I
FALDH 4 F2	TAG GGA TAA CAG GGT AAT GAC TAA CCC TAC TTC CTC ATA	I-*Sce* I
FALDH 4 R2	CTT AGT GGT TGT GGC AAC AGT GTA CGT AG	SEP
LacZ F1	TAG GGA TAA CAG GGT AAT ATG ACC ATG ATT ACG G	I-*Sce* I
LacZ R1	ATT ACC CTG TTA TCC CTA TTA TTT TTG ACA CCA GAC C	I-*Sce* I

Table 3.6 Vectors used for construction of PT and PlacT-cassettes

Vectors	Description	Source / Reference
pGEM®- T Easy	Cloning vector with T-overhangs used for cloning of *Taq* DNA polymerase amplicons	Promega, USA
YIp356R	*Saccharomyces cerevisiae* shuttle vector containing a *lac*Z gene	Dr. J. Albertyn (University of the Free State, South Africa)

3.5 Constructing PT and PlacT-cassettes with the Enzyme-free cloning method

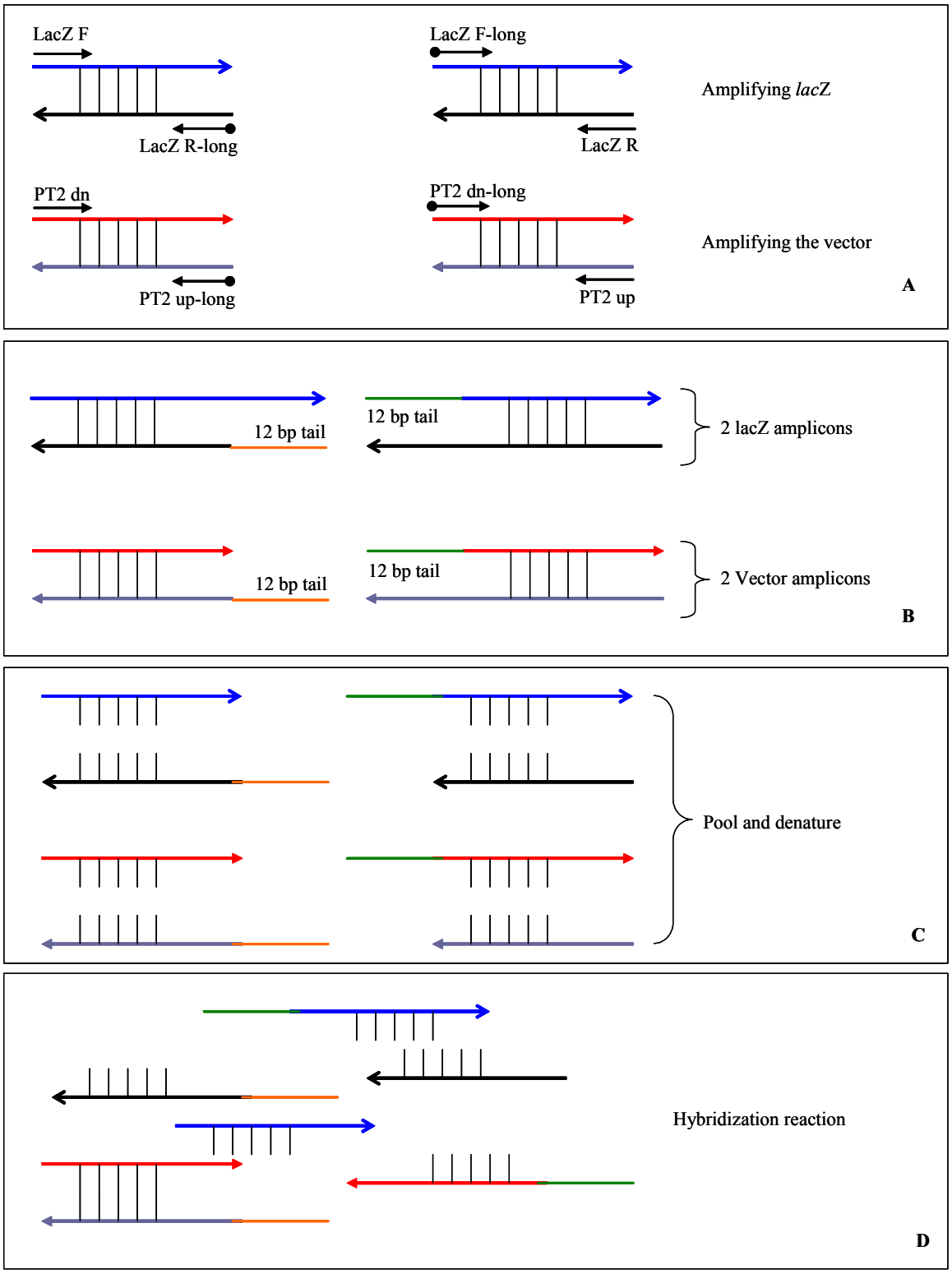
3.5.1 Experimental setup: Enzyme-free cloning

The enzyme free cloning method as described by Tillet & Neilan (1999) was used in this study with slight modifications. In short, the protocol calls for two sets of two parallel PCRs to be performed: one set for the amplification of the gene and the other set for the vector. The gene and vector amplicons are then pooled, denatured and allowed to hybridize to yield a construct with a cloned gene without the use of ligase.

In Figure 3.2 (A), the *lac*Z gene was amplified from YIp356R with two sets of primers (Table 3.7) of which two contained a 12 bp tail. The vector (harboring a promoter- and terminator region cassette, section 3.4.1) was also amplified with two sets of primers (Table 3.7) of which two primers contained a 12 bp tail that was complementary to the 12 bp tail from the *lac*Z amplicons. The resulting *lac*Z and vector amplicons Figure 3.2 (B) contained the incorporated 12 bp tails (orange and green lines). All amplicons were gel purified and the two *lac*Z amplicons and two vector amplicons were then pooled separately.

A denaturation and hybridization reaction was performed using equimolar quantities of each pooled component (Fig. 3.2 C and D) in 50 uL hybridization buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.4 and 1 mM EDTA, pH 8.0). Hybridization was carried out at 95˚C for 3 min followed by four cycles at 65˚C for 2 min and 25˚C for 15 min.

The resulting amplicon contains a vector backbone with a cloned *lac*Z gene, flanked by a promoter- and terminator region which was transformed into competent XL-10 Gold *E. coli* cells.



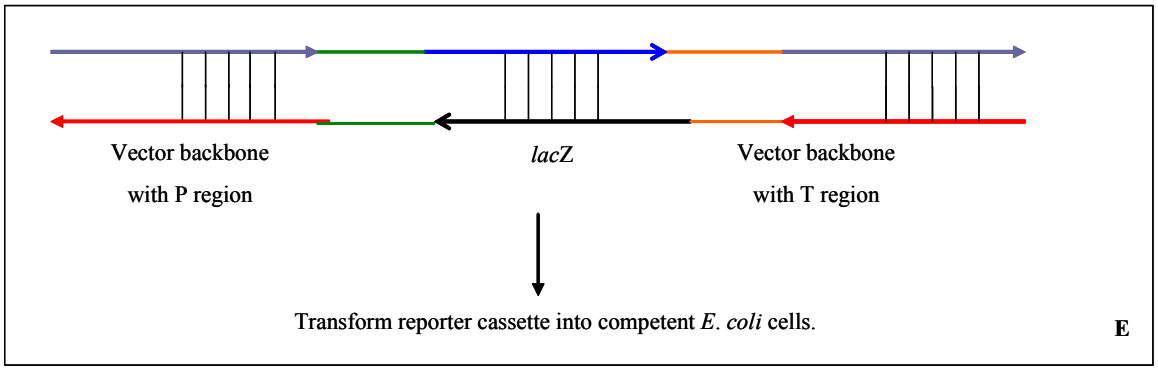


Fig. 3.2 Enzyme-free cloning method for constructing a reporter system by using *lacZ* as the reporter gene.

Table 3.7 Primers used for enzyme-free cloning of PlacT-cassettes

Primer name	Sequence in 5′ – 3′ orientation	Introduced basepairs & application
LacZ F	ATG ACC ATG ATT ACG GAT TCA CTG	Enzyme-free cloning
LacZ R	TTA TTT TTG ACA CCA GAC ACT GG	Enzyme-free cloning
LacZ F-long	GGT ACG GAC TAT ATG ACC ATG ATT ACG GAT TCA CTG	12 bp tail
LacZ R-long	GAA CTA GGA CAT TTA TTT TTG ACA CCA GAC CAA CTG G	12 bp tail
PT2 dn	TAT GAT ATT TTT ACT AAC CAC AGT TCG CC	Enzyme-free cloning
PT2 up	GAT TGG CGT TGT GTT CAA ATG TT	Enzyme-free cloning
PT dn-long	ATG TCC TAG TTC TAT GAT ATT TTT ACT AAC CAC ATT TCG CC	12 bp tail
PT up-long	ATA GTC CGT ACC GAT TGG CGT TGT GTT CAA ATG TT	12 bp tail

Table 3.8 Vectors used for enzyme-free cloning of PlacT-cassettes

Vectors	Description	Source / Reference
pGEM®-T Easy:: PT cassette 2	Cloning vector containing the PT-cassette of *FALDH* 2	This study
YIp356R	*Saccharomyces cerevisiae* shuttle vector containing a *lac*Z gene	Dr. J. Albertyn (University of the Free State, South Africa)

3.6 Expression studies of the four *FALDH* promoters using the pINA781 vector

It was decided to clone each of the four putative *FALDH* promoters (p*FALDH*) of each into the integrative vector pINA781 (Madzak *et al*., 1999). This vector had two useful components: a built-in *lac*Z gene and a pBR docking site that is used for directed, single copy integration into the genome of *Y. lipolytica* Po1g. Integration is facilitated by linearizing the vector with *Not*I. Figure 3.3 illustrates the main components of the vector.

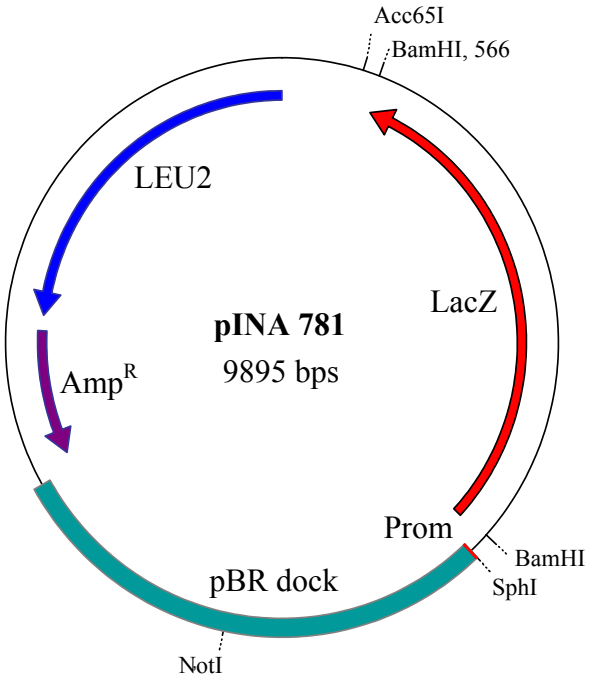


Fig. 3.3 The mono-copy, integrative pINA781 vector containing the pBR dock, minimal leucine promoter and *lacZ* gene used for expression studies in *Y. lipolytica*.

3.6.1 Cloning the p*FALDH* into pINA781 to study expression

The four putative p*FALDH* were amplified from the genome of *Y. lipolytica* E150 with Supertherm (Roche) according to the standard cycling profile as described in Table 3.2 (section 3.3.2). Primers used in the amplification procedure incorporated a *Sph*I and *Bam*HI on the 5′ and 3′ side respectively (Table 3.9). Amplicons were loaded onto a 1% agarose gel and assessed as described in section 3.3.4. Purified promoter amplicons were subcloned into pGEM®-T Easy with an insert:vector ratio as recommended by the manufacturer. Clones were screened with *Eco*RI to confirm positive ligation.

Before the pINA781 vector could be used for cloning purposes it had to be re-amplified to facilitate the removal of an additional *Bam*HI site present at position 566 (Fig. 3.3). This was accomplished by performing a divergent PCR with oligonucleotide primer set p781-F and p781-R to delete the *Bam*HI 566 site. Amplification was performed with Expand Long Template (Roche) and the standard cycling profile as described in Table 3.3 (section 3.3.2) was followed except that elongation was performed at 8 min. Amplicons were then treated with 5U *Dpn*I for 1 hour at 37˚C and then purified directly from solution with the GFX™ PCR DNA and gel band purification kit from Amersham. The purified samples were then treated with 10 Weiss units of T4 DNA Ligase as described by the manufacturer.
Ten micro liters of the ligation mixture was transformed into competent XL-10 Gold *E. coli* cells and plasmid extraction was performed as described in section 3.3.5 and 3.3.6.
Clones were screened with *Bam*HI and *Not*I in a double digestion reaction and assessed on a 1% agarose gel to select clones of which the *Bam*HI 566 site was removed.

Positive clones containing the promoter of each isozyme in pGEM®-T Easy and the pINA781 vector (with the abolished *Bam*HI 566 site) were both treated with a double digestion of *Bam*HI and *Sph*I. The corresponding fragments were excised from the gel and purified. The promoter fragments were ligated to the digested pINA781 vector using an insert to vector ratio of 3:1. Ten micro liters of each ligation reaction was transformed into competent XL-10 Gold *E. coli* cells and plasmid extraction was performed as

described in section 3.3.5 and 3.3.6. All positive clones were preserved at -20˚C until further usage.

A flowchart of the cloning strategy is provided in Figure 3.4.

Table 3.9 Oligonucleotide primers used for cloning *FALDH* promoters into pINA781

Primer name	Sequence in 5′ – 3′ orientation	Restriction sites introduced & Application
Prom 1F	GCA TGC GCA ACA TCA TTT TTT GGA TCA TC	*Sph*I
Prom 1R	GGA TCC TTC ATT GTG GCG GAA GTT GTA CAC CC	*Bam*HI
Prom 2F	GCA TGC TTTT GAA TAT TTG TTT AGT CAT TTG CTA TGT TAT TTC CAA TCA CCC	*Sph*I
Prom 2R	GGA TCC TTC ATG ATT GGC GTT CTG TTC AAA TGT TAG TAA ATA TTA ATA TTC GTG C	*Bam*HI
Prom 3F	GCA TGC TCG TTA CCC TGC CTT GAA CGT C	*Sph*I
Prom 3R	GGA TCC TTC ATT CAG TGG GGG GAG AGC G	*Bam*HI
Prom 4F	GCA TGC CTC AAG CGA CTT TCT GTG GAA ACA CG	*Sph*I
Prom 4R	GGA TCC TTC ATT TTT ATT GGT GGT GTG TTT GAG GAA GGA GG	*Bam*HI
p781-F	CCC CTG CCC GGT TAT ATT ATT TTT GAC AC	Divergent PCR
p781-R	AAC TAC GGA ACT TGT GTT GAT GTC TTT GC	Divergent PCR

Table 3.10 Vectors used for cloning *FALDH* promoters into pINA781

Vector	Description	Source / Reference
pGEM®- T Easy	Cloning vector with T-overhangs used for cloning of *Taq* DNA polymerase amplicons	Promega, USA
pINA781	Integrative vector harboring the pBR322 docking site for directed, single-copy integration into the genome of Po1g	Madzak *et al*., 2000

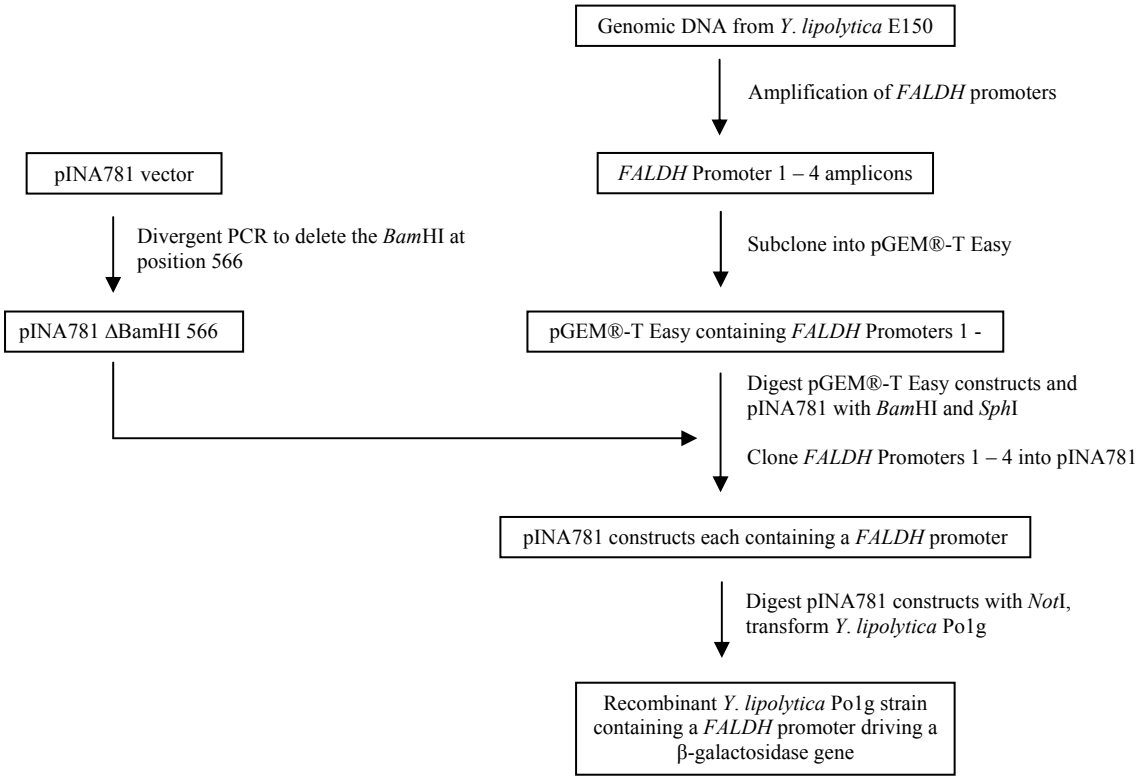


Fig. 3.4 Cloning of the four p*FALDH* into pINA781

3.6.2 Transformation of *Y. lipolytica* Po1g with pINA781 containing putative p*FALDH*

The pINA781 vector containing the native p*FALDH* of each isozyme was transformed into competent *Y. lipolytica* Po1g cells with the lithium acetate method as described by Barth & Gaillardin, (1996). One microgram of pINA781 plasmid was linearized with *Not*I, in order for the pBR dock (Fig. 3.3) to integrate into the genome of the yeast by means of homologous recombination, and transformed to *Y. lipolytica* Po1g. The transformed *Y. lipolytica* Po1g cells were plated out on selective media to select for leucine prototrophs (section 3.3.5). After 3 days of growth at 28˚C, colonies were streaked out on to fresh selective media to create a second progeny.

3.6.3 Preliminary screening of promoter activity

Presumptive positive transformants were streaked on minimal media containing 1.7 g/L YNB (without amino acids and without ammonium sulphate), 1% carbon source and X-gal (40 µg/ml). Hydrophobic carbon sources like dodecane, hexadecane and oleic acid were sonicated togther with 1% Tween 80. Glucose and glycerol were also included as carbon sources. The YNB media was made up in 50 mM phosphate buffer, pH 7.4 and supplemented with 15 g/L bacteriological agar. Plates were incubated for 3 days at 28˚C or until a blue colour reaction was observed. After every transformation, 10 *Y. lipolytica* transformants were tested. All transformants that were subjected to the colour screening were also cross checked by PCR to confirm integration. The *FALDH* promoter and *lacZ* fusions (p*FALDH*::*lacZ*) were amplified with primers p781-F and the forward primer of each promoter primer with Expand Long Template (Roche) as described in Table 3.3 with the exception that the elongation step was performed for 3 min at 68˚C.

3.6.4 Growth of the recombinant *Y. lipolytica* Po1g strains for assay purposes

Recombinant *Y. lipolytica* Po1g strains containing pINA781 p*FALDH*::*lacZ* were pre-cultured in YPD media (section 3.2) for 16 hours at 28˚C on an orbital shaker until a cell density of *ca*. 1.0 x 10^8 cells/mL. The cells were harvested by centrifugation at 5000 x *g* for 10 min and the resulting pellet was washed twice with 50 mL sterile YNB culture media buffered in 50 mM phosphate buffer, pH 6.8. The cells were counted on a standard hemacytometer and inoculated into 50 mL media in 500 mL flasks to a final concentration of 6 x 10^6 cells/mL. The media contained the following components: 1.7 g/L YNB (without amino acids and ammonium sulphate), 4 g/L ammonium chloride, 0.1 g/L uracil and 0.32 g/L leucine. Glucose, glycerol or hexadecane served as carbon sources with the addition of alternating combinations of glucose and glycerol with hexadecane. One milliliter samples were taken from triplicate shake flasks at 5, 10, 15, 20, 25, 30, 48 and 58 hours. Cultures containing only hexadecane were first cultured for 30 hours at 28 ˚C with agitation since *Y. lipolytica* W29 derivatives experience a long lag phase on *n*-alkanes before generating significant biomass. After 30 hours, the normal 58 hour sampling, as described above, commenced as usual. Optical density readings were measured at OD_{620} to monitor growth and cells were harvested at 13 000 x *g*.

Supernatants were discarded and the resulting pellets were frozen at -20˚C until a β-galactosidase assay was performed.

3.6.5 Determinations of β-galactosidase acitivity

Pellets were resuspended in 0.5 mL buffer Z (100 mM Sodium phosphate, pH 7.0, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol) containing protease inhibitor cocktail (Roche) and kept on ice. The cells were then treated with Y-PER® (Pierce) for 20 min while being gently agitated at room temperature and then centrifuged at 14 000 x *g* for 10 min. The supernatant was removed and glass beads were added followed by cell breaking for 3 min (30 sec breaking followed by a 30 sec cooling period on ice) and the cell suspension transferred to clean tubes. This suspension was centrifuged at 13 000 x *g* for 3 min to pellet the cell debris. The supernatant was aspirated and used for a β -galactosidase activity assay and for specific protein activity assay. The colour reaction of the β-galactosidase assay was started by adding 0.2 mL of 4 mg/mL σ-nitrophenyl-β-D-galactopyronoside (ONPG) which was equilibrated together with the supernatant at 37 ˚C prior to the assay. Samples were incubated for 2 min and the reaction was stopped by adding 1 M Na_2CO_3. The yellow colour (formation of σ-nitrophenol) was measured at 405 nm with a 96 well microtiter plate reader (Labsystems iEMS Reader MF, Finland). The β-galactosidase activity was calculated as follows:

$$\text{Activity.mL}^{-1} = OD_{405} \times 1000 / (V \times t)$$

Where : V = volume of supernatant (mL)
t = duration of incubation (min)

The activities are calculated as a percentage of the highest activity observed for the *POX2* promoter during growth on hexadecane.

3.6.6 Construction of standard curves

A standard curve of absorbance (405 nm) vs. σ-nitrophenol concentration was constructed by dissolving σ-nitrophenol (Fluka) in buffer Z and preparing serial dilutions. The dilutions were done in triplicate and the data was subjected to linear regression to construct the curve.

For the total protein determination assay, a standard curve of absorbance (540 nm) vs. bovine serum albumin concentrations was prepared with a Micro BCA™ Protein Assay Kit (Pierce) as specified by the manufacturer. Data was generated in triplicate and linear regression was performed to obtain the best fit.

3.7 Evaluating the toxicity effects of triple and quadruple *FALDH Y. lipolytica* mutants when cultured on 1-dodecanol and 1-dodecanal

Triple and quadruple deletion mutants of *Y. lipolytica* Po1d were obtained from Dr. P.R. Matatiele (University of the Free State, South Africa). These mutants were constructed using the Cre-*lox* recombination system as adapted for *Y. lipolytica* by Fickers *et al*., (2003). This was performed by Matatiele (2005) to help establish the role of the four *FALDH* isozymes in *n*-alkane metabolism The *FALDH* ORF of each of the four isozymes was deleted by replacing it with an auxotrophic *URA3* marker to yield the following deletion mutants: *FALDH* Δ234:*URA3*, *FALDH* Δ134:*URA3*, *FALDH* Δ124:*URA3*, *FALDH* Δ123:*URA3* and *FALDH* Δ1234:*URA3*. In this study, we wanted to see the toxicity effects the *FALDH* deletions would have on *Y. lipolytica* when cultured on 1-dodecanol and 1-dodecane.

The following experiment is a modified version of work conducted by Smit *et al*., (2004). Triple and quadruple *FALDH* deletion mutants were precultured in YPD medium for 15 hours. Optical densities were determined at λ = 620 nm with a microtitre plate reader (Labsystems iEMS Reader MF, Finland) and the inocula were adjusted to an OD_{620} = 1.0 in 1 mL YPD media. From the standardized inocula, dilutions of 5x, 50x and 500x were prepared.

YPD_2 media containing 2% (w/v) glucose, 1% (w/v) peptone, 1% (w/v) yeast extract and 2% (w/v) bacteriological agar was sterilized and cooled at 60˚C for 15 min before 1-dodecanol and 1-dodecanal was added. The media was briefly stirred and plates were poured immediately. The final concentrations of the fatty alcohol and its corresponding aldehyde were 0%, 0.002%, 0.01% and 0.025%. Once the plates were dry, 2.5 µL aliquots of the *Y. lipolytca* cell dilutions were spotted onto the solid media with the aid of a grid and left to completely dry. The plates were enclosed with parafilm and incubated at 28˚C for 2 days.

3.8 Establishing subcellular localization in *Y. lipolytica* using GFP and putative *FALDH* localization sequences

It was decided to use the integrative vector: JMP5 (Fig. 3.5) which had no promoter and had a mutagenesis cassette that consisted of an *ura3d*1 marker flanked by two *zeta* elements that facilitate random integration. The strategy was to clone the inducible *ICL1* promoter (p*ICL1*) in frame with a GFP protein (with an omitted stop codon) and to fuse a putative localization sequence (LSs) from each *FALDH* (containing a stop codon) to the 3′ end of the GFP gene. If the p*ICL1* drove transcription, the GFP would be localized by the putative LSs and could then be visualized in the cell due to GFP fluorescence.

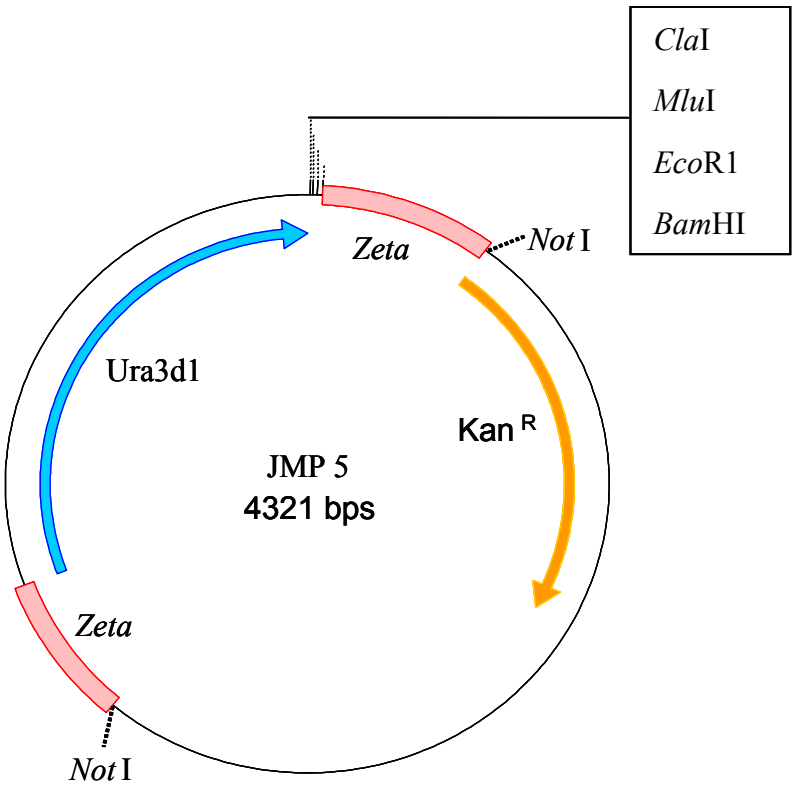


Fig. 3.5 The integrative JMP5 vector that facilitates random integration illustrating the restriction enzymes used for cloning and the mutagenesis cassette.

3.8.1 Fusing the p*ICL1* to GFP

The p*ICL1* was amplified from the expression vector: JMP21-p*ICL*1-CPR which harbored a p*ICL1*. The promoter was excised with *Bam*HI and *Eco*RI and the corresponding fragment was purified from a gel. The fragment was cloned into JMP5 and clones were screened with *Bam*HI and *Eco*RI.

Before the GFP was used for further studies, the functionality of this protein was assessed by transforming the pYES2 vector, containing the GFP, into *Saccharomyces cerevisiae* CEN PK42 as described by Chen *et al*., (1992). Resulting colonies were inoculated into

YPD media and cultured for 12 hours with agitation. Green fluorescence was confirmed with a Zeiss epifluorescence microscope (West Germany) with filter set no. 18.

The GFP protein was amplified from the *S. cerevisiae* expression vector: pYES2 with oligonucleotide primers GFP-F and GFP-R with Supertherm (Roche) as described in Table 3.2 (section 3.3.2). The resulting amplicon was purified and cloned into JMP5 p*ICL1*. Positive clones were identified by means of a double digestion screening with *Bam*HI and *Mlu*I.

3.8.2 Fusing the putative *FALDH* LSs to the JMP5 p*ICL1* GFP chimera

3.8.2.1 Modeling the localization sequences from *R. norvegicus* FALDH

Topology predictions of the four FALDH amino sequences were performed with the PSORT II software using default settings (http://psort.ims.u-tokyo.ac.jp/form2.html) to determine whether these proteins possessed NH_2- terminal LSs. Since the software was unable to detect any NH_2-terminal LSs the presence of COOH-terminal LSs were investigated. To assist in this exercise, the well known COOH-termimal tail of the *R. norvegicus* FALDH was used as a reference. This choice was based on the fact that the *R. norvegicus* FALDH possesses a COOH-terminal tail that is well studied and understood (Masaki *et al*., 1994, 2003). The last 35 amino acids of this particular FALDH serves as a hydrophobic anchor for the protein when inserted into the ER membrane.

The amino acid sequence of the *R. norvegicus* FALDH (accession number: AF289813), was used in a hydrophobic domain analyses by means of the scoring system designed by Kyte and Doolittle (http://us.expasy.org/tools/protscale.html). The last 35 amino acids (excluding the stop codon) were used for the Kyte and Doolittle hydrophobicity plot of the *R. norvegicus* FALDH. The last 35 amino acids of each translated *Y. lipolytica FALDH* were also used to construct Kyte and Doolittle plots. The parameter settings for the analyses were; window size 9, relative weight of window edges 100% and weight variation model, linear.

3.8.2.2 Cloning the four *FALDH* LS into JMP p*ICL1* GFP

Each *FALDH* LS was amplified from the genome of *Y. lipolytica* E150 with oligonucleotide primers (Table 3.11) incorparting a *Cla*I and *Mlu*I restriction site. The amplicons were purified from a 2% agarose gel and assessed with a low range molecular marker (Fermentas). Purified amplicons were subcloned into pGEM®-T Easy with an insert to vector ratio as specified by the manufacturer. Clones were screened with *Eco*RI and putative positive clones were sequenced with both SP6 and T7 universal oligonucleotide primers.

Plasmid maps of the subcloned putative LSs in pGEM®-T Easy (constructed with pDRAW32, version 1.1.87) indicated that the *Cla*I introduced on the primers would yield a DNA adenine methylase (Dam) sensitive amplicon with a resulting adenine methylated *Cla*I palindrome after propagation in *E. coli*. Consequently, sequenced clones that tested positive were transformed into competent Dam negative *E. coli* cells (New England Biolabs). Transformations and DNA mini-extractions were carried out as described in section 3.3.5 and 3.3.6.
The LSs were excised from pGEM®-T Easy by means of a double digestion with *Cla*I and *Mlu*I and cloned into JMP5 p*ICL1* GFP to yield a JMP5 p*ICL1* GFP LS-chimera. Figure 3.6 provides a flowchart of the cloning strategy used to create the JMP5-chimeras.

3.8.2.3 Transforming *Y. lipolytica* E150 to establish subcellular localization

The JMP5 p*ICL1* GFP LS 1-4 constructs were linearized with *Not*I (Fig. 3.5) to facilitate random integration via the *zeta* elements. Linearized vectors were transformed into competent *Y. lipolytica* E150 cells as described by Barth & Gaillardin (1996) and selection was done on solid media as specified in section 3.3.5. Resulting clones were re-streaked on the same media to create a second progeny. To induce the p*ICL1*, cells were streaked out on both rich medium (10 g/L yeast extract, 10 g/L peptone) and minimal YNB medium (1.7 g/L without amino acids and ammonium sulphate, 5 g/L ammonium chloride dissolved in phosphate buffer, pH 6.8 (Mlíčková *et al*., 2004) containing 1%

absolute ethanol. When ethanol was provided in liquid form, it was added aseptically to sterile media. For ethanol in vapour phase, 200 µL ethanol was dispersed (Paryshna & Barth, 2003) on sterile 90 mm, no. 32 filter papers (Whatman). Plates were incubated at 28˚C for 2 days. In addition, colonies were also picked up and induced in the rich as well as the minimal media as described above with ethanol provided in liquid form to serve as carbon source. Cultures were shaken with aeration at 28 ˚C for 24 h.

Fluorescence was intially checked under a Carl Zeiss Axioscope epifluorence microscope (West Germany) by using filter set no. 18. This was followed by confocal laser scanning microscopy (Nikon TE 2000 C1 confocal, Japan).

Table 3.11 Oligonucleotide primers used to construct JMP5 p*ICL1* GFP LS-chimeras

Primer name	Sequence in 5′ – 3′ orientation	Restriction sites introduced
GFP F	GGA TCC ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC	*Bam*HI
GFP R	GAC GCG TTT TGT ATA GTT CAT CCA TGC CAT GTG TAA TCC C	*Mlu*I
LS1 F	GAC GCG TAT TCC TCG AAC CGG CCG AC	*Mlu*I
LS1 R	GAT CGA TCT ACT TAA TAA ACA CCC GAC ATA ATC TGA GCA ATG GTG	*Cla*I
LS2 F	GAC GCG TAA GGC TGT CCA GGG TAG TCT AGC	*Mlu*I
LS2 R	GAT CGA TTC ATA AGA AAA TTC CTG ACT TCA AAC TGG CTG C	*Cla*I
LS 3F	GAC GCG TAG ACA ACC CGG ATG GGT CG	*Mlu*I
LS 3R	GAT CGA TCT AGT TGA AGA GTC TCG ACC AAA ATC CTC TC	*Cla*I
LS 4F	GCA TGA CGC GTC GAA GCA TGG TCA AGA ACC TGC T	*Mlu*I
LS 4R	GCA TGA TCG ATC TAG AGC AGA GCC TTG GCA CC	*Cla*I

Table 3.12 Vectors used for construction of JMP5 p*ICL1* GFP LS-chimeras

Vectors	Description	Source / Reference
pGEM®- T Easy	Cloning vector with T-overhangs used for cloning of *Taq* DNA polymerase amplicons	Promega, USA
JMP21-pICL-CPR	Expression vector constructed from p67RYL and JMP21 containing a cytochrome P450 reductase expressed under the ICL promoter	Nthangeni *et al*. 2004
pYES2	Inducible, expression vector for *S. cerevisiae* containing a URA3 marker and 2µ origin for high copy maintenance	Prof. G. Lindsay (UCT)
JMP5	Integrative vector carrying the mutagenesis cassette (*zeta-ura3d1-zeta*)	Pignéde *et al*., 2000

3.8.2 4 Summary of cloning strategy

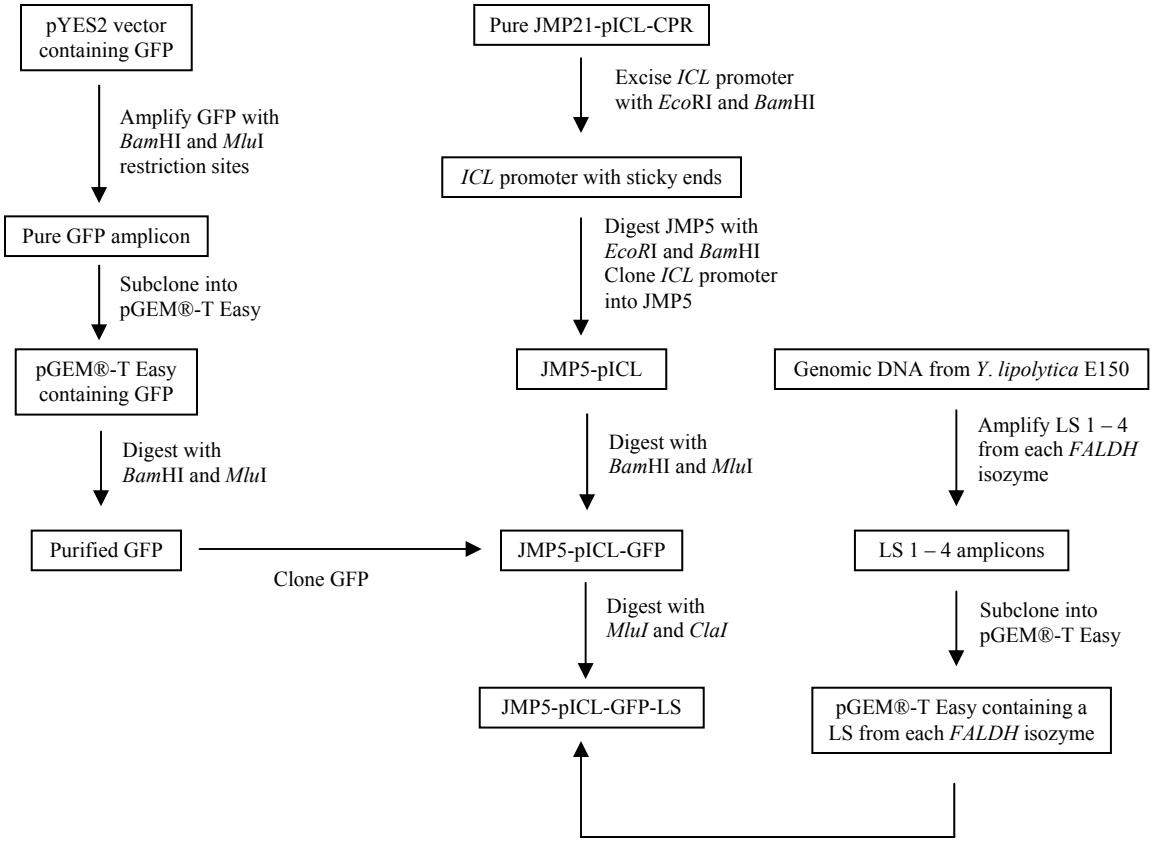


Fig 3.6 Cloning of the p*ICL1*, GFP and four putative *FALDH* localization LSs into JMP5 to study localization in *Y. lipolytica*.

3.9 Cloning of the GFP LS – fragment into pKOV136

The GFP LS - fragment was amplified from the JMP5 p*ICL1* GFP LS construct with primers containing *Bam*HI and *Acc*65I restriction sites to facilitate cloning into pKOV136. The amplicons in question were obtained with Supertherm that was subjected to a cycling profile as described in Table 3.2. Purified amplicons were subcloned into pGEM®-T Easy and presumptive positive clones were purified. The GFP LS - fragment was excised from pGEM®-T Easy by means of a double digestion of *Bam*HI and *Acc*65I according the manufacturer's recommendations and compatible sticky ends in pKOV136

were obtained with an identical double digestion. The GFP LS - fragment was cloned into pKOV136 containing a constitutive transcription elongation factor -1α promoter (p*TEF*) Cloning into pKOV136 was performed with an insert to vector ratio of 5:1 and transformations in *E. coli* and DNA mini-extractions were carried out as described in section 3.3.5 and 3.3.6. Purified pKOV136 GFP LS constructs were transformed into *Y. lipolytica* Po1g as described by Barth & Gaillardin (1996) and selection for leucine prototrophs was performed on media as described in section 3.3.5.

Fluorescence was checked as described in section 3.8.2.3. Before checking fluorescence, recombinant strains were cultured with aeration in YPD media at 28 ˚C for 24 hours to achieve constitutive expression of the GFP LS – fragment in *Y. lipolytica*.

Table 3.13 Oligonucleotide primers used for cloning the GFP LS – fragment into pKOV136

Primer name	Sequence in 5′ – 3′ orientation	Restriction sites introduced
GFP F	GGA TCC ATG AGT AAA GGA GAA GAA CTT TTC ACT G	*Bam*HI
JMP5 R	GGT ACC CGC CCA AGC TTA TCG AT	*Acc*65I

Table 3.14 Vectors used for cloning the GFP LS – fragment into pKOV136

Vectors	Description	Source / Reference
pGEM®- T Easy	Cloning vector with T-overhangs used for cloning of *Taq* DNA polymerase amplicons	Promega, USA
pKOV136	pINA781 with the β-galactosidase gene replaced with the TEF promoter and *LIP2* terminator from pINA1313	Madzak *et al*., 2000 Nicaud *et al*., 2002

Chapter 4

Results

4.1 Constructing Promoter/Terminator- and Promoter-*lacZ*-Terminator-cassettes

4.1.1 Amplification of Promoter/Terminator-cassettes using SEP

The construction of the promoter - terminator (PT) cassette entailed the amplification of each promoter and terminator of each isozyme from genomic DNA. Figure 4.1 illustrates PT-cassettes of *FALDH1* and *2* constructed by means of the Sticky-end PCR (SEP) method as described by Maftahi *et al*., (1996) and Fickers *et al*., (2003). The cassettes were assessed on a 1% (w/v) agarose gel and the expected amplicon sizes were 2103 bp and 1871 bp for the *FALDH1* and *2* PT-cassettes respectively (Fig. 4.1). PT-cassettes 3 and 4 were already constructed by Matatiele (2005).

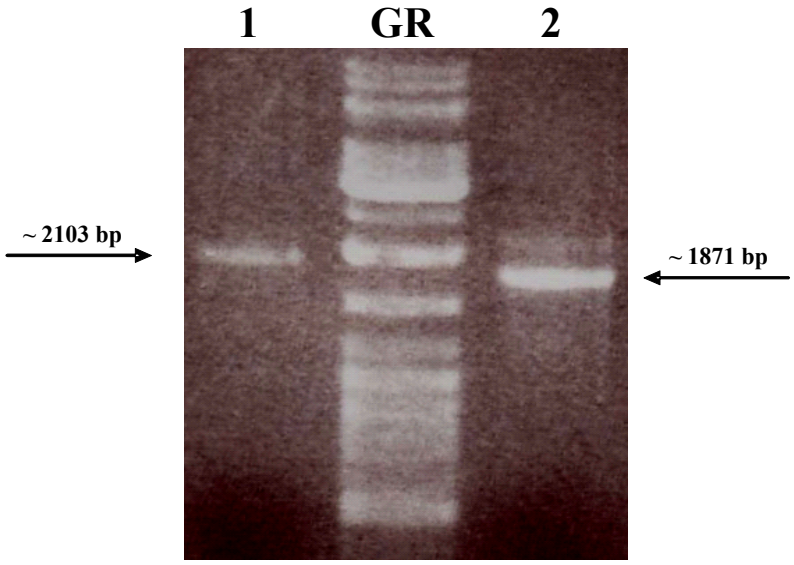


Fig. 4.1 Gel electrophoresis of SEP amplified PT 1 and 2 cassettes ligated to each other by means of I-*Sce* I sticky ends. **Lane 1** = *FALDH1* PT-cassette, **GR** = GeneRuler™ molecular weight marker and **Lane 2** = *FALDH2* PT-cassette.

4.1.2 Construction of Promoter-*lacZ*-Terminator-cassettes for expression studies

Following the construction of the PT-cassettes, these constructs together with the *lacZ* amplicon (amplified from the yeast shuttle vector YIp356R) were digested with I-*Sce* I to yield cohesive sticky ends for ligation. Figure 4.2 illustrates a successful ligation experiment that yielded the **P**romoter-***l****ac*Z-**T**erminator (PlacT)-cassettes for *FALDH1* and *3*. Following *Eco*RI digestion the pGEM®-T Easy is liberated (2997 bp band) from the PlacT-cassette together with 2 bands representing the *lacZ* gene + *FALDH* terminator (*ca*. 4.0 kb band) as well as the *FALDH* promoter (*ca*. 1.1 kb). Even after numerous attempts the PlacT-cassettes for *FALDH2* and *4* could not be successfully constructed.

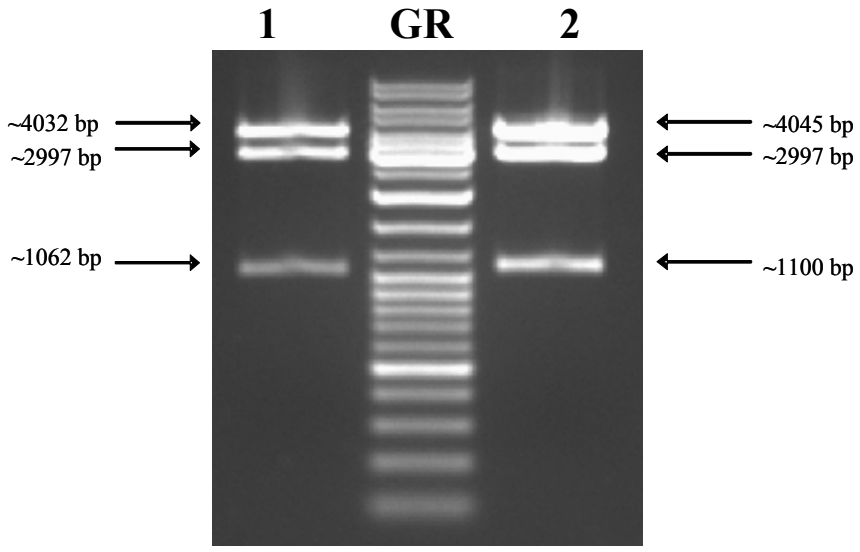


Fig. 4.2 Agarose gel electrophoresis of PlacT-cassettes 1 and 3 digested with *Eco*RI. **Lane 1** = *FALDH1* PLACT-cassette, **GR** = GeneRuler™ molecular weight marker and **Lane 2** = *FALDH3* PLACT-cassette. The band of 2997 bp represents the backbone of pGEM®-T Easy while the 1.1 kb bands represent the *FALDH* promoter portion of the construct. The bands in the 4.0 kb range represent the *lacZ-FALDH* terminator portion of the construct.

4.2 Enzyme free cloning experiments

Since we achieved varied degrees of success with the SEP method and results were not reproducible, we opted for another method to construct PlacT-cassettes in order for us to

study the expression of these *FALDH* isozymes under transcriptional control of their native promoters.

The enzyme-free cloning method as described by Tillet and Neilan (1999) entails two sets of PCR reactions in which two 'gene amplicons' (in this case the *lac*Z gene) and two 'vector amplicons' (in this case pGEM®-T Easy containing a PT-cassette) are obtained. This population of amplicons are then pooled in equimolar quantities and hybridized to yield a PlacT-cassette. No data is presented from the hydridization step since no PlacT-cassettes could be constructed. However Figure 4.3 illustrates the two *lac*Z and two vector amplicons that comprise the first half of the enzyme free cloning procedure.

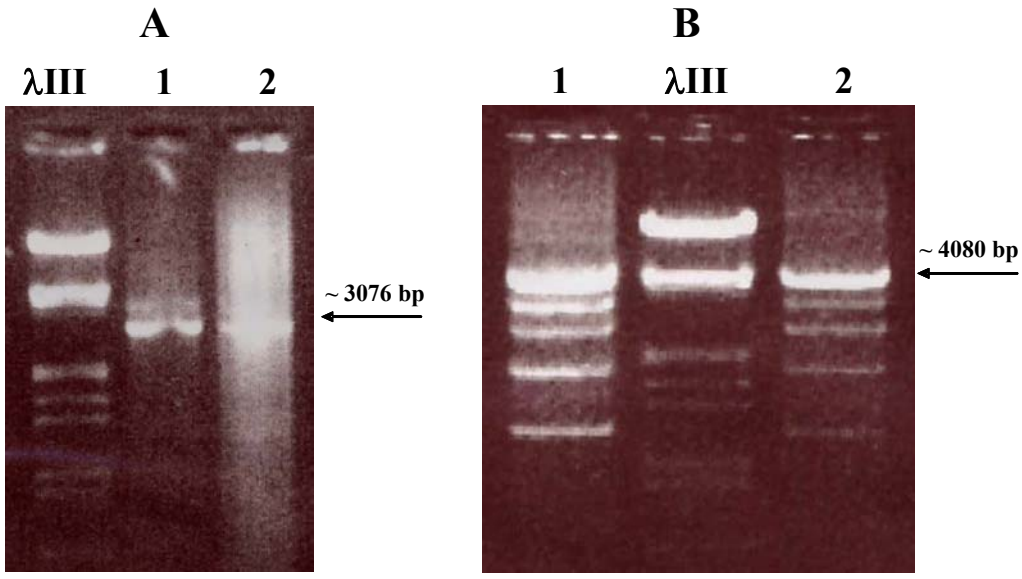


Fig. 4.3 **(A) Lanes 1** and **2** contain the *lac*Z amplicons that possess complementary tails to those of the vector amplicons. **(B) Lanes 1** and **2** represent the vector amplicons each containing a complementary tail to those of one of the *lac*Z amplicons. The sizes of the amplicons were estimated with the Lambda III DNA molecular weight marker (**λ III**).

4.3 Expression studies with pINA781 in *Y. lipolytica* Po1g

Due to the inability to construct a PlacT-cassette with either SEP or Enzyme-free cloning, for all four the *FALDH* genes it was decided to use the pINA781 vector as basis for an expression vector. This vector contains a *lac*Z gene driven by a minimal leucine promoter

as well as the pBR322 docking site for directed, single copy integration into the genome of *Y. lipolytica* strain Po1g. All four *FALDH* promoters (encompassing a region of *ca*. 1000 bp upstream from the initiation codon) from the putative *FALDH*s were amplified from genomic DNA (Fig. 4.4 A and B). The primers used in the amplification procedure incorporated a *Sph*I and *Bam*HI restriction site on the 5′ and 3′ side respectively of each promoter amplicon.

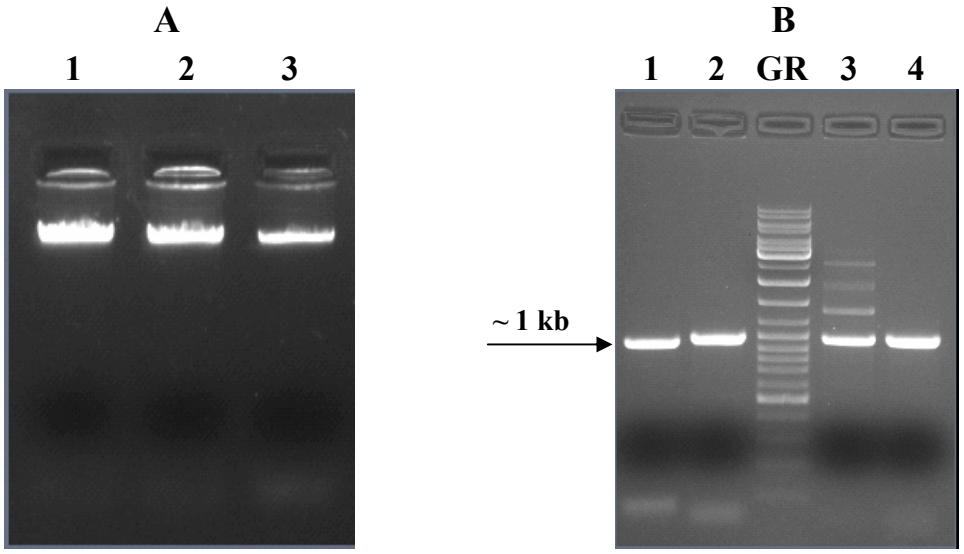


Fig. 4.4 **(A)** Genomic DNA extracted from *Y. lipolytica* E150. **Lanes 1 – 3** represent three independent extractions from the same culture broth. **(B)** Promoter amplicons as amplified from genomic DNA. **Lane 1** = *FALDH1* promoter, **Lane 2** = *FALDH2* promoter, **GR** = GeneRuler™ molecular weight marker, **Lane 3** = *FALDH3* promoter and **Lane 4** = *FALDH4* promoter.

4.3.1 Cloning of the putative *FALDH* promoter regions into pINA781

Following digestion of the pINA781 with *Sph*I and *Bam*HI (which both removed the minimal leucine promoter and linearized the vector), the promoter regions were directionally cloned into the pINA781 expression vector to yield a construct that drives transcription of a β-galactosidase gene through the four *FALDH* promoters (p*FALDH*s). Verification of the correct clones was performed with a double digestion using *Sph*I and *Bam*HI to yield a liberated promoter region and the pINA781 backbone (Fig. 4.5 A and B). Figure 4.5 (C) represents an example of a successfully constructed expression vector containing p*FALDH1*.

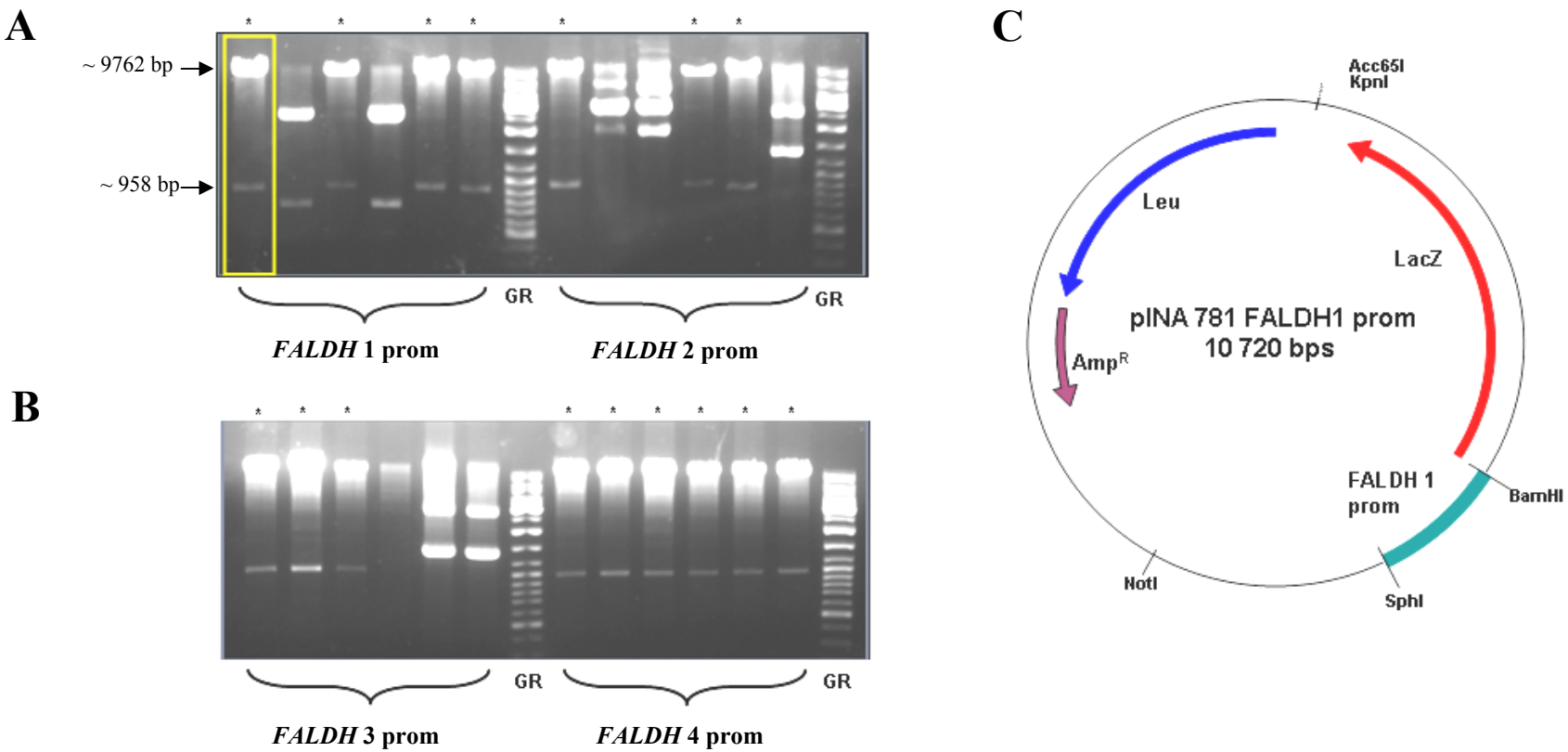


Fig. 4.5 Digestion with *Bam*HI and *Sph*I of pINA781 containing the four cloned putative *FALDH* promoters. Correct digestion profiles are annotated with an asterisk. Digestion profile of pINA781 with *FALDH* promoters *1* & *2* **(A)** and *3* & *4* **(B)**. **GR** = GeneRuler™. **(C)** Plasmid map of the mono-integrative vector pINA781 harboring a *Bam*HI, *Sph*I-cloned putative promoter from *FALDH* 1 (yellow bordered digestion profile in A) driving the β-galactosidase gene. The vector also comprises of a pBR322-dock which can be linearized with *Not*I to facilitate single copy genomic integration in *Y. lipolytica* Po1g.

4.3.2 Confirming genomic integration and assessing promoter activity

After transformation of *Y. lipolytica* Po1g with the pINA781 vectors (containing the various putative *FALDH* promoters), a two-part strategy was followed to verify integration and to assess promoter activity. Firstly, genomic DNA was extracted from the recombinant *Y. lipolytica* Po1g strains and the entire promoter and β-galactosidase gene (*lacZ*) was amplified (*ca*. 4 kb) from the genomic template (Fig. 4.6). The second leg of the strategy was to streak out the above mentioned transformed *Y. lipolytica* Po1g strains as well as appropriate controls onto minimal medium (containing 40 µg/ml X-gal) to detect possible promoter activity. The plates contained one of the following carbon sources (1% v/v): dodecane, glucose, glycerol, hexadecane or oleic acid (Fig. 4.7 A and B). Integration could be confirmed with a PCR of p*FALDH*s 1 -3 fused to *lacZ* but not of p*FALDH4*. In each case, 10 transformants from each experiment was tested for integration and promoter activity.

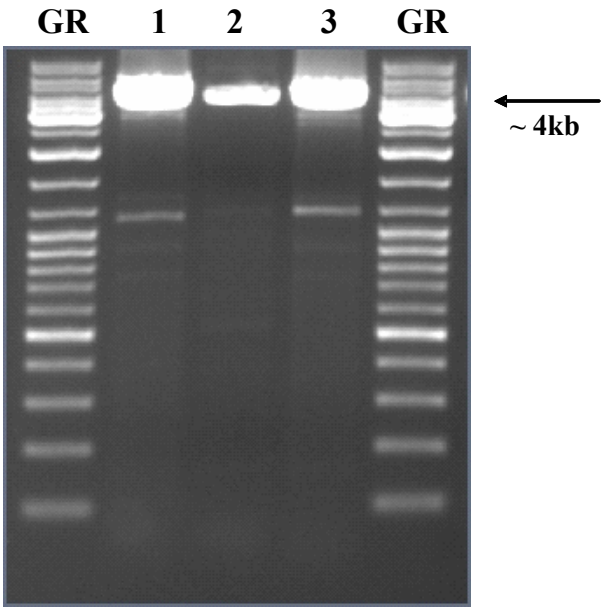


Fig. 4.6 Amplicons of the *FALDH* promoter::*lacZ* fusion to confirm genomic integration in *Y. lipolytica* Po1g. **Lane 1** = *FALDH1* promoter ::*lacZ*, **Lane 2** = *FALDH2* promoter ::*lacZ*, **Lane 3** = *FALDH3* promoter ::*lacZ* and **GR** = O'GeneRuler™ molecular weight marker.

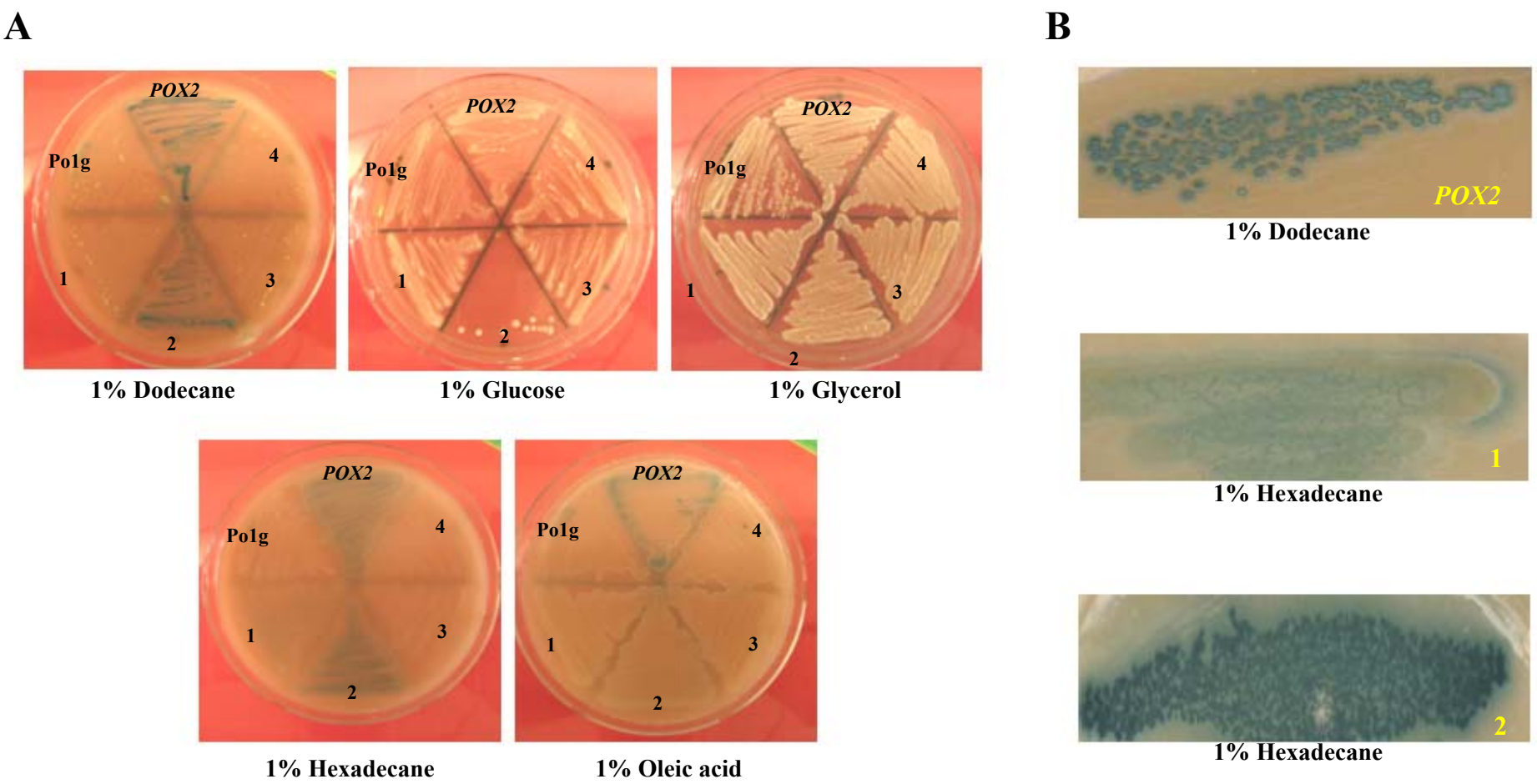


Fig. 4.7 **(A)** Recombinant *Y. lipolytica* Po1g strains containing the integrated pINA781 *FALDH* promoter fusions cultured for 6 days at 28˚C on YNB media containing 40 µg/ml X-gal and 1% (v/v) carbon source. All these strains were streaked from an YPD preculture. **(B)** Recombinant *Y. lipolytica* Po1g strains containing the integrated pINA781 *FALDH* promoter fusions cultured for 3 days at 28˚C on YNB containing 40 µg/mL X-gal and 1% (v/v) carbon source. All these strains were streaked directly from a glycerol stock. The promoter that drives transcription of β-galactosidase in each recombinant strain in **(A)** and **(B)** is abbreviated: **POX2** = *POX*2 promoter, **Po1g** = Negative control, **1** = *FALDH1* promoter, **2** = *FALDH2* promoter, **3** = *FALDH3* promoter and **4** = *FALDH4* promoter.

4.3.3 Monitoring promoter activity as a function of β-galactosidase expression

Recombinant strains of *Y. lipolytica* Po1g, containing the integrated pINA781 vector with the putative p*FALDH*s fused to *lacZ*, were cultured in minimal medium containing various carbon sources as described in Chapter 3, section 3.6.4. Strains were cultured for 60 hours and biomass was harvested at fixed time intervals. These samples were used for: construction of growth curves, a β-galactosidase assay and a total protein concentration determination.

For the β-galactosidase assay, a standard curve of absorbance (405 nm) vs. σ-nitrophenol was constructed (Fig. 4.8) to indicate were the absorbance would deviate from linear kinetics. The β-galactosidase assay was performed as described in Chapter 3, section 3.6.5.For the total protein determination, the Micro BCA™ Protein Assay Kit (Pierce) was used. Bovine Serum Albumin (BSA) standards were prepared and a standard curve of absorbance (504 nm) vs. protein concentration was constructed (Fig. 4.9).

4.3.3.1 Standard curve for β-galactosidase assay

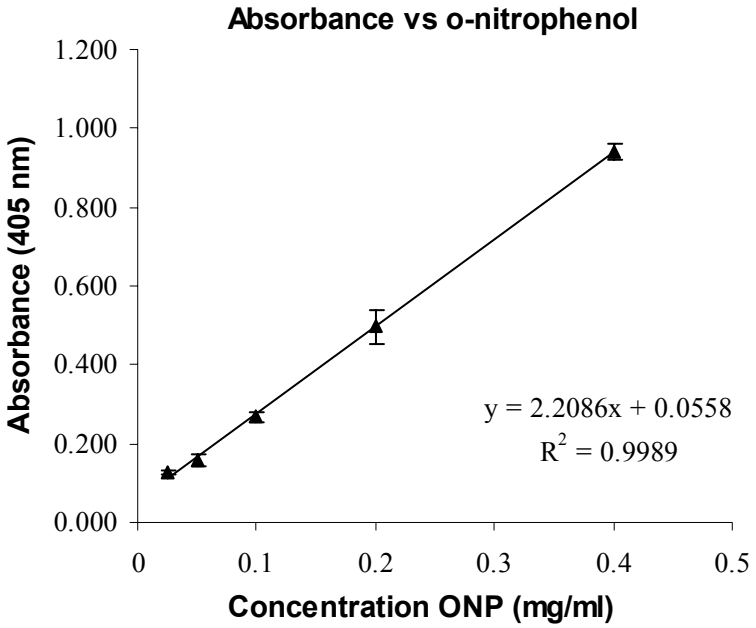


Fig. 4.8 Standard curve of absorbance vs. σ-nitrophenol. Data is representative of standards prepared in triplicate and then averaged. Standard deviations are indicated as bars on the graph.

4.3.3.2 Standard curve for total protein determination

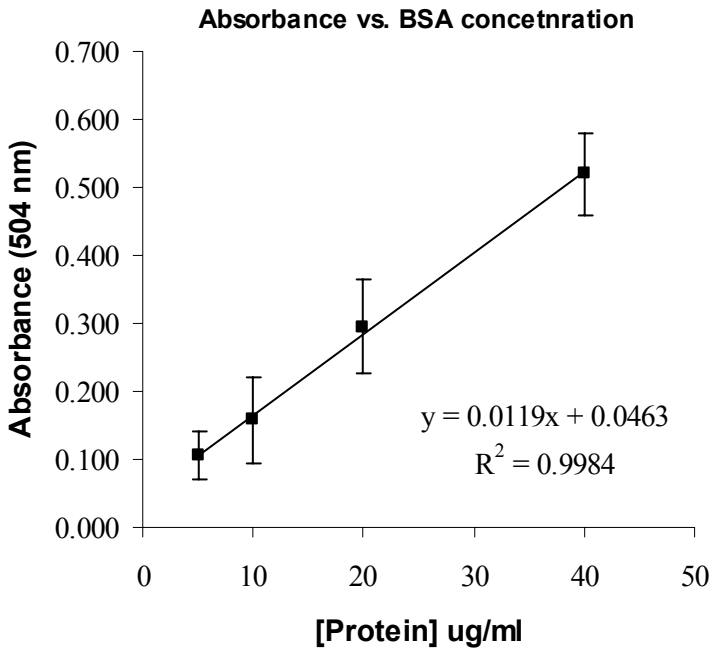


Fig. 4.9 Standard curve of absorbance vs. BSA concentration. Data is representative of averaged triplicate samples and standard deviations are indicted on the graph as bars.

4.4 Growth of recombinant *Y. lipolytica* strains in minimal liquid medium containing various carbon sources to study p*FALDH* expression

Recombinant *Y. lipolytica* strains were cultured in buffered minimal YNB medium containing various carbon sources to study expression of the putative *FALDH* promoters (Fig. 4.10 A and B). Strains cultured in hexadecane, displayed poor growth and exhibited a very long lag phase and this consequently complicated standardization of the experiment. The recombinant *Y. lipolytica* Po1g strains contained the integrated pINA781 expression vector containing the putative p*FALDH1 – 4* that are fused to the *lacZ* gene. Although experiments were prepared and already in progress, a time constraint resulted in this part of the study to be abandoned. Figure 10 represents two such growth studies of *Y. lipolytica* Po1g strains with a positive control (containing p*POX2*) and p*FALDH2*

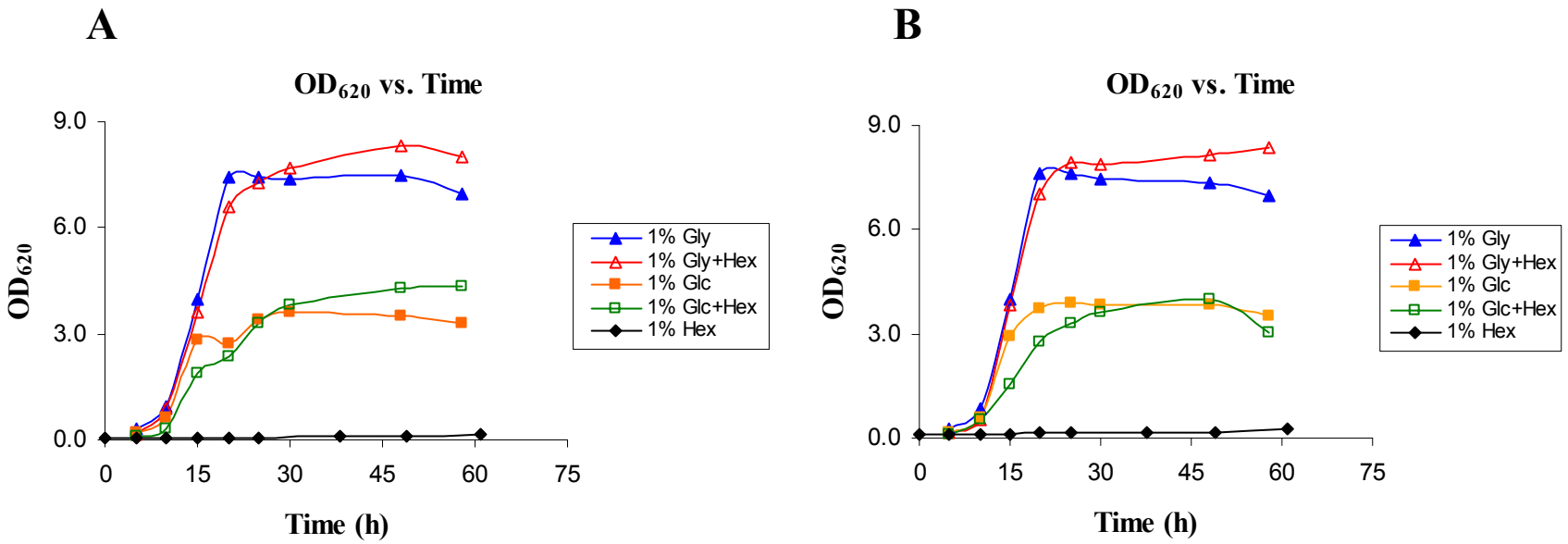


Fig. 4.10 Recombinant *Y. lipolytica* strains cultured in phosphate buffered (pH 6.8) minimal YNB media containing various carbon sources, with agitation for 60 hrs at 28˚C. Recombinant *Y. lipolytica* strains contain either the inducible *POX2* **(A)** or putative *FALDH2* promoter **(B)** both driving transcription of the β-galactosidase gene. Abbreviations in figure legend: **Glc** = Glucose, **Gly** = Glycerol, **Hex** = Hexadecane.

4.5 Culturing triple and quadruple *FALDH* deletion mutants of *Y. lipolytica* Po1d

Triple and quadruple *FALDH* deletion mutants of *Y. lipolytica* Po1d were created by Matatiele (2005) as described by Fickers *et al*., 2003. Strains were precultured in rich YPD broth and the inocula were standardized to an OD_{620} = 1.0. Serial dilutions of the cultures were prepared in YPD_2 broth and the diluted cultures were then spotted onto YPD_2 plates (Smit *et al*., 2004) containing increasing concentrations of dodecanol and its corresponding aldehyde: dodecanal or laurinaldehyde (Fig. 4.11 A and B). This experiment was performed in the hope that the results would provide additional clues as the role of these *FALDH*s in the detoxifying process of fatty alcohols and their aldehydes and to establish their role in their *n*-alkane degradation pathway. Growth was observed for all the strains and the substrates did not seem to exert any toxic effects on the yeast cells.

A

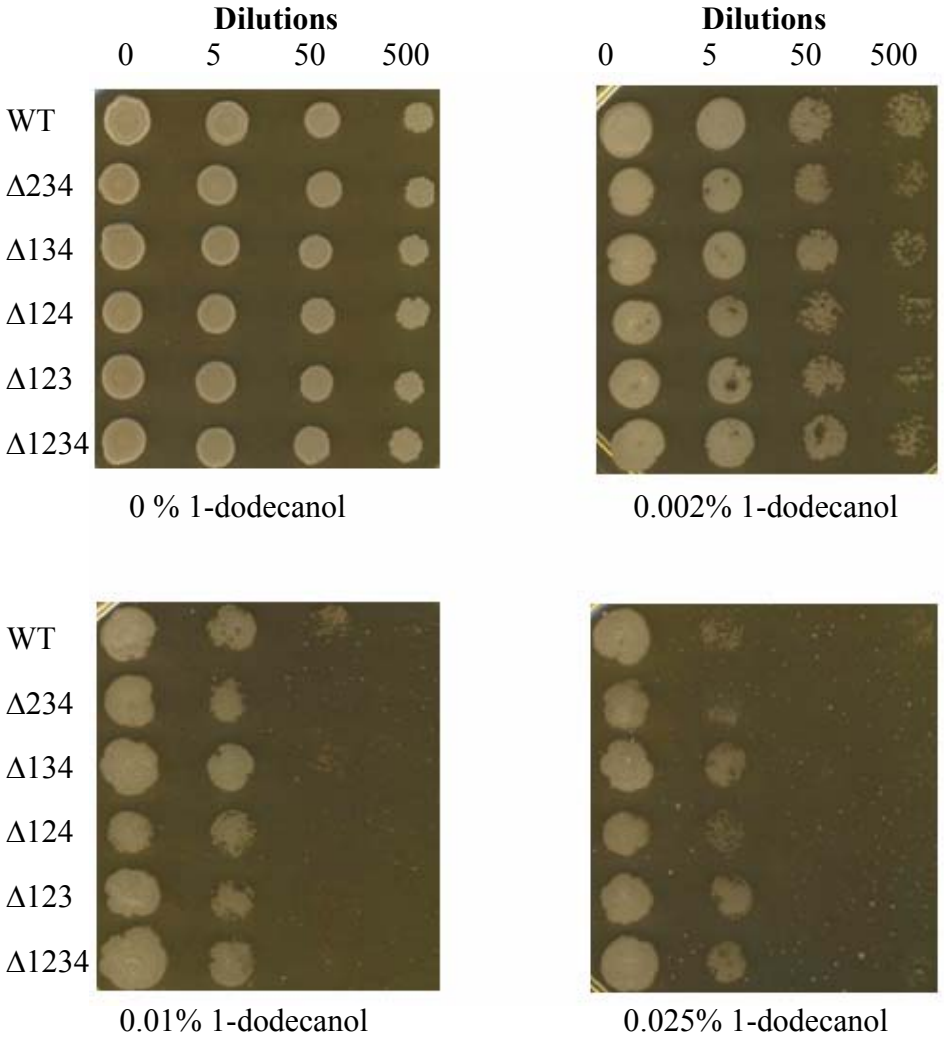


Fig. 4.11 **(A)** *Y. lipolytica* Po1d cell suspensions spotted on YPD_2 media containing 0%, 0.002%, 0.01% and 0.025% 1-dodecanol cultured for 2 days at 28˚C. Abbreviations: **WT** = Wild type *Y. lipolytica* Po1d, **Δ234** = Deletion mutant having only *FALDH1* intact, **Δ134** = Deletion mutant having only *FALDH2* intact, **Δ124** = Deletion mutant having only *FALDH3* intact, **Δ123** = Deletion mutant having only *FALDH4* intact and **Δ1234** = Deletion mutant devoid of any *FALDH*s.

B

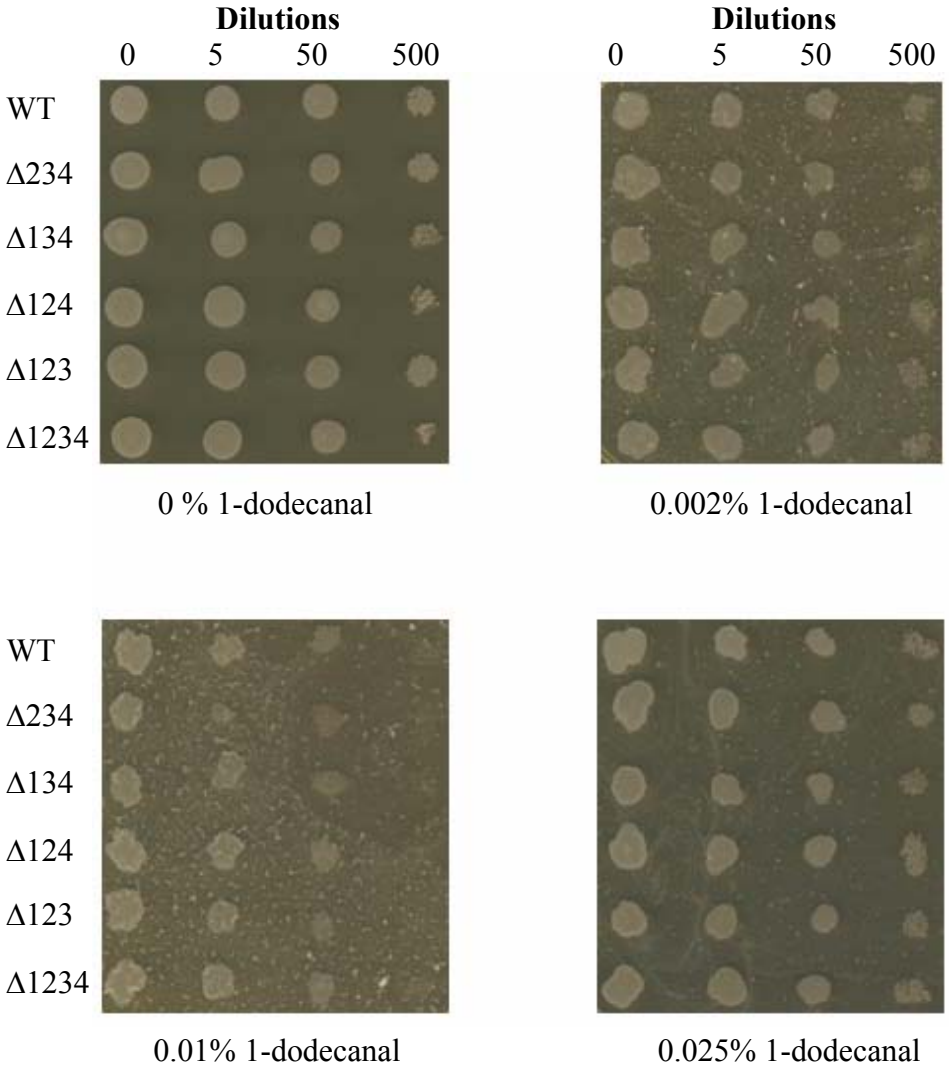


Fig. 4.11 **(B)** *Y. lipolytica* Po1d cells suspensions spotted on YPD_2 media containing 0%, 0.002%, 0.01% and 0.025% 1-dodecanal cultured for 2 days at 28˚C. Abbreviations: **WT** = Wild type *Y. lipolytica* Po1d, Δ**234** = Deletion mutant having only *FALDH1* intact, Δ**134** = Deletion mutant having only *FALDH2* intact, Δ**124** = Deletion mutant having only *FALDH3* intact, Δ**123** = Deletion mutant having only *FALDH4* intact and Δ**1234** = Deletion mutant devoid of any *FALDH*s.

4.6 Attempts at subcellular localization with putative *FALDH* LS

4.6.1 *R. norvegicus* microsomal *FALDH*: the point of reference

Before any cloning experiments could commence, we had to establish whether the *Y. lipolytica FALDH*s possessed NH_2-terminal or COOH-terminal localization sequences(LSs). Available topology software on the internet, including PSORT II, did not detect any NH_2-terminal LSs. This suggested that the *FALDH*s might have COOH-terminal LSs. The *FALDH*s of *Homo sapiens*, *Mus musculus* and *Rattus norvegicus* have well documented COOH-terminal localization sequences that serve as hydrophobic protein anchors in the membrane. In the case of *R. norvegicus*, the last 35 amino acids are responsible for localization and insertion of the *FALDH* protein into the membrane of the ER. Masaki *et al*., 1994 annotated the microsomal *FALDH* of *R*. norvegicus as follows: a stem, hydrophobic domain and a hydrophilic tail. In order to visualize these annotations and to compare the *R. norvegicus FALDH* to the *FALDH*s of *Y. lipolytica* it was decided to construct Kyte and Doolittle hydrophobicity plots by using the ProtScale software (Chapter 3, section 3.7.2).

Each hydrophobicity plot represents the last 35 amino acids of the COOH-terminal of each *FALDH* protein (Fig. 4.12). For comparison purposes each *Y. lipolytica FALDH* was placed next to the *R. norvegicus FALDH* as reference. The stem, hydrophobic domain and hydrophilic tail have been annotated on the *FALDH* of *R. norvegicus*. The *FALDH*s of *Y. lipolytica* have only been divided up by means of dashed lines and have not been annotated as described by Masaki *et al*., (1994). These are merely suggestions by the author by using the *R. norvegicus FALDH* as a model as illustrated in Figure 4.12.

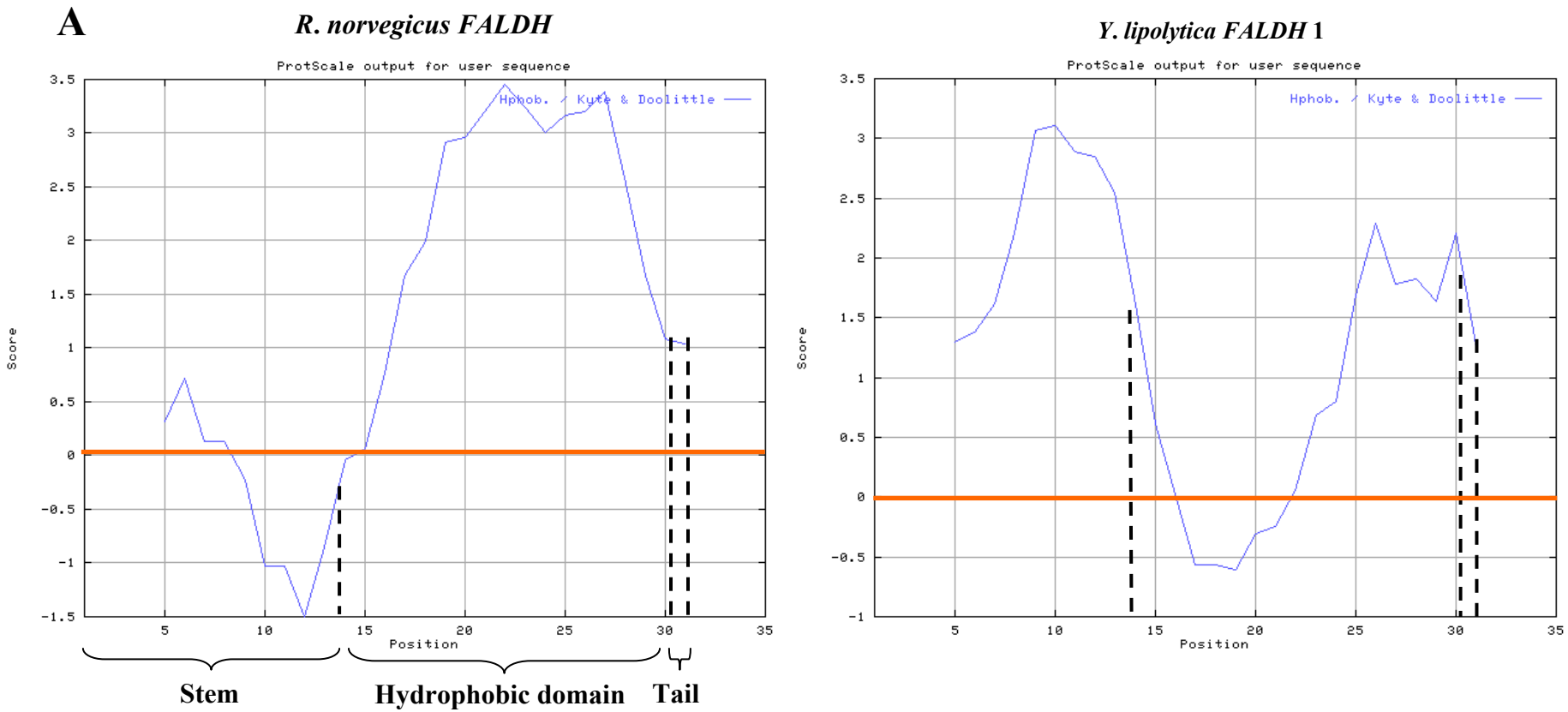


Fig. 4.12 **(A)** Kyte and Doolittle hydrophobicity plots representing the last 35 amino acids of the *R. norvegicus FALDH* and *Y. lipolytica FALDH* 1. Window size: 9 amino acid residues, Relative weight of window edges: 100 % and Weight variation model: linear.

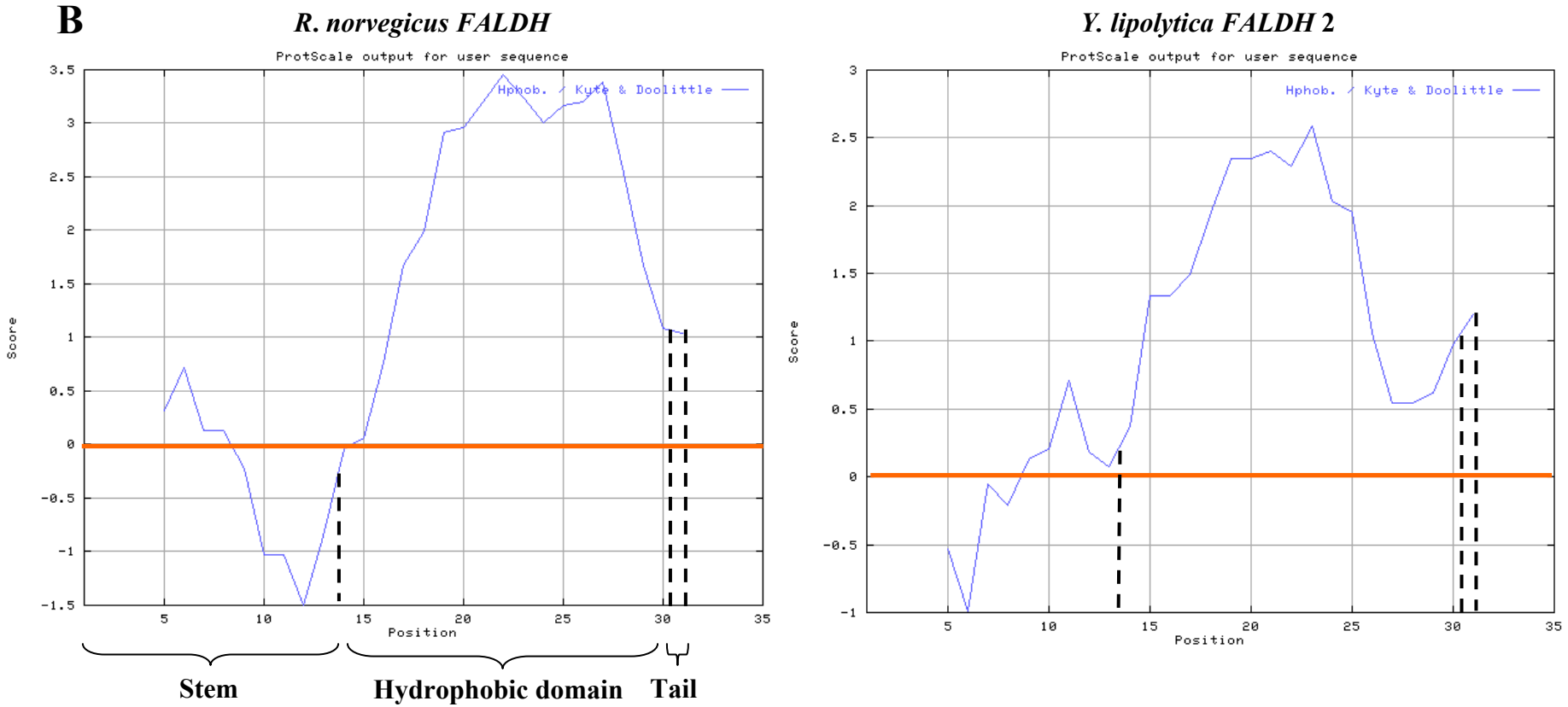


Fig. 4.12 **(B)** Kyte and Doolittle hydrophobicity plots representing the last 35 amino acids of the *R. norvegicus FALDH* and *Y. lipolytica FALDH* 2. Window size: 9 amino acid residues, Relative weight of window edges: 100 % and Weight variation model: linear.

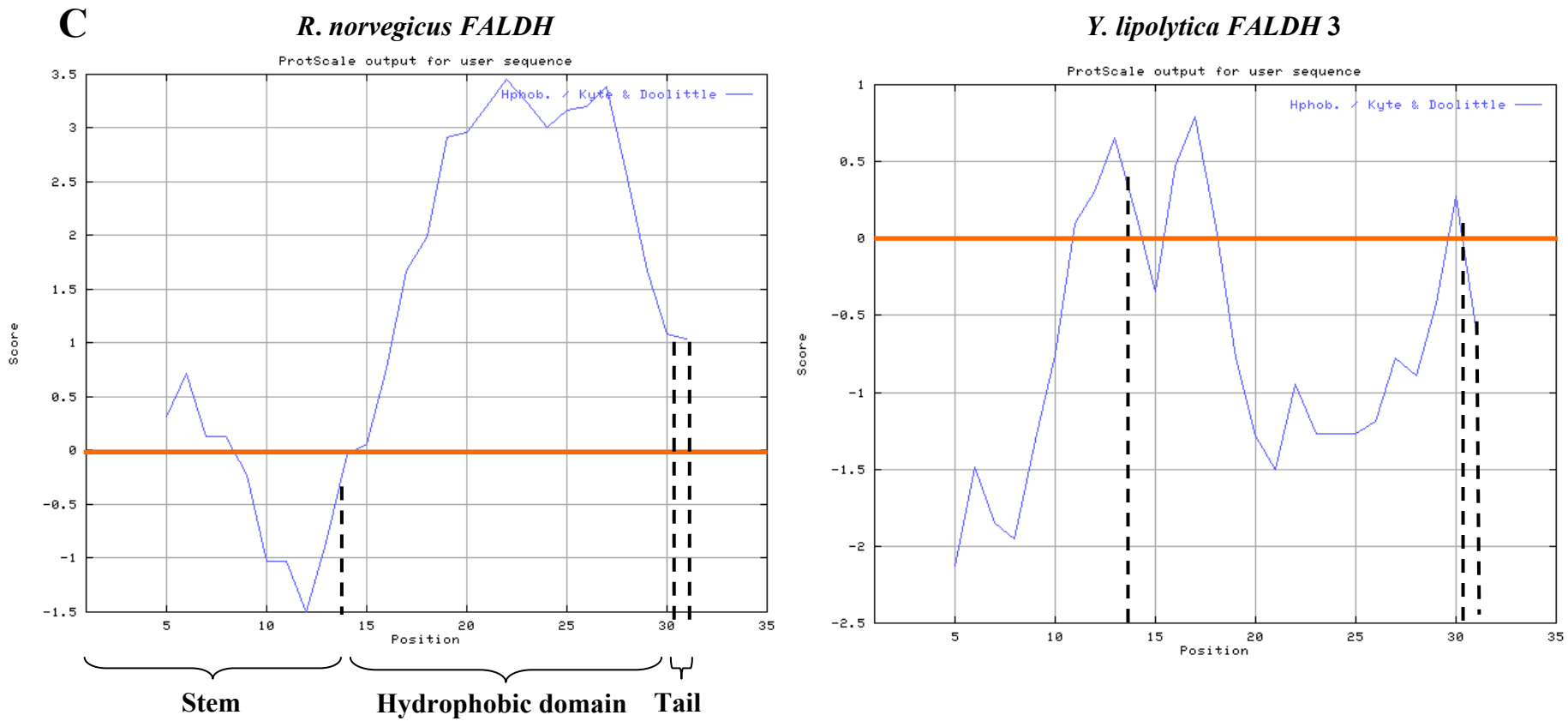


Fig. 4.12 **(C)** Kyte and Doolittle hydrophobicity plots representing the last 35 amino acids of the *R. norvegicus FALDH* and *Y. lipolytica FALDH* 3. Window size: 9 amino acid residues, Relative weight of window edges: 100 % and Weight variation model: linear.

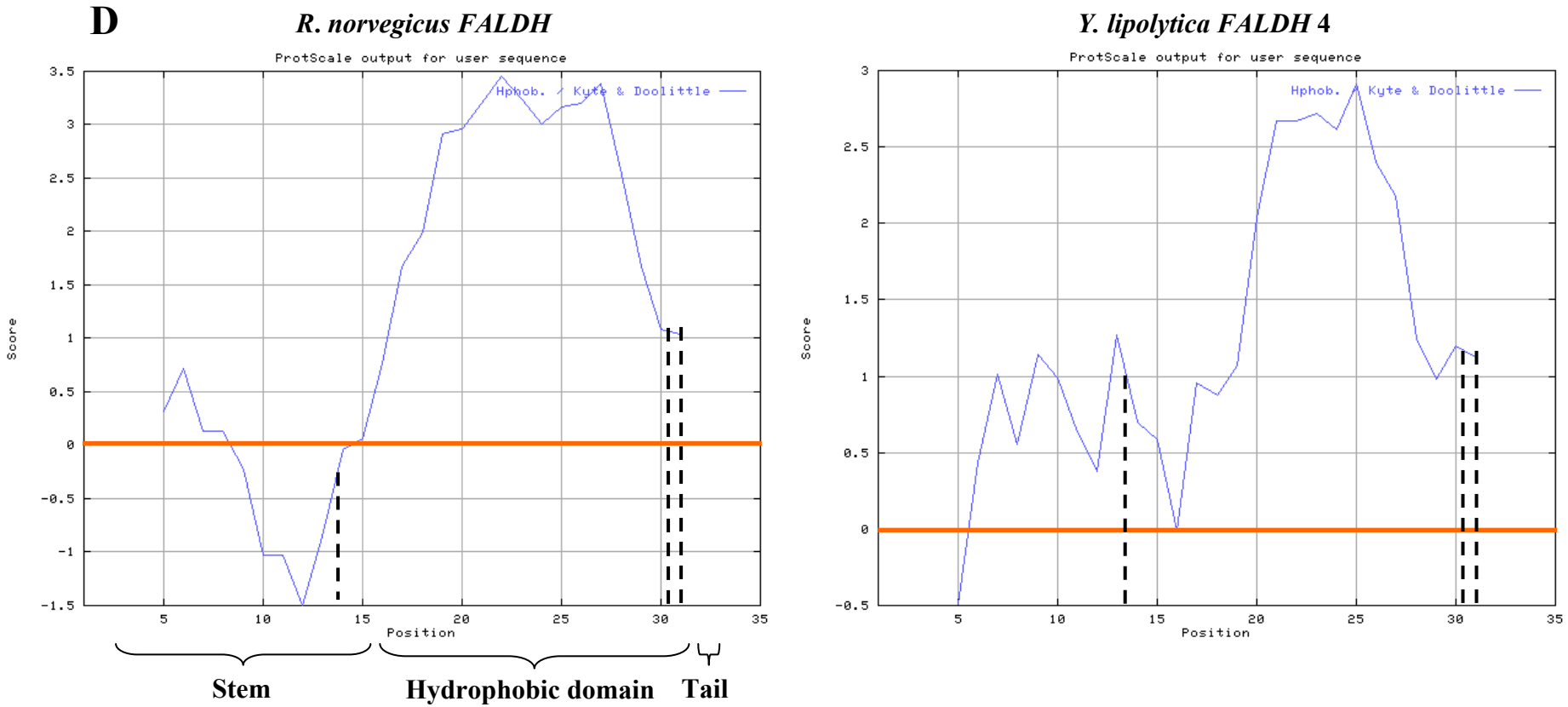


Fig. 4.12 **(D)** Kyte and Doolittle hydrophobicity plots representing the last 35 amino acids of the *R. norvegicus FALDH* and *Y. lipolytica FALDH* 4. Window size: 9 amino acid residues, Relative weight of window edges: 100 % and Weight variation model: linear.

4.6.2 Constructing chimeras of JMP5

To determine localization of the four *FALDH* gene products the integrative vector JMP5 was used as an expression vector. This vector consists of an *ura3d*1 marker flanked by two *zeta* elements that facilitate random integration. The approach was to clone the inducible *ICL1* promoter (p*ICL1*) in frame with a GFP protein and to fuse the putative localization sequence from each *FALDH* (containing a stop codon) to the 3′ end of the GFP gene. Thus, each chimera of JMP5 consists of an inducible p*ICL1*, a GFP from *Aequorea victoria* and a putative *Y. lipolytica FALDH* localization sequence (LS). The chimeras are all identical, except for the fact that each has its own unique *FALDH* LS. For this reason, only one set of results from one such a chimera is illustrated in this section of the chapter. The p*ICL1* was directionally cloned with *Eco*RI and *Bam*HI. The GFP (with an omitted stop codon) was cloned in frame with the p*ICL1* with *Bam*HI and *Mlu*I. Each putative localization signal was directionally cloned to the JMP5 p*ICL1* GFP-fusion with *Mlu*I and *Cla*I. Figure 4.13 illustrates each component used in the construction of the JMP5-chimeras up and to the point where the LSs have to be cloned.

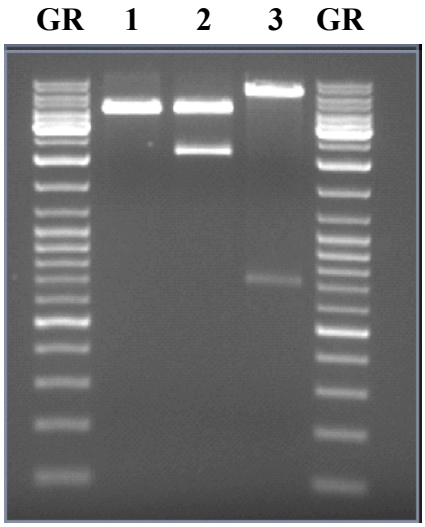


Fig. 4.13 Gel electrophoresis of different components used for constructing JMP5 chimeras. **Lane 1** = JMP5 vector (4 321 bp) linearized with *Eco*RI, **Lane 2** = JMP5 with cloned p*ICL1* digested with *Eco*RI and *Bam*HI to yield the vector backbone (5 019 bp) and the p*ICL1* (2 164 bp), **Lane 3** = JMP5 p*ICL1*

construct with the cloned GFP that was digested with *Bam*HI and *Mlu*I to yield the JMP5 *ICL* portion (6 463 bp) and the GFP gene (720 bp). **GR** = O'GeneRuler™ molecular weight marker.

It was decided to use 150 bp of the COOH-terminal of each *Y. lipolytica FALDH* which translates to a 50 amino acid LS. The putative LSs were amplified from *Y. lipolytica* E150 genomic DNA and assessed on a 2% (w/v) agarose gel (Fig. 4.14). The LSs (containing their native stop codons) were directionally cloned in frame to the GFP with *Mlu*I and *Cla*I. However, before this specific cloning step could commence, the subcloned LS as well as the JMP5 p*ICL1* GFP construct had to be re-amplified in DNA adenine methylase negative (dam^-) *E. coli* since the *Cla*I palindrome was dam sensitive. After transformation to dam^- *E. coli* followed by plasmid purification, clones were digested with *Cla*I to evaluate the dam^- status of the palindrome (Fig. 4.15).

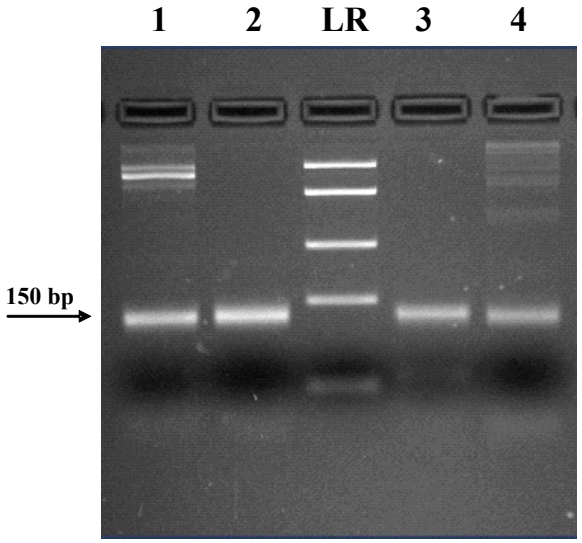


Fig. 4.14 Amplicons of the putative *FALDH* LS from *Y. lipolytica* E150. **Lane 1** = Localization sequence 1, **Lane 2** = Localization signal 2, **LR** = Low Range ladder, **Lane 3** = Localization sequence 3 and **Lane 4** = Localization sequence 4.

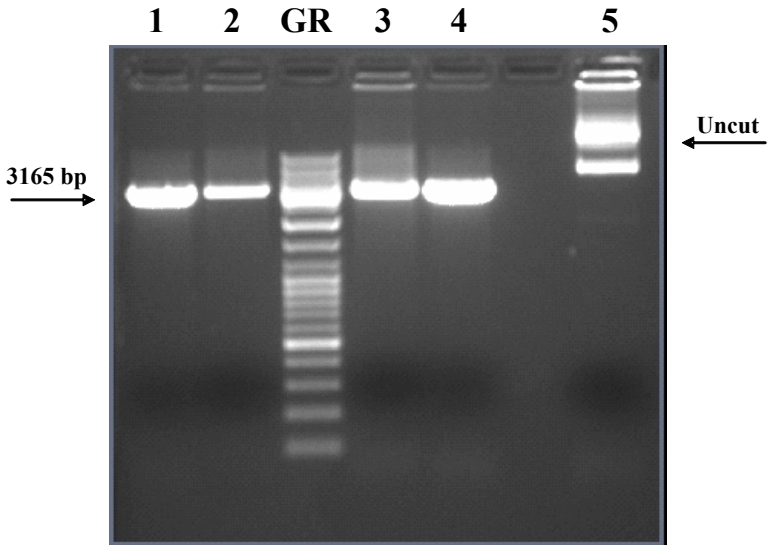


Fig. 4.15 *ClaI* digestion of subcloned LS to confirm abolishment of the methylation on the *Cla*I palindrome. **Lane 1 – 4** = Linearized pGEM®-T Easy with subcloned LS from *FALDH* 3. **Lane 5** = Uncut pGEM®-T Easy vector and **GR** = O'GeneRuler molecular weight marker.

In order to confirm successful subcloning of the putative LS into pGEM®-T Easy, the clones were subjected to nucleotide sequencing (Fig. 4.16). Below is an alignment of the sequenced subcloned LS 3 with that of the known sequence. The 'N' and the'G' (both indicated by an asterisk) were removed after the electropherogram confirmed that the 'N' was a cytosine and the 'G' was not present. This resulted in a 100% alignment. The integrity of the introduced restriction sites, *Cla*I and *Mlu*I were also both verified with the electropherogram.

```
LS 3     1                                                                    0
Seq LS3  1   NCNNNNNNNNNNNNNNNCNATNGGCCGATCNAAGCTCCGGCCCCATGGCGGCCGCGGGAATTCGAT 65
                                        *
LS 3     1           CTAGTTGAAGAGTCTCGACCAAAATCCTCTCACGTCACCCTCCCGGGGAAAACCAGG 57
Seq LS3  66  TGATCGATCTAGTTGAAGAGTNTCGACCAAAATCCTCTCACGTCACCCTCCCGGGGAAAACCAGG 130
               ClaI
                   *
LS 3     58  GCTGGGG-ACCACCATTCTCCGGACAAACTTCTCGTTGCTCTTGTTGTACGGAGGATACCGTTTC 121
Seq LS3  131 GCTGGGGGACCACCATTCTCCGGACAAACTTCTCGTTGCTCTTGTTGTACGGAGGATACCGTTTC 195

LS 3     122 TTCTGCAGCATTTCGACCCATCCGGGTTGTCT                                 153
Seq LS3  196 TTCTGCAGCATTTCGACCCATCCGGGTTGTCTACGCGNCAATCANTAGTGAATTCGCGGCCNCCT 260
                                             MluI
```

Fig. 4.16 Nucleotide alignment of the sequenced, subcloned LS from *FALDH3* in pGEM®-T Easy with that of the known sequence of *FALDH3*. Alignments were performed with DN*Assist* version 2.2. Abbreviations: **LS3** = Known sequence of Localization sequence 3 and **Seq LS3** = Sequenced Localization sequence 3.

After the LSs were cloned into the JMP5 p*ICL1* GFP construct, confirmation of successful cloning was obtained by amplifying the LS from the construct. This PCR-approach was taken since the LSs were not always visible on an agarose gel after being liberated with *Cla*I and *Mlu*I. Figure 4.17 illustrates such an amplification.

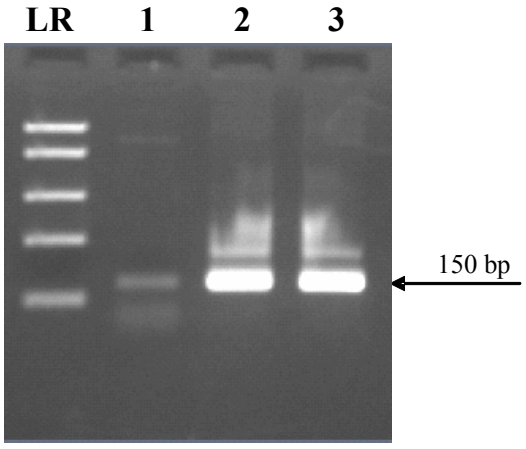


Fig. 4.17 A 2% (w/v) agarose gel illustrating amplification of LS 3 cloned into JMP5 p*ICL1* GFP. **LR** = Low Range molecular weight ladder, **Lanes 1 – 3** = Localization seqeunce 3 amplicons from three independent clones.

The complete chimerical molecule containing the LS from *FALDH3* designated: JMP p*ICL1* GFP LS 3 is illustrated in Figure 4.18. The complete constructs containing the 4 different LSs were linearized with *Not*I to liberate the bacterial moiety from the integrative vector. Consequently only the yeast component of the vector was transformed into *Y. lipolytica* E150 as described by Barth & Gaillardin (1996).

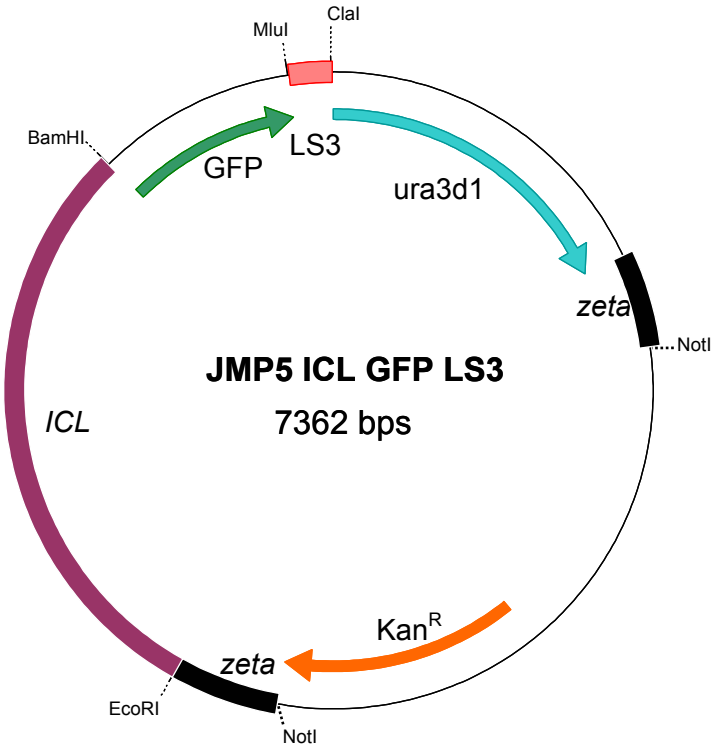


Fig. 4.18 Plasmid map of a successfully constructed JMP5 *ICL* GFP LS3 chimera. The bacterial moiety, containing the kanamycin resistance marker, was liberated by *Not*I and the yeast component was transformed into *Y. lipolytica* E150 where random genomic integration occurred via the *zeta* elements.

4.6.3 Induction of the p*ICL1* and epifluorescence microscopy

Positive clones were cultured on plates with minimal YNB medium and ethanol was provided in liquid (1% v/v) as well as vapour form to serve as inducer. Induction was followed by epifluorescence and confocal laser scanning microscopy by using the filter set that provides light in the UV-range. Unfortunately no fluorescence could be observed

even after inspecting several clones from each chimera. An untransformed *Y. lipolytica* E150 strain served as a negative control. We did however notice light emission that was of a granular nature on the surface of the cells (Fig. 4.19 A and B). We ascribed this emission of light to auto-fluorescence since the negative control also displayed the same trend of light emission.

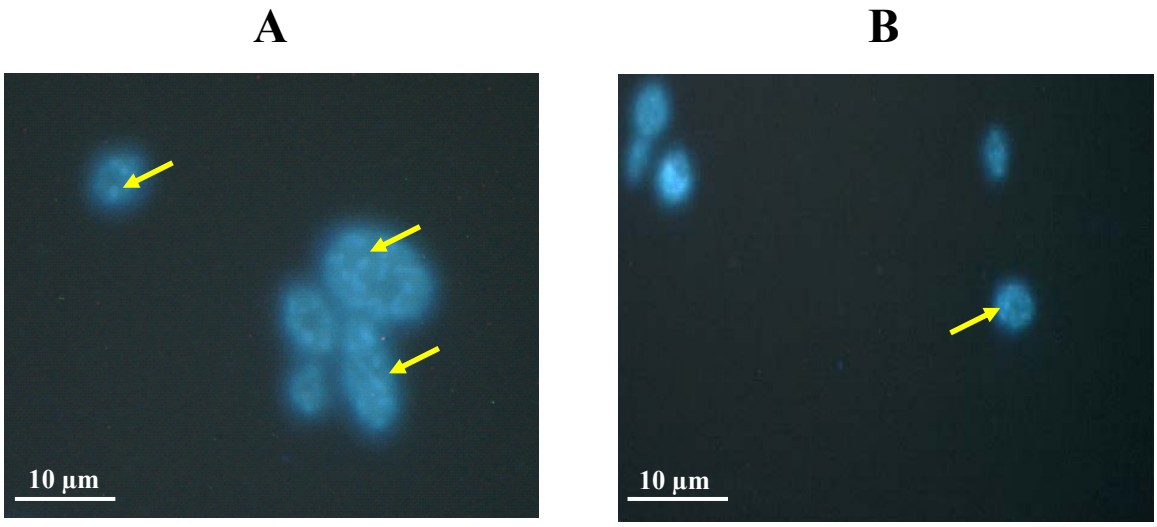


Fig. 4.19 Confocal laser scanning microscopy images of *Y. lipolytica* E150 cells exposed to light of 400 nm. **(A)** Recominant *Y. lipolytica* E150 transformed with the JMP5 p*ICL1* GFP LS3 construct. **(B)** Negative control (untransformed *Y. lipolytica* E150). The granular-like light emission is indicated on the micrograph with yellow arrows.

4.6.4 Using the pKOV136 vector for visualizing fluorescence

The GFP-LS fragment of the JMP5 p*ICL1* GFP LS chimeras was amplified and the restriction sites *Bam*HI and *Acc*65I were introduced to facilitate directional cloning into pKOV136. The GFP-LS fragment was cloned to the 3′-end of the constitutive *TEF* promoter (Fig. 4.19). Positively identified clones were digested with *Not*I to linearize the pKOV136 integration vector. The linearized vector was transformed into *Y. lipolytica* Po1g. Single copy, genomic integration was facilitated by the pBR322 docking system in *Y. lipolytica* Po1g.

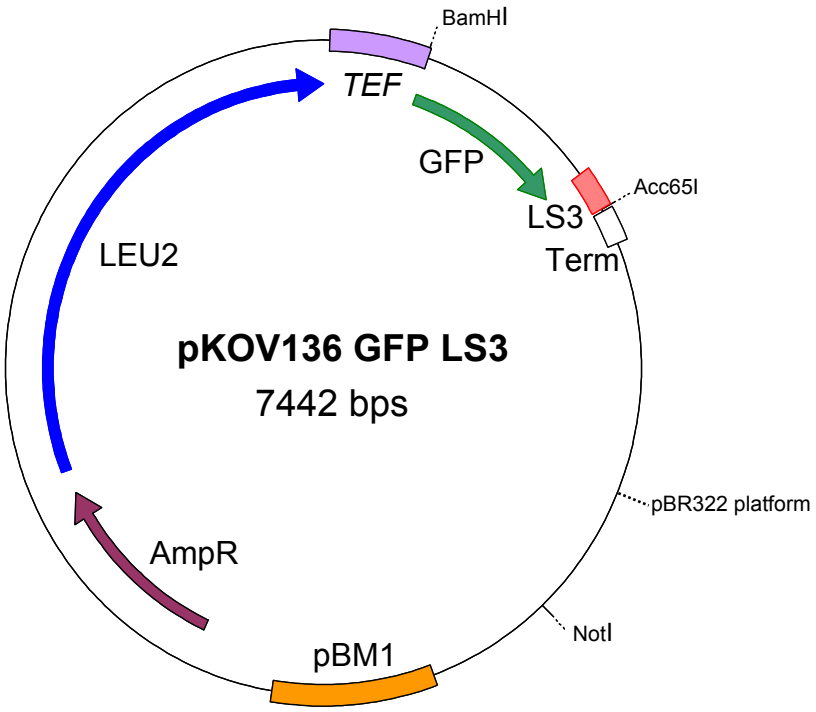


Fig. 4.20 Plasmid map illustrating the pKOV136 vector with the cloned GFP-LS3 fragment from JMP5 *ICL* GFP LS3. This vector comprises a constitutive *TEF* promoter, *LIP*2 terminator and a pBR322 docking platform that facilitates single copy genomic integration in *Y. lipolytica* Po1g once linearized with *Not*I.

Positive clones (*i.e.* leucine prototrophs) were selected and cultivated in rich YPD medium to express the GFP in pKOV136. This was followed was confocal laser scanning microscopy. An untransformed *Y. lipolytica* Po1g was included as a negative control.

An identical trend of granular-like light emission was observed when micrographs were visually inspected (results not shown). This was also attributed to auto-fluorescence.

Chapter 5
Discussion

This chapter will comprise of a discussion with regard to the methods employed in Chapter 3 and the results presented in Chapter 4. Each section of Chapter 4 will be discussed separately with plausible explanations, comments or suggestions. (Note: To ease reading, the *Yarrowia lipolytica FALDH*s and their promoters will be abbreviated as *Yl*FALDH [when referring to the protein] and p*YlFALDH* respectively for the purpose of this discussion).

5.1 Construction of Promoter-*lacZ*-Terminator cassettes for *FALDH* genes

The nucleotide sequences received for the four putative *Yarrowia lipolytica FALDH* isozymes (*YlFALDH*) comprised of the ORF, flanked by a promoter- and terminator region (from this point forward it will be referred to as the 'promoter' and 'terminator'). It was decided to construct promoter-terminator cassettes (PT-cassettes) from each *YlFALDH* and to use these PT-cassettes to construct an expression cassette by cloning a β-galactosidase gene (*lacZ*) gene between the promoter and terminator to yield a Promoter-*lacZ*-Terminator (PlacT) –cassette. The PT-cassettes were constructed with the Sticky-end PCR (SEP) method (Maftahi *et al*., 1996) and the PlacT-cassettes were constructed with the Enzyme-free cloning method (Tillet & Neilan, 1999).

Only the PT-cassettes for *YlFALDH1* and *2* were constructed since the PT-cassettes for *YlFALDH3* and *4* were already constructed in a study of Matatiele (2005). The PT-cassettes constructed in this study were also provided to the study of Matatiele (2005) to aid in the completion of a full set of *YlFALDH* deletion mutants. In exchange for PT-cassettes 1 and 2, all the *YlFALDH* deletion mutants constructed by Dr. P.R. Matatiele, were made available for this study.

5.1.1 Sticky-End Polymerase Chain Reaction (SEP)

Section in question: Chapter 4, section 4.1

The aim of this study was to construct a reporter cassette consisting of a large *FALDH* Promoter region, a reporter gene (*lacZ*) and a large *FALDH* Terminator region (*i.e*. a PlacT-cassette). The SEP method seemed to be less time consuming since it eliminated the need for extra sub-cloning steps and the fact that this method had previously been used by Nicaud *et al*., 1998 and Matatiele (2005) for *Y. lipolytica*, made this method more attractive for this study. Supporting our choice was the successful implementation of this method for rapid gene disruption and marker rescue in *Y. lipolytica* as illustrated by Fickers *et al*., 2003 and Matatiele (2005).

In the study by Maftahi *et al*., 1996, the Promoter-Terminator cassette (PT-cassette) was separated by a 16 bp sequence containing the rare restriction site of *Asc*I. In this study we opted for the meganuclease: I-*Sce* I which had previously been used by Fickers *et al*., 2003 and Matatiele (2005). The 18 bp cutter, I-*Sce* I seemed an obvious choice because of the rare occurrence of its recognition sequence in genomes. According to the manufacturer, an 18 bp recognition site will, statistically, occur once in 6.9×10^{10} bp or once in every 20 human genomes. This rarity of the I-*Sce* I recognition sequence makes it ideal for use in large cloning vehicles requiring unique restriction sites e.g. such as yeast artificial chromosome vectors. Another advantage that influenced our choice was the fact that I-*Sce* I is not a palindrome and this implied no orientation problems when cloning commenced.

PT cassettes were successfully constructed for *YlFALDH1* and *YlFALDH2*. These cassettes were also used by Matatiele (2005) for the construction of disruption cassettes for the deletion of the *YlFALDH* genes. Previously constructed PT cassettes for *YlFALDH3* and *YlFALDH4* were obtained from Dr. Matatiele.

The use of I-*Sce* I to insert *lacZ* into the PT cassettes of all four *YlFALDH* genes did however have limitations. The most important problem was that the method was not reproducible. In Chapter 4 (section 4.1.2, Fig. 4.2) only PlacT-cassettes 1 and 3 could be successfully constructed.

Several workers in our laboratory had noticed that various plasmids, including ones containing the PT-cassettes, re-ligated back onto themselves after I-*Sce* I digestion. Another possibility could also be that the digestion was incomplete or totally unsuccessful, but this seemed highly unlikely since positive controls were always included. Attempts were made to dephosphorylate the digested constructs with calf intestine alkaline phosphatase and to perform blunt-end cloning, but with no success. After digestion the I-*Sce* I recognition sequence yields a deeply buried 5' recessed end. We propose that the deeply tucked away 5′ recessed end makes it difficult for the alkaline phosphatase to remove the phosphate group. This might explain why the PT-cassettes were often lacking the reporter gene after attempts to clone it into the cassette. Since drawbacks occurred at sub-cloning level, it was decided to abandon this protocol entirely and to adopt a new strategy.

5.1.2 Enzyme free cloning: rapidly cloning amplicons independent of vector restriction sites

Section in question: Chapter 4, section 4.2

The previous method only yielded two successfully constructed PlacT-cassettes. We wanted a complete set of PlacT-cassettes and thus experimented with another method. Since our biggest limitation with the previous study was with the restriction enzyme and with the re-ligation of vectors back onto themselves (as stated in section 5.1), we opted for the enzyme-free cloning method as described by Tillet & Neilan, (1999). The method described by Tillet & Neilan, (1999) is a hybrid between the 'hetero-stagger PCR cloning' method (Liu, 1996) and the ligase independent cloning (LIC) procedure (Aslanidis *et al*., 1990). This method had several lucrative advantages of which the most cardinal ones were:

(1) There was no need for multiple sub-cloning steps, (2) no need for post-amplification enzymatic treatments, (3) the method is not constrained by the need for the presence of suitable restriction enzyme sites within the cloning vector and (4) this is a ligase-independent system which would save money and of course time. This method appeared to be superior in comparison with well established methods *e.g*. blunt-end cloning, TA-cloning, LIC and *in vivo* cloning since they all required extensive enzymatic treatment of the amplicon or vector (Aslandidis *et al*., 1990; Liu, 1996).

When this method was employed, the aim was still to create a PlacT-cassette to study the expression of the *lacZ* gene under control of the *YlFALDH* promoters. Tailed PCR primer sets were used on all four the PT-cassettes to create complementary staggered overhangs on both the insert (*lac*Z) and vector (PT-cassette in pGEM®-T Easy) by a post-PCR denaturation-hybridization reaction. For a more comprehensive description of the method see Chapter 3, section 3.5.

No problems were encountered when creating the various amplicons necessary for constructing the PlacT-cassettes (Chapter 4, section 4.2, Fig. 4.3) but problems started to arise when the separate components (*lac*Z and PT-cassettes) were pooled together in equimolar quantities. We experienced a recurring problem with the PT-cassettes since the only result we obtained, after attempting to clone the tailed *lac*Z into the cassettes, were 'empty' cassettes.

We ruled out the possibility that the engineered tails were incorrect since we used the same tails as described by Tillet & Neilan, (1996). The only plausible explanation was that there was carry-over contamination from the template plasmid (PT-cassette in pGEM®-T Easy) that was being transformed into the competent *Escherichia coli* cells. To circumvent this problem we treated our amplicons with *Dpn*I which is a restriction enzyme that recognizes a 4 bp palindromic sequence that is methylated *in vivo*. This enzyme cuts numerous times and since our amplicons were created *in vitro*, only our template plasmid (amplified in *E. coli*) would be methylated and consequently be digested by *Dpn*I. This did however not alter the outcome of our results and the experiment continued to fail.

Although the enzyme-free cloning procedure seems extremely elegant there are also drawbacks: (1) the cost for synthesizing oligonucleotide primers are higher, since long tails have to be added and (2) the increase in primer length may cause the formation of primer dimers (Liu, 1996). This method was also abandoned since the success rate was extremely low.

5.2 Use of pINA781 to monitor expression of the putative *FALDH* promoters

Section in question: Chapter 4, section 4.3

5.2.1 Why did we choose the pINA781 integration vector?

Since both the SEP method and the enzyme-free cloning procedure did not prove successful in our study, we re-strategized and decided to use the integrative pINA781 reporter plasmid (Madzak *et al*., 1999). This pINA781 vector (Chapter 3, section 3.6, Fig. 3.3) was derived from pINA354 (Blanchin-Roland *et al*., 1994) and contains the translational fusion of a minimal leucine promoter and β-galactosidase gene (minimal p*LEU2*::*lac*Z), cloned with *Sph*I-*Bam*HI. This vector possessed two very valuable assets: (1) it already harbored a fully functional β-galactosidase gene, (2) it contained a 430 bp alkaline extracellular protease terminator (*XPR*2t) and (3) it possessed a pBR322 docking platform that enables the plasmid to homologously integrate into the yeast genome as a single-copy at a designated locus (Madzak *et al*., 1999, 2004).

The pINA781 construct is a shuttle vector that is maintained in the *Y. lipolytica* cell by means of homologous integration into the yeast chromosome. These specific vectors yield high transformation frequencies when linearized within the homology region *e.g*. pBR322 docking platform linearized with *Not*I. In more than 80% of the cases, a single complete copy of the vector will be integrated at the chosen site (Barth & Gaillardin, 1996). Integrative vectors are also extremely stable and do not require the maintenance of selective pressure as with episomal vectors. It has been illustrated by Hamsa & Chattoo, (1994) that even after 100 generations under non-selective conditions; the vectors were retained without any rearrangements.

It thus seemed appropriate for us to use this integrative vector in our expression studies when one takes into account the advantages this specific shuttle vector holds. In particular, the site specific single-copy integration seemed advantageous to our study since it allows for comparisons of different promoters controlling expression of the same reporter gene.

5.2.2 Cloning the putative *YlFALDH* promoters in frame to *lacZ*

The *YlFALDH* promoters (*ca*. 1 kb) were amplified (Chapter 4, section 4.3, Fig. 4.4) and fused with the *lacZ* in pINA781 by means of directional cloning with *Sph*I and *Bam*HI (Chapter 4, section 4.3, Fig. 4.5). When the minimal p*LEU2* was excised from pINA781, the ATG from *lacZ* was also removed (Fig. 5.1). Consequently, the reverse primers of each *YlFALDH* promoter region did not only possess a *Bam*HI restriction site but also a synthetic ATG that was engineered to be in frame with *lacZ*. When fusion of the *YlFALDH* promoter with *lacZ* was complete, the introduced ATG was next to the 10th codon (GTC) of the *lacZ* gene.

*Sph*I *Bcl*I
GCATGCACTGATCACGGGCAAAAGTGCG**TATATATACAA**GACGCTTTGCCAGCCAGCAGA

TTTTCACTCCACACACCAACATCACACATACAACCACACACATCCACA **ATG** GAA CCC

GAA ACT AAG GAT CCC GTC GTT TTA-→ *lacZ*
*Bam*HI 10 11 12

Fig. 5.1 Nucleotide sequence of pINA781 annotated with the *Sph*I and *Bam*HI sites that were used for cloning of the *FALDH* promoter regions. The initiation codon of *lacZ* (present in pINA781) is highlighted in green while the TATA-box of the minimal p*LEU2* is highlighted in pink. The first original codon from *lacZ* (codon 10) has also been annotated. (Adapted from: Madzak *et al*., 1999).

5.2.3 Plate assays: a preliminary screening tool

Qualitative plate assays were performed with recombinant *Y. lipolytica* Po1g strains containing the pINA781 vector harboring a *YlFALDH* promoter (*pYlFALDH*) that is cloned to the 5′-end of a *lacZ* gene. The pINA781 integrative plasmid's integration into the genome of the yeast was confirmed for recombinant strains containing p*YlFALDH1-3* (Chapter 4, section 4.3.2., Fig. 4.6). The integration of p*YlFALDH4* could not be confirmed.

Plate assays made a valuable contribution to this study and the power of such a simple method should never be underestimated. The plate assay used in this study was only for qualitative purposes and one should always bear in mind that certain kinetics will differ when culturing the same strains on solid media *vs*. when cultured in liquid media with agitation. Factors like oxygen transfer, homogeneity of nutrient concentrations in the media, inocula size and time required to generate biomass are just some of the factors that will differ between the two culturing states.

The plates in this study were used to monitor a colour reaction. If the recombinant *Y. lipolytica* Po1g strains did indeed harbor a fully integrated and functional pINA781 p*YlFALDH*-construct, then the putative *YlFALDH* promoter region would drive transcription of the *lacZ* gene. If this was the case, the colonies would turn blue if 5-bromo-4-chloro-3-indolyl- β-D-galactoside (X-gal) was provided in the medium. The medium used was buffered with 50mM phosphate buffer but the pH was elevated to 7.4 and not adjusted to the normal pH 6.8 (Nicaud *et al*., 1998; Juretzek *et al*., 2000). This was necessary for the colour reaction to occur properly.

5.2.4 Are the promoters inducible or constitutive?

After a fixed incubation period we observed that only recombinant strains with the putative promoters of *YlFALDH1* and *YlFALDH2* driving transcription of *lacZ* displayed a blue colour (Chapter 4, section 4.3, Fig. 4.7 A and B). In the case of p*YlFALDH2, lacZ* expression was strong and repeatable, but in the case of p*YlFALDH1* only a faint blue colour could be observed when cultures were streaked directly from frozen glycerol stocks. A preliminary

conclusion was made that these two putative promoters are induced by *n*-alkanes, since they only turned blue in the presence of dodecane (C_{12}) and hexadecane (C_{16}) as sole carbon source. Induction of *lacZ* in the different strains was also tested on plates containing *n*-alkanes and glucose or *n*-alkanes and glycerol (results not shown). Transformants with p*YlFALDH1* and p*YlFALDH2* also turned blue but only at a much later stage – most probably when the glucose and glycerol were depleted. It is thus possible that glucose and glycerol may repress these promoters, but studies will have to be performed in liquid media to corroborate this statement. Judging from the intensity of the blue colour, we made the conclusion that the putative p*YlFALDH*2 drives transcription of *lac*Z more strongly than the p*YlFALDH*1 (Chapter 4, section, 4.3, Fig. 4.7 B).

Color development on X-gal containing agar plates has previously been used to qualitatively estimate and compare promoter strengths on various carbon sources (Müller *et al*., 1998 and Juretzek *et al*., 2000). The positive control in this experiment was a recombinant *Y. lipolytica* Po1d strain containing the *POX2* promoter (p*POX2*) driving transcription of the *lac*Z gene (Juretzek *et al*., 2000). In accordance with literature the strains containing the p*POX2* exhibited induction on dodecane, hexadecane and oleic acid with colonies turning blue on media containing these carbon sources (Chapter 4, section 4.3., Fig. 4.7 A). p*POX2* is considered a strong inducible promoter (Juretzek *et al*., 2000; Madzak *et al*., 2004). It is therefore of interest that the transformant containing p*YlFALDH2* apparently displayed the same levels of *lacZ* expression as the p*POX2* strain on dodecane and hexadecane. However, p*YlFALDH2* was apparently not induced by oleic acid. No blue colour appeared when the p*POX2* strain was cultured on glucose and glycerol, although according to Juretzek *et al*., (2000), glucose and glycerol do not repress the p*POX2* completely.

It is interesting to note that induction of p*YlFALDH1*, p*YlFALDH2* and p*POX2* was apparently different when strains were pre-cultured on YPD plates than when strains were streaked out directly from glycerol stocks (Chapter 4, section 4.3, Fig. 4.7 A and B). Pre-culturing on YPD plates delayed growth and induction of the promoters on *n*-alkane containing YNB plates, with the blue color becoming visible within three days after streaking directly from frozen stocks but only after six days when streaked from YPD plates.

The transformants containing the putative *FALDH3* and *FALDH4* promoters did not display any β-galactosidase expression (*i.e*. cultures did not turn blue). Genomic integration of the p*YlFALDH3-lacZ* fusion could be confirmed by PCR, but not integration of p*YlFALDH4-lacZ* (Chapter 3, section 3.6.3), In the case of the p*YlFALDH4-lacZ* fusion transformations were repeated several times as well as genomic DNA extractions and PCR. The possibility of defective primer pairs were ruled out since these primers amplified the correct amplicon from the vector when it was used as template. The only plausible explanation is that there is simply no genomic integration. However, it is then difficult to explain how the recombinant strains became leucine prototrophs and were able to grow well when selecting on minimal medium lacking leucine.

Matatiele (2005) had investigated induction of *YlFALDH3* and *YlFALDH4* with Northern hybridizations. These experiments had indicated weak induction of *YlFALDH4* by hexadecane and hexadecane with pristane, but no induction of *YlFALDH3* was observed, even when H222 cells were cultured on C_{12}, C_{16} and C_{22} *n*-alkanes. It should be noted that these Northern hybridizations were carried out with total RNA isolated from *Y. lipolytica* H222 wild-type cells and not from the W29 derivative: *Y. lipolytica* Po1g. Perhaps it is premature to conclude that *lacZ* expression can not be induced in the transformants with the p*YlFALDH4*. Studies should rather be performed in liquid media with a sensitive, quantitative assay procedure. It was however possible to conclude from the plate assays that none of the p*YlFALDH*s are strong constitutive promoters and that p*YlFALDH1* and p*YlFALDH2* are induced by *n*-alkanes, but not by oleic acid, glucose or glycerol.

5.2.5 Expression studies in liquid minimal media

In order for us to study the promoter expression properly, the recombinant Po1g strains were cultured in buffered YNB media as described in Chapter 3, section 3.6.4. The biggest obstacle we faced was that the recombinant strains experienced extremely long lag phases (up to 30 hours) when cultured in *n*-alkanes as sole carbon source. This hindered comparison between the *YlFALDH* promoters, since all recombinant strains needed to be cultured under exactly the same conditions. For example after culturing the Po1g strain containing p*YlFALDH2* for 30

hours in hexadecane with glycerol, the culture is already in stationary phase with OD_{620} = 7.9 (Chapter 4, Fig. 4.10 B). When comparing the same recombinant strain in hexadecane media after 30 hours, the culture has an OD_{620} < 0.1 and is still in the lag phase (Chapter 4, Fig.4.10 A). Recombinant strains performed better when glucose or glycerol was used in conjunction with the *n*-alkanes. This seemed reasonable since the strains preferentially utilized the glucose and the glycerol before assimilating the *n*-alkanes.

In addition to the long lag phase, poorer biomass yields were also observed. Mauersberger *et al*., (2001) noted that W29 exhibited a very long lag phase or even no growth at all when replica plated from glucose to *n*-alkane media. The W29 strain also displayed a very long lag phase (30 – 35 hours) when cultured in liquid medium supplemented with *n*-alkanes ranging in carbon length from C_{10} to C_{16}. In addition, the authors also noted that amino acid-auxotrophic strains (Leu^-, Met^-, Lys^- and His^-) exhibited a reduced growth rate on hydrocarbons in minimal medium when ammonium salts were used as nitrogen source. It thus became evident that studies involving β-galactosidase assays of cell free extracts from cultures grown in *n*-alkane containing liquid media will require time consuming optimization. Although β-galactosidase and protein assays were set up the idea to test β-galactosidase activity of cell free extracts from liquid cultures had to be abandoned due to lack of time.

5.2.6 Why promoter-*lacZ* fusions might fail to detect induction of a promoter

In the case of the p*YlFALDH3* both Northern hybridizations (Matatiele, 2005) as well as *lacZ* fusions had failed to detect induction, while in the case of p*YlFALDH4* Northern hybridizations (Matatiele, 2005) and *lacZ* fusions apparently gave opposite results. It should however be kept in mind that the reporter gene approach might miss or underestimate induction of a promoter. We propose the following arguments for the obtained results of p*YlFALDH3* and *4*.

Eukaryotic promoters require multiple sites for proteins to be associated with the functioning of even the smallest *cis*-acting elements (Olesen & Guarente, 1990). In yeast cells, upstream activating sequences (UASs) are necessary for successful transcription. The UASs function in either orientation and at a variable distance with respect to the TATA-box and transcription

start site. Transcriptional activators target the UASs when transcription commences and the inactivation or absence of these activators results in most of the repression phenomena observed in yeasts. In addition to this, the presence of negative regulatory sites, or upstream repression sequences (URSs) have also been identified and noted for repressing transcription (Nishizawa *et al*., 1989; Nehlin *et al*., 1991).

The inducible alkaline extracellular protease promoter (p*XPR*2) from *Y. lipolytica* is the best studied promoter (Blanchin-Roland *et al*., 1994; Madzak *et al*., 1999) and probably the most extensively employed (Nicaud *et al*., 1989; Müller *et al*., 1998). Functional dissection of the p*XPR2* revealed that the TATA-box and two major UASs were essential for promoter activity under conditions of repression or full induction (Blanchin-Roland *et al*., 1994). Deletion of both the distal UAS1 and proximal UAS2 greatly reduced *XPR2* expression without abolishing its regulation. The authors suggested that these UASs are targets for transcriptional factors required for assisting specific regulatory proteins.

The UAS1 was found to be poorly affected by environmental conditions and was consequently used to design a hybrid promoter, composed of four tandem UAS1 copies upstream from a minimal p*LEU2* that was reduced to its TATA-box (Madzak *et al*., 1999). This recombinant promoter, designated as hp4d, was found to be almost independent from environmental conditions and was able to drive a high quasi-constitutive expression in virtually any medium. The hp4d promoter was not repressed by carbon and nitrogen sources, did not require the addition of peptones in the culture medium and was insensitive to extracellular pH. Some of the best results obtained with the hp4d promoter, in terms of production, using mono-copy vectors were 160 mg/mL for active prochymosin which was several fold higher than obtained with previously available *Y. lipolytica* promoters (Madzak *et al*., 2000, 2004). The above statements and results illustrate just how important and potentially powerful UASs are in driving transcription.

It has to be stated that no functional analyses were performed on any of the *YlFALDH* promoters. So the promoters have not been dissected or genetically altered in any way. The promoter regions (*ca*. 1 kb) were only amplified and cloned in front of the *lacZ* reporter gene.

The chosen size of the promoter region was arbitrary and 1 kb is not considered that large for a promoter – the inducible isocitrate lyase 1 promoter (p*ICL1*) is *ca*. 2.1 kb in size. Since the *YlFALDH* promoter region 3 was not functionally dissected, we propose that the UAS might not have been included in the nucleotide sequence we obtained. As noted in section 5.2.6, if no UASs are available there will be no transcription. Alternatively, we propose that possible URSs may also play a role in repressing p*YlFALDH3* and *4*. It is once again recommended that promoter studies be performed in liquid media in conjunction with a sensitive, quantitative assay.

5.3 GFP as a fluorescent source to determine subcellular localization

The main factors contributing to our choice to utilize GFP as a reporter gene in this study was firstly: we had a pYES2 shuttle vector containing the gene encoding the GFP protein. This vector was a gift from Prof. G. Lindsey (University of Cape Town, South Africa) and was used in his laboratory for localization studies in *Saccharomyces cerevisiae*. Secondly, we had an epifluorescent as well as a confocal laser scanning microscope at our disposal to study the fluorescence.

5.3.1 Why is GFP a popular reporter system?

GFP is an extrememly stable protein (238 amino acids) from the luminescent jellyfish, *Aequorea victoria* (Prasher *et al*., 1992). The fluorescent properties are unaffected by prolonged treatment of with 6 M guanidine HCl, 8 M urea or 1% SDS. Even after a two day treatment with various proteases such as trypsin, chymotrypsin, papain, subtilisin, thermolysin and pancreatin at concentrations up to 1 mg/ml fail to alter the intensity of GFP fluorescence (Bokman & Ward, 1981). GFP is stable in neutral buffers up to 65˚C and displays a broad pH stability from 5.5 – 12. This fluorescent protein has been produced in a number of heterologous cell types and there appears to be little requirement for specific additional factors for post-translational modification of the protein. GFP has been successfully used in localization studies for transgenetic plants (von Arnim *et al*., 1998; Davis & Vierstra, 1998), mammalian cell lines (Masaki *et al*., 2003) and in yeast cells, including *Y. lipolytica* (Du *et al*., 1999; Geraghty *et al*.,

1999; Koller *et al*., 1999; Hurtado & Rachubinski, 2002; Augstein *et al*., 2003; Parshyna & Barth, 2003 and Masaki *et al*., 2003).

The principle advantage of GFP reporter systems is the ability to detect fluorescence in living specimens with real-time kinetics. Consequently, this has established GFP as a powerful reporter for the analysis of gene expression and protein localization in a wide variety of experimental designs (Yang *et al*., 1998). The introduction of *A. victoria* GFP was a breakthrough in the studies on the localization of proteins and organelles in a large number of species, including yeasts.

The GFP-SKL (SKL = last three carboxy-terminal amino acids that direct proteins to the peroxisomes) fusion is widely used as a fluorescent tag of peroxisomes and has proven to be a valuable biological research tool (Nazarko *et al*., 2005). It thus did seem as if GFP was an adequate candidate for localization studies in *Y. lipolytica*. Quite recently, GFP was employed in a study by Parshyna & Barth, 2003 in which they studied the detection of biosynthesis and degradation of peroxisomes by using a β-galactosidase-GFP-SKL chimeric protein under control of the p*ICL1*.

5.3.2 Prediction of subcellular localization of *YlFALDH* gene products

Section in question: Chapter 4, section 4.4

5.3.2.1 Putative localization sequence from the *YlFALDH*s

In order for us to study the subcellular localization of the *YlFALDH1 - 4*, we had to utilize the native localization sequence (LS) from each isozyme. Online software (*e.g*. PSORT version 2.0) did not detect any NH_2-termimal LS or potential peptidase cleavage sites in any of the four isozymes. It was also observed that all the available software is designed to look for only NH_2-terminal LS. So our reasoning was that if no NH_2-termimal LSs were detected, it probably meant that these isozymes may very well have COOH-terminal LSs. In support of this, literature was obtained stating that the last 35 amino acids of *FALDH*s at the COOH-terminal

were responsible for localizing the proteins. In addition the last 35 amino acids of these COOH-terminal LSs also serve as a hydrophobic protein anchor (Masaki *et al*., 1994; Rizzo *et al*., 2001; Vasiliou *et al*., 2001; Masaki *et al*., 2003; Demozay *et al*., 2004). This has been well documented for the *FALDH*s of *Homo sapiens* (Rizzo *et al*., 2001), *Mus musculus* (Rizzo *et al*., 2001) and *Rattus norvegicus* (Masaki *et al*., 1994). The fact that no NH_2-terminal LSs were predicted and that *FALDH* COOH-terminal LS were so well documented supported our theory that the four *YlFALDH* isozymes of *Y. lipolytica* might possess similar LS at their COOH-termini.

As a starting point, hydrophobicity plots of the last 35 amino acids of each *YlFALDH* were constructed. The last 35 amino acids of the *R. norvegicus FALDH* (*RnFALDH*) served as reference (Chapter 4, section 4.6, Fig. 4.12 A – D). Masaki *et al*., 1994 divided the *R. norvegicus* 35mer into three sections: a stem, hydrophobic domain and a hydrophilic tail. In Figure 4.12 (A – D) the *R. norvegicus* 35mer has been annotated according the Masaki *et al*., 1994. Upon visual inspection of *YlFALDH2* and *4*, similar trends are observed in comparison with the *RnFALDH*. The trend that the hydrophobic domain follows in the reference as well as the *YlFALDH*s is very striking. Although *YlFALDH1* and *3* do not follow a similar trend, this does not unequivocally prove that these *YlFALDH*s do not have COOH-terminal LSs.

In an attempt to determine the localization of the *YlFALDH* proteins it was decided to clone the last 150 bp of the COOH-terminal *YlFALDH*s (Chapter 4, section 4.6.2, Fig. 4.14) to the 3′ end of the GFP followed by integration of the completed construct into *Y. lipolytica*. In addition there was a need to allow expression of the FALDH::GFP fusion and this was accomplished with the use of p*ICL1* since we had a construct with such a promoter available for subcloning purposes (Nthangeni *et al*., 2004). The JMP5 integrative vector was used for localization studies (Wang *et al*., 1999; Pignéde *et al*., 2000) since it had no promoter and this provided the opportunity to clone the promoter of choice into the construct. In addition the JMP5 vector (Chapter 3, section 3.8, Fig. 3.5) possessed *zeta* elements that provided at least 100 potential integration sites per genome when using *Y. lipolytica* strains like E150 (Madzak *et al*., 2004). Zeta-based auto-cloning vectors have been used successfully to overexpress *Y. lipolytica* extracellular lipase (Pignéde *et al*., 2000) and to express heterologous genes (Juretzek *et al*.,

2001). It is interesting to note that *zeta* sequences are able to enhance non-homologous integration into *Y. lipolytica* strains devoid of the Ylt1 retrotransposon. This feature provided a non-homologous process of transformation for *Y. lipolytica*. This method can be used particularly in the Po1d strain and its improved derivatives like Po1f and Po1h (Madzak *et al*., 2004).

5.3.3 Cloning the p*ICL1* into the promoter lacking JMP5 vector

The p*ICL1* is a well regulated promoter and drives transcription of the isocitrate lyase of the anaplerotic glyoxylate cycle. It is induced when carbon sources such *n*-alkanes, fatty acids, ethanol and acetate are used. The promoter is repressed to a basal level when cells are grown in the presence of glucose (Juretzek *et al*., 2000).

Before cloning the p*ICL1* we noticed that there was an intron that was part of the promoter. It has been reported by Juretzek and co-workers (2000) that the intron had no effect on expression levels when a p*ICL1* was fused to the 5′-end of the *E. coli lac*Z gene. The intron also had no effect on expression levels of the *Y. lipolytica* cytochrome P450 reductase CPR gene (Nthangeni *et al*., 2004). At the time of this dissertation's completion, no literature could be obtained that stated the contrary result with regard to the omitting of the intron from the p*ICL1*. The restriction enzymes that we selected (*Eco*RI and *Bam*HI) for cloning and construction of our JMP5-chimeras, would excise the intron. Since introns do not form part of a functional transcript, we did not foresee any problems with the induction of the promoter once the intron was removed. Figure 5.2 illustrates the intron in p*ICL1* and the relevant restriction sites (communication Dr. M.B. Nthangeni).

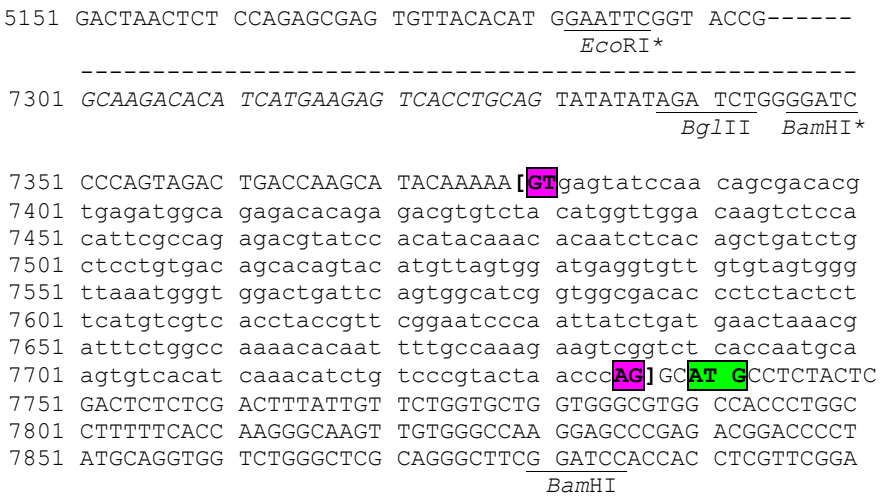


Fig. 5.2 An extract of the nucleotide sequence of p*ICL1* with its intron. The 2106 bp sequence, of the p*ICL*1, omitted in this figure, is illustrated by dashed lines. The *Eco*RI and *Bam*HI (marked with an asterisk) sites used for cloning into JMP5 are illustrated. The intron is illustrated in lower case letters and is captured with square brackets. The conical branch points are illustrated as pink boxes and the start codon of the *Y. lipolytica* cytochrome P450 reductase CPR gene is annotated as a green box.

As stated in Chapter 4, section 4.6.4, no fluorescence was observed when the p*ICL1* was cloned in frame with the GFP. We propose some of the possible reasons why this was the case:

In its native form, the intron of the p*ICL1* is correctly spliced. We have now removed this intron before splicing would commence. It is therefore possible that the intron provides a specific distance between the tranciptional start and the binding site for the RNA polymerase. If this was the case, then the specific sequence in between is not relevant but the distance the intron covers might be of importance. The random integration was not confirmed with a PCR and integration in our mind seemed successful once the recombinant E150 strain became a uracil prototroph. It might prove to be insightful to amplify the randomly integrated JMP5 plasmid as to rule out the possibility of failed integration into the yeast chromosome. In the future this study could be repeated with the 'intact' p*ICL1* without the removal of the intron.

5.3.4 What could have affected the absence GFP fluorescence?

When the GFP was amplified from the pYES2 vector, the reverse primer was designed in such a way as to omit the stop codon. This was necessary since the localization sequences (LS) needed to carry their native stop codons and translation needed to occur up to the LS stop codon and not the GFP. This was the first genetic manipulation introduced into the *GFP* gene (Chapter 4, section 4.6.2, Fig. 4.13). The cloning of the putative 150 bp *FALDH* LS to the 3′ end of the GFP could be seen as the second alteration. Since the GFP had no stop codon and the LS (with a stop codon) was cloned in frame to the GFP, it created an open reading frame (ORF) with a 150 bp extension (GFP + LS). LS of this length have traditionally been fused with *lac*Z when studying localization (Masaki *et al*., 1994; Augstein *et al*., 2003) but it was also used successfully with GFP when Masaki and co-workers (2003), performed post-translational targeting of a rat microsomal aldehyde dehydrogenase (msALDH) tail-anchored GFP protein to the ER. In the study by Masaki *et al*. (2003), the 105 bp (35 amino acids) of the msALDH LS was fused to the 3′ end of GFP and was successfully localized to the ER membrane.

The instrinsic fluorescence of the GFP protein is due to a unique covalently attached chromophore which is formed post-translationally within the protein upon cyclisation and oxidation of residues 65 – 67, serine-tyrosine-glycine (Cody *et al*., 1993). Could it be that the fusion of the LS perhaps altered the chromophore folding? Was this perhaps why we could not detect any fluorescence? It may prove useful to obtain the emission spectra for the wild-type GFP and to compare that with the emission spectra of the GFP chimera in JMP5. Spectral data might provide some insights towards the folding of the GFP protein.

We also noticed that the JMP5 vector contained no terminator. This might play a pivotal role in the successful expression of our GFP chimera. Localization studies in *Y. lipolytica* with *lac*Z and GFP both utilized vectors that harbored a terminator. For example, in the study by Augstein *et al*. (2003), a glyoxylate pathyway repressor 1 terminator (*GPR*1 term) was used and in the study by Nazarko *et al*. (2005), the alkaline extracellular protease 2 terminator (*XPR*2 term) was included in the pYEG1 vector. In the study conducted by Neuvéglise *et al*.,

1998, a shuttle mutagenesis system for tagging genes, was developed with a GFP mutant (S65T) that was fused to the *XPR2* term. In the future, one should consider placing a terminator at the 3′-end of the chimera. Whether this terminator would be of a *YlFALDH* origin or from other genes remains to be answered.

Complete JMP5 chimeras comprised of a: inducible p*ICL*1, GFP (without a stop codon) and a putative *YlFALDH* LS. *Y. lipolytica* E150 cells transformed with the JMP5 chimerical vector were cultured with ethanol as inducer. Ethanol was initially provided in vapour form but we latter switched to liquid form since the alcohol evaporated too rapidly from the sterile filter paper even when petri dishes were enclosed with parafilm. In addition, very poor growth was observed when ethanol was provided in vapour form. Unfortunately no speculation can be made about whether these *YlFALDH* LS would localize the recombinant GFP to specific targets in the yeast cell since no visible fluorescence was detected.

5.3.5 Cloning of the GFP-LS fragment into pKOV136

Initially when no fluorescence was observed, we first suspected that the GFP may have been defective. The pYES2 vector, from which the GFP was originally amplified from, was transformed into *S. cerevisiae* CEN PK49. The promoter was induced with galactose and fluorescence was observed (results are not shown). This ruled out the possibility of a defective GFP protein. However, it has to be stated once again, that the GFP in pYES2 was not genetically altered as in the case of the GFP in the JMP5-chimera.

We re-strategized and decided to select the pKOV136 vector that was constructed in our laboratory by Michel Labuschagne. This vector (Chapter 4, section 4.6.4, Fig. 4.20) is a single copy, integrative vector that utilizes the pBR322 docking platform for homologous integration into the yeast genome. Other features of this vector include: a constitutive transcription elongation factor-1α promoter (p*TEF*), a bacterial moiety containing the β-lactamase gene conferring ampicillin resistance, a *LEU2* auxotrophic selection marker and the lipase 2 terminator (*LIP*2 term). This vector seemed more advantageous than JMP5 since transcription was constitutive and in addition it possessed a terminator.

Our failure to illustrate fluorescence in the JMP5-chimera was, in our view, based on two problems: we suspected that the inducible p*ICL1* was not driving transcription and even if it was, we suspected that the absence of a terminator was also a contributing factor. We thus amplified the GFP-LS fragment from the JMP5-chimeras and cloned it into pKOV136. After culturing the recombinant *Y. lipolytica* Po1g strain in rich YPD media, the cells were inspected for fluorescence with a confocal laser scanning microscope. We observed the emission of light that was distributed in an almost granular-like fashion on the cell surface when using both the JMP5- and pKOV136-chimeras. This was initially thought to be fluorescence but after evaluating the negative control (untransformed E150 strain); an identical pattern of light emission was observed (Chapter 4, section 4.6.3, Fig. 4.19). The conclusion was drawn that it was most probably auto-fluorescence and not fluorescence by GFP.

No fluorescence was observed when *Y. lipolytica* Po1g was transformed with the JMP5-chimeras as well as for strains transformed with pKOV136. In the case of the JMP5-chimera, there was an inducible p*ICL1* and no terminator. Contrasting to this was the pKOV136 vector that contained a constitutive p*TEF* and a *LIP*2 term. It thus could be that neither the promoters nor the lack of or presence of a terminator was responsible for the absence of GFP fluorescence.

A possibility to consider might be the incorrect folding of this GFP-LS protein due to the addition of the 150 bp LS from the *YlFALDH*s to the 3′-end of the GFP protein. The chromophore of GFP is formed within the protein post-translationaly and GFP fusion proteins possess a ‘tight fold’ when expressed *in vivo*. It could be that the addition of the 150 bp LS caused the protein to fold the chromophore more inward *i.e*. not exposing it to the cytoplasm. Horie *et al*. 2002 attached 50 amino acids (150 bp) of Tom5 (a mitochondrial LS) to the COOH-terminal of a GFP protein and emphasized that the 150 bp of Tom5 should always face the surface of the fluorescent active molecule since GFP fusion proteins ‘fold tightly’ when expressed *in vivo*.

5.4 Culturing *YlFALDH* deletion mutants to assess their role in *n*-alkane metabolism

Tripple and quadruple deletion *YlFALDH* mutants were obtained from the study of Matatiele (2005). These *YlFALDH* deletion mutants were used by Matatiele (2005) to determine the role of *YlFALDH*s in n-alkane metabolism and the effect the deletions of these isozymes would have when cultured *n*-alkanes like *e.g*. decane, tetradecane, hexdecane and octadecane. From the growth studies performed by Matatiele (2005), it was concluded that the *YlFALDH* isozymes do not seem to be essential for the growth of *Y. lipolytica* on *n*-alkanes as sole carbon source since growth was achieved with even the quadruple *YlFALDH* deletion mutant. Matatiele (2005) ultimately concluded that all *YlFALDH* deletion mutants were still able to grow slowly on *n*-alkane media and suggested the presence of another mechanism of survival.

The aim of this experiment was to study the toxicity effects of 1-dodecanol and 1-dodecanal when the triple and quadruple *YlFALDH* deletion mutants were cultured on these substrates.. These results would then contribute to the existing *YlFALDH* information gathered from the study of Matatiele (2005) and consequently help us to understand the role of these four isozymes better.

The deletion mutants were cultured on rich YPD_2 medium (Smit *et al*., 2004) that contained either 1-dodecanol or its corresponding fatty aldehyde. These two substrates are toxic for *Y. lipolytica* and inhibit growth (Green *et al*., 2000; Smit *et al*., 2004). Consequently the cells have to eliminate these toxic substrates to alleviate the toxicity exerted by these substrates. In a study of *H. sapiens FALDH* by Willemsen and co-workers (2001) it was postulated that the deficiency of *FALDH* may lead to an accumulation of fatty alcohols or aldehyde modified macromolecules with structural consequences for cell membrane integrity and elevated concentrations of biologically highly active lipids that can be lethal to the cell.

*FALDH*s are responsible for the oxidation and detoxification of exogenous as well as endogenous aldehydes (Lin *et al*., 2000; Demozay *et al*., 2004). If the *YlFALDH*s are responsible for detoxification of the fatty aldehydes, then the deletion mutants would provide us with that answer. Undiluted triple and quadruple deletions had no effect on growth even in

the presence of 0.025% 1-dodecanol and 1-dodecanal. Even when the quadruple deletion mutants were diluted 500 fold, satisfactory growth was still obtained on the highest 1-dodecanol concentration (0.025%) (Chapter 4, section 4.5, Fig. 4.11). At a concentration of ≥0.01% it appears as if the alcohol has a more pronounced negative effect on growth than its aldehyde at the same concentration when cells were diluted 50 and 500 fold.

From this experiment it seems evident that the deletion of the *YlFALDH* isozymes has no effect on the growth of these deletion strains when cultured on toxic 1-dodecanol and 1-dodecanal. Further more it does not appear as if these *YlFALDH*s play a role in detoxifying the fatty aldehyde or the alcohol.

Ultimately we are concluding that these putative *YlFALDH* do not seem to play a role in *n*-alkane metabolism since they do not seem essential for growth on *n*-alkanes (Matatiele, 2005) and they do not seem to play a role in detoxifying fatty alcohols and their corresponding aldehydes.

5.5 Future research

In future it might prove useful to utilize a wild-type strain *e.g*. H222 for expression studies in minimal liquid medium containing *n*-alkanes as sole carbon source. The lag phase is only *ca*. 10 hours and biomass yields are much better as compared to W29 derivatives for example. Po1g. In addition one could also use shake flasks that are baffled or have notches to increase proper dispersion of the *n*-alkanes during cultivation. Zeta-based vectors could also be used instead of pBR322-based vectors and do not necessarily have to be transformed into strains with *zeta*-elements. If W29 derivatives are still preferred, then zeta-based vectors may also be used in Po1d, Po1f and Po1h strains since zeta sequences enhance non-homologous integration into *Y. lipolytica* strains devoid of Ylt1 retrotranposons. It should however be kept in mind that the long lag phase on *n*-alkanes is still present in these strains.

Future studies on the localization of these *YlFALDH*s could be performed with Enhanced Yellow Fluorescent Protein (EYFP). This is a GFP derivative and has been used with great

success when studying peroxisome dynamics in *Y. lipolytica* (Bascom *et al*., 2003; Nazarko *et al*., 2005). This EYFP gene is available from BD Biosciences Clontech, Palo Alto, CA.

If ER localization of these *YlFALDH*s is strongly suspected, spectral overlaps could be utilized for visualizing the localized *YlFALDH*s as well as the ER by using single fluorescein isothiocyanate filter set as described by Nazarko *et al*. (2005).

Alternatively, if one does not wish to employ fluorescence, an indirect method could also be utilized *e.g*. fractionation studies in which *lac*Z activity are measured. Alternatively chloramphenicol acetyltransferase assays could also be used for determining subcellular localization. Both these methods have been used successfully in the localization of the rat *FALDH* (Masaki *et al*., 1994).

Chapter 6
Conclusions

6.1 Expression studies of the four putative *Yarrowia lipolytica FALDH* promoters

In our initial experiments we attempted to construct a reporter system to study expression of the four *Yarrowia lipolytica FALDH* promoters (p*YlFALDH*). For this purpose the Sticky-end Polymerase Chain Reaction method (SEP) and Enzyme-free cloning method were employed. In the case of the SEP method the cardinal problem was the re-ligation of the plasmids back onto themselves which hampered subcloning. With the Enzyme-free cloning method, suspected contaminating template plasmid repeatedly resulted in failed cloning and transformation attempts. Both these methods were elegant and required less subcloning steps and post-amplification treatment but the results could not be reproduced and varied too much. It appears as if directional cloning with ligase was more advantageous in this part of the study.

Expression studies with the reporter vector: pINA781 proved to be more successful. The pINA781 vector was very suited for our needs since it already harbored a functional *lacZ* and since its original minimal p*LEU*2 could easily be removed and replaced with the p*YlFALDH*s. This vector also integrated into the yeast genome at a known locus and in single-copy fashion which made this vector ideal for comparison studies between different p*YlFALDH*s expressing the same reporter gene (*lacZ*).

With the aid of pINA781 and qualitative plate assays we concluded that p*YlFALDH1* and *2* were induced by dodecane and hexadecane but not by oleic acid. These two p*YlFALDH*s appeared to be repressed by glucose and glycerol. From the qualitative plate assays we also observed that transcription strength was higher in p*YlFALDH2* when compared to p*YlFALDH1*. From a qualitative point of view it also seems as if

p*YlFALDH2* has the same level of transcription as the strongly inducible *POX2* promoter (p*POX2*). No promoter activity could be observed (judged by a blue colour formation) when the strain containing the p*POX2* was cultured in glucose and glycerol although it has been stated that glucose and glycerol do not repress the p*POX2* completely (Juretzek *et al*., 2000). Promoter activity was observed when the p*POX2* containing strain was cultured in oleic acid and this was in accordance with literature (Juretzek *et al*., 2000).

It is interesting to note that the strains containing p*YlFALDH1* and *2* displayed different patterns of induction depending from which source they were streaked out. Strains streaked directly from glycerol stocks displayed blue colour formation (*i.e*. induction of the promoter) after 3 days but when identical strains were streaked from YPD plates, visible induction was only observed after 6 days. This may indicate the possibility that residual glucose represses p*YlFALDH1* and *2* stronger than glycerol.

p*YlFALDH3* did not display any activity on *n*-alkanes. Even after confirmation of successful genomic integration in *Y. lipolytica*, no p*YlFALDH3* activity could be detected with the qualitative plate assay. This finding was supported by the Northern hybridization studies performed by Matatiele (2005) who reported no *YlFALDH3* activity when the recombinant strains were cultured on *n*-alkanes. No induction could be illustrated for p*YlFALDH4* when cultured with *n*-alkanes but it has to be emphasized that confirmation of successful genomic integration in *Y. lipolytica* was never obtained. If we suppose that genomic integration was successful, we still could not observe any promoter activity and that contradicts the results of the Northern hybridization studies performed by Matatiele (2005) who reported *YlFALDH4* activity when strains were cultured in *n*-alkane containing media.

We concluded that recombinant strains should be cultured in minimal, buffered liquid media with *n*-alkanes as sole carbon source and be studied with a sensitive and quantitative assay procedure to eliminate the discrepancies between our results and that of Matatiele (2005) with the regard to p*YlFALDH4*. Such a study would also be useful to

quantify promoter strength of the *YlFALDH* promoters in order for comparison with other inducible as well as constitutive *Y. lipolytica* promoters.

Ultimately we concluded that p*YlFALDH1* and *2* are induced by dodecane and hexadecane but not by oleic acid, glucose and glycerol. It also appears from the plate assays that p*YlFALDH1* and *2* are not strong constitutive promoters.

During the expression studies the recombinant *Y. lipolytica* Po1g strains experienced a long lag phase in buffered minimal medium when *n*-alkanes served as sole carbon source. This has been well documented for W29-strains (Mauersberger *et al*., 2001). Since *Y. lipolytica* Po1g is a W29 derivative this explained the long lag phase of these strains. A similar problem was encountered with the Po1d strain when Juretzek *et al*., 2000 performed a comparative study of different *Y. lipolytica* promoters. The long lag phase complicated the standardization of our expression study and made comparisons of the promoters on different carbon sources difficult.

6.2 Attempts at localization with putative *Yarrowia lipolytica* localization seqeunces

The four *YlFALDH*s apparently do not possess NH_2-terminal LS since numerous software programs failed to detect it. Based on comparisons with the mammalian FALDHs we suggest that the *YlFALDH*s could have COOH-terminal LSs that could localize them to the membranes of the ER. This statement is justified if one looks at the *FALDH* COOH-terminal localization sequences (LSs) of *Homo sapiens* (Rizzo *et al*., 2001), *Mus musculus* (Rizzo *et al*., 2001) and *Ratttus norvegicus* (Masaki *et al*., 1994; 2003).

The last 35 amino acids on the COOH-terminal of these FALDH proteins are responsible for localization to the ER membrane and serve as a hydrophobic protein anchor. When comparing Kyte and Doolittle hydrophobicity plots of the last 35 amino acids on the COOH-terminal from *R. norvegicus* to that of the *YlFALDH*s, similar hydrophobicity trends are noticed for especially the translated *YlFALDH2* and *4* genes. The

hydrophobicity plots for *YlFALDH1* and *3* gene products do not compare well with that of *R. norvegicus* FALDH but this does not unequivocally prove that *YlFALDH1* and *3* do not possess LSs at their COOH-termini.

For localization studies in *Y. lipolytica* the zeta-based integrative JMP5 vector was used as well as the pBR322-based integrative pKOV136 vector.

When JMP5-chimeras were completed they comprised of: an inducible *ICL1* promoter (p*ICL1*) and a GFP with a putative 150 bp (50 amino acids) *YlFALDH* LS from each isozyme. We could not demonstrate fluorescence of the GFP when the JMP5-chimeras were transformed into *Y. lipolytica* E150. We did however observe a granular-like light emission on the surface of the cells but this was attributed to auto-fluorescence since the untransformed negative control also displayed the same light emission trend. Since the intron was removed from the inducible p*ICL1*, we suspect that the length of the intron and not the sequence information might play a role in the binding of the RNA polymerase to facilitate successful transcription. We also noted that the absence of a terminator in the JMP5-chimeras might also play a fundamental role in successful expression since other studies with GFP (involving localization) utilized plasmids with terminators of *e.g.* the *XPR2* and *LIP2* genes from *Y. lipolytica*. We also suspected that the GFP source might be defective and that this was the possible cause of failed green fluorescence. The source of the GFP in this study was from the pYES2 shuttle vector used for localization studies in *Saccharomyces cerevisiae*. This GFP was not defective since fluorescence was demonstrated when the shuttle vector was transformed into *S. cerevisiae* CEN PK49.

No fluorescence was observed with the pKOV136 vector which comprised of a constitutive *TEF* promoter (p*TEF*), GFP-LS fusion (amplified from a JMP5-chimera) and a *LIP*2 terminator (*LIP*2 term). When transformed cells were evaluated under a confocal laser microscope, identical patterns of light emission were observed as for the JMP5-chimeras.

It could also be that neither promoters, nor the lack of or presence of a terminator played a role in the failure to produce green fluorescence. Even when pKOV136, which has a constitutive promoter and *LIP*2 term, was used for localization studies, fluorescence could not be detected. The failed attempt at fluorescence could also be ascribed to the improper folding of the GFP's chromophore. Fluorescence was observed with the 'native' GFP protein in the pYES2 vector since the gene was not genetically altered. But one has to keep in mind that the GFP used in the JMP5-chimeras and in pKOV136 had no stop codon and contained a 50 amino acid sequence that was very hydrophobic in nature. It could be that the addition of this extra 50 amino acid sequence caused the GFP's chromophore to fold improperly and consequently render the protein incapable of fluorescence.

6.3 Do the putative *Yl*FALDHs play a vital role in *n*-alkane metabolism?

In the study of Matatiele (2005) growth experiments with *n*-alkanes, as only carbon source, were performed with single, double, triple and quadruple *YlFALDH* deletion mutants. From this study it was concluded that the four *Yl*FALDH isozymes do not seem to be essential for growth on *n*-alkanes. This raised the question of whether these *Yl*FALDHs play a role in *n*-alkane metabolism and it they do not, what is their function?

In this study we evaluated the same triple and quadruple *YlFALDH* deletion mutants from the study of Matatiele (2005) on plates containing 1-dodecanol and 1-dodecanal. These substrates are considered toxic for *Y. lipolytica* if they are not metabolized and consequently detoxified. FALDHs catalyze the oxidizion of fatty aldehydes and as a result the FALDH protein also performs a detoxifying function. From our toxicity experiments we concluded that the *Yl*FALDHs do not seem to play a role in detoxifying these substrates since even the quadruple *YlFALDH* deletion mutant's growth matched that of the wild-type which has all four *YlFALDH* isozymes intact. We concluded that there are possibly other enzymes *e.g.* cytochrome P450s that work independently or in concert with the *Yl*FALDH isozymes to detoxify 1-dodecanol and 1-dodecanal.

If one takes into account the above results of Matatiele (2005) and from the results from this study, it does not seem as if the four *Yl*FALDH isozymes play an essential role in either in the growth of *Y. lipolytica* on *n*-alkanes or in the detoxification of fatty alcohols and aldehydes. From a gene expression point of view it can be concluded from this study that the promoters of *YlFALDH1* and *2* are induced by *n*-alkanes but this provides a contradiction with regard to the above mentioned statements.

Consequently the following questions arise:

(1) If these four *Yl*FALDH isozymes are not essential for growth on *n*-alkanes, what enzymatic mechanism does *Y. lipolytica* employ to oxidize the fatty aldehyde intermediates of *n*-alkane metabolism?
(2) What enzymes play a detoxifying role when fatty alcohols and aldehydes are given to *Y. lipolytica*?
(3) What is the precise function and regulation of these *Yl*FALDH isozymes and do they play a role in *n*-alkane metabolism?
(4) Are there perhaps more *Yl*FALDH isozymes present in the genome of *Y. lipolytica*?

It is clear that considerable more research will have to be done in order to elucidate some of these questions. Quantification of expression and the successful identification of subcellular localization of these *Yl*FALDH proteins will help contribute to the little knowledge we have and in the understanding of the *Yl*FALDH isozymes and about yeast FALDHs in general.

Chapter 7

Summary

The dimorphic fungus *Yarrowia lipolytica* is an *n*-alkane assimilating yeast. During *n*-alkane oxidation toxic fatty aldehydes are formed that are further oxidized by fatty aldehyde dehydrogenases (*FALDH*) to carboxylic acids that then enter the β-oxidation pathway.

Very little research emphasis has been placed on *FALDH*s in yeasts and their precise role in *n*-alkane metabolism. This study aimed at contributing to the limited knowledge of yeast *FALDH*s and in particular the four putative *Y. lipolytica FALDH*s (*YlFALDH1 - 4*) that were recently identified in the fully sequenced genome of *Y. lipolytica* E150. The contribution made from this study to *YlFALDH*s was with reference to their promoter expression and subcellular localization.

The promoter and terminator of each *YlFALDH* was initially used to construct reporter cassettes in conjunction with β-galactosidase (*lacZ*) by utilizing the Sticky-end PCR (SEP) and Enzyme-free cloning methods. These two methods proved to be unsuitable for the expression study. The promoter region of each *YlFALDH* was then cloned into the pINA781 expression vector containing *lacZ* to study the expression further. With the aid the pINA781 integrative vector and qualitative plate assays, with 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal), it was observed that the promoters of *YlFALDH1* and *2* were inducible by dodecane and hexadecane but not by oleic acid, glucose or glycerol. The promoter of *YlFALDH2* also seemed to display the same level of transcriptional strength as the inducible *POX2* promoter. Induction of the *YlFALDH3* and *4* promoters was not observed.

Localization of the *YlFALDH* proteins was studied with the aid of green fluorescent protein (GFP) from *Aequorea victoria* and putative localization sequences (LS) from

each *YlFALDH* isozyme. The putative *Y. lipolytica* LSs comprised of the last 150 bp of the COOH-terminal of the *YlFALDH* proteins. These LSs were modeled from *Rattus norvegicus* FALDH that possesses a 35 amino acid hydrophobic protein anchor at its COOH-terminal.

For localization studies, chimerical JMP5 molecules were created with an inducible *ICL1* promoter, GFP and putative *Y. lipolytica* LS from each isozyme. Chimerical molecules were also constructed with a pKOV136 vector that contained a constitutive *TEF* promoter, a GFP-LS fragment (from a JMP5-chimera) and *LIP*2 terminator. No fluorescence was observed with epifluorescence or confocal laser microscopy when either of the JMP5- or pKOV136-chimeras were transformed into *Y. lipolytica* E150 and Po1g respectively. Consequently the subcellular localization could not be identified.

Keywords: *Yarrowia lipolytica*, fatty aldehyde dehydrogenase (FALDH), *n*-alkanes, expression studies, *FALDH* promoters, subcellular localization, localization sequences (LS)

Chapter 8

Opsomming

Die dimorfiese fungus *Yarrowia lipolytica* is ‘n *n*-alkaan assimilerende gis. Tydens oksidasie van *n*-alkane word toksiese vet aldehiede gevorm wat verder ge-oksideer word deur vet aldehied dehidrogenase (VALDH) om karboksielsure te vorm wat dan die β-oksidasie weg binne gaan.

Betreklik min klem is in die verlede geplaas op VALDHs in giste en hulle presiese rol in die metabolisme van *n*-alkane. Hierdie studie het gepoog om ’n bydrae te lewer tot die beperkte kennis van gis VADLH deur die vier putatiewe *Y. lipolytica VADLH*s (*YlVADLH1-4*) te bestudeer wat onlangs in die volkome bepaalde nukleotied volgorde van *Y. lipolyitca* E150 geïdentifiseer is. Die bydrae tot hierdie beperkte kennis van gis *VADLH* is gemaak m.b.t. uitdrukkings studies van die *YlVALDH* gene en die sellulêre lokalisering van die *Yl*FALDH proteïene in *Y. lipolytica*.

Die promoter en termineerder van elke *YlFALDH* was aanvanklik gebruik om rapporterings kassette te konstrueer m.b.v. die β-galaktosidase geen (*lacZ*) deur die ”Sticky-end PCR” (SEP) en die ”Enzyme-free cloning” metodes te gebruik. Die laasgenoemde twee metodes het nie herhaalbare en bevredigende resultate gelewer nie. Die promoter en termineerder van elke *YlVALDH* was toe gekloneer in die pINA781 uitdrukkings vektor (wat ’n *lacZ* geen bevat) om die uitdrukking van hierdie *YlVALDH*s verder te bestudeer. Met behulp van die pINA781 uitdrukkings vektor en kwalitatiewe plaat proewe, met 5-brom-4-chloro-3-indoliel-β-D-galaktosidase (X-gal), het ons vasgestel dat die promoters van *YlVALDH1* en *2* geïnduseer word deur dodekaan en heksadekaan, maar nie deur oleïensuur, glukose of gliserol nie. Die promoter van *YlVALDH2* blyk dieselfde vlak van uitdrukkings sterkte te vertoon as die van die sterk induseerbare *POX2* promoter. Induksie van die *YlVALDH3* en *YlVALDH4* promoters kon nie waargeneem word nie.

Lokalisering van die *Yl*VALDH proteiene was bestudeer m.b.v. ”Green Fluorescent Protein” (GFP) van die jellievis *Aequorea victoria* en putatiewe lokaliserings volgordes (LV) van elke *YlVALDH* wat gekloneer was aan die 3′-end van die GFP. Die putatiewe *Y. lipolytica* LVs het bestaan uit die laaste 150 basis pare van die COOH-terminus van die *YlVALDH* gene. Hierdie LVs van die *YlVALDH*s was gemodelleer op grond van die laaste 35 amino sure (aan die COOH-terminus) van *Rattus norvegicus* se VALDH wat dien as ’n hidrofobiese anker vir die proteïen in die membraan van die ER.

Die lokalerings studies was uitgevoer met chimere van JMP5 en pKOV136. Die JMP5-chimere het bestaan uit die induseerbare *ICL1* promoter, die ”GFP” en ’n putatiewe *YlVALDH* LV. Die pKOV136 –chimere het bestaan uit ’n konstitiewe *TEF* promoter, ’n ”GFP”-LV fragment (van ’n JMP5-chimeer) en ’n *LIP2* termineerder. Geen fluoresensie van die ”GFP” kon waargeneem word toe beide die JMP5- en pKOV136-chimere onderskeidelik in *Y. lipolytica* E150 en Po1g getransformeer is nie. Gevolglik kon lokalisering ook nie geïdentifiseer word nie.

Sleutelwoorde: *Yarrowia lipolytica*, vet aldehied dehidrogenase (VALDH), *n*-alkane, uitdrukkings studies, *VALDH* promoters, sellulêre lokalisering, lokaliserings volgordes (LV)

Literature cited

Ahvazi, B., Coulombe, R., Delarge, M., Vedadi, M., Zhang, L., Meighen, E. and Vrielink, A. (2000). Crystal structure of the $NADP^+$-dependent aldehyde dehydrogenase from *Vibrio harveyi*: structural implications for cofactor specificity and affinity. *Biochem J* **349**, 853-861.

Aslanidis, C. and de Jongh, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nuleic Acids Res* **18**, 6069-6074.

Augstein, A., Barth, K. Gentsch, M., Kohlwein, S.D. and Barth, G. (2003). Characterization, localization and functional analysis of Gpr1p, a protein affecting sensitivity to acetic acid in the yeast *Yarrowia lipolytica*. *Microbiology* **149**, 589-600.

Barathi, S. and Vasudevan, N. (2001). Utilization of petroleum hydrocarbons by *Pseudomonas fluorescens* isolated from a petroleum-contaminated soil. *Environ Int* **26**, 413-416.

Barns, S.M., Lane, D.J., Sogin, M.L., Bibeau, C. and Weisburg, W.G. (1991). Evolutionary relationships among pathogenic *Candida* species and relatives. *J Bacteriol* **173**, 2250-2255.

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

Barth, G. & Gaillardin, C. (1996). *Yarrowia lipolytica*. pp. 313-388. In: K. Wolf (ed.) Nonconventional yeasts in biotechnology. A Handbook. Springer-Verslag, Berlin, Germany.

Blanchin-Roland, S., Cordero Otero, R. and Gaillardin, C. (1994). Two upstream activation sequences control the expression of the *XPR2* gene in the yeast *Yarrowia lipolytica*. *J Biol Chem* **273**, 30903-30908.

Bognar, A. and Meighen, E.A. (1978). An induced aliphatic aldehye dehydrogenase from the bioluminescent bacterium *Vibrio harveyi*, Purification and Properties. *J Biol Chem* **253**, 446-450.

Bokman, S.H. and Ward, W.W. (1981). Renaturation of *Aequorea victoria* green fluorescent protein. *Biochem Biophys Res Commun* **101**, 1372-1380.

Borgese, N., Colombo, S. and Pedrazinni, E. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. *J Cell Biol* **161**, 1013-1019.

Byers, D. and Meighen, E.A. (1984). *V. harveyi* aldehyde dehydrogenase: Partial reversal of aldehyde oxidation and its possible role in the reduction of fatty acids for bioluminescent reaction. *J Biol Chem* **259**, 7109-7114.

Casaregola, S., Neuvéglise, C., Lépingle, A., Bon, E., Feynerol, C., Artiguenave, F., Wincker, P. and Gaillardin, C. (2000). Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. *FEBS Lett*. **487**, 95-100.

Chen, D.-C., Yang, B.C. and Kuo, T.T. (1992). One step-transformationof yeast in stationary phase. *Curr Genet* **21**, 83-84.

Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G. and Ward, W.W. (1993). Chemical –structure of the hexapeptide chromophore of the *Aequorea victoria* green fluorescent protein. **Biochemistry 32**, 1212-1218.

Davis, S.J. and Vierstra, R.D. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plan Mol Biol* **36**, 521-528.

de Laurenzi, V. Rogers, G.R. Hamrock, J.D., Marekov, L.N., Dompton, J.G., Markova N. and Rizzo, W.B. (1996). Sjögren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nat Genet* **12**, 52-57.

Demozay, D., Rocchi, S., Mas, J-C, Grillo, S., Pirola, L. Chavey, C. and van Obberghen, E. (2004). Fatty aldehyde dehydrogenase: Potential role in oxidative stress protection and regulation of its gene expression by insulin. *J Biol Chem* **279**, 6261-6270.

Dickinson, F. & Wadforth, C. (1992). Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. *Biochem J* **282**, 325-331.

Du, W., Huang, Z., Flaherty, J.E., Wells, K. and Payne, G.A. (1999). Green fluorescent protein as a reporter to monitor gene expression and food colonization by *Aspergillus flavus*. *Appl Environ Microbiol* **65**, 834-836.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker P. and Souciet, J.L. (2004). Genome evolution in yeasts. *Nature* **430**, 35–44.

Eirich, L., Craft, D., Steinberg, L., Asif, A., Eschenfeldt, W., Stols, L., Donnelly, M. and Wilson, C. (2004). Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain ATCC 20336. *Appl Environ Microbiol* **70**, 4877-4879.

Ekundayo, E.O. and Obuekwe, C.O. 2000. Hydrocarbon utilization in yeast isolates found to grow in association with petroleum in a polluted ultisol of midwestern Nigeria. *Environ Monit Assess* **63**, 381-387.

Endrizzi, A., Pagot, Y., Le Cleinche, A., Nicaud, J-M. and Belin, J-M. 1996. Production of lactones and peroxisomal β-oxidation in yeasts. *Critical Rev Biotechnol* **16**, 301-329.

Fickers, P., Benetti, P-H, Waché, Y. Marty, A., Mauersberger, S., Smit, M.S. and Nicaud, J-M. (2005). Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Research* **5**, 527-543.

Fickers, P., Le Dall, M-T, 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.

Flores, C-L & Gancedo, C. (2005). *Yarrowia lipolytica* mutants devoid of pyruvate carboxylase activity show an unusual growth phenotype. *Eukaryot Cell* **4**, 356-364.

Geraghty, M.T., Bassett, D., Morrell, J.C. Gatto, G.J., Bai, J., Geisbrecht, B.V., Hieter, P. and Gould, S.J. (1999). *Proc Natl Acad Sci USA* **96**, 2937-2942.

Hamsa, P.V. and Chattoo, B.B. (1994). Cloning and growth-regulated expression of the gene encoding the hepatitis B virus middle surface antigen in *Yarrowia lipolytica*. *Gene* **143**, 165-170.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Hempel, J., Perozich, J., Chapman, T., Rose, J., Boesch, J.S., Liu, Z.J., Lindahl, R. and Wang, B.C. (1999). Aldehyde dehydrogenase catalytic mechanism. A proposal. *Adv Exp Med Biol* **463**, 53-59.

Horie, C., Suzuki, H., Sakaguchi, M. and Mihara, K. (2003). Characterization of signal that directs C-tail-anchored proteins to mammalian mitochondrial outer membrane. *Mol Biol Cell* **13**, 1615-1625.

Hurtado, C.A.R. and Rachubinski, R.A. (2002). Isolation and characterization of *YlBEM1*, a gene required for cell polarization and differentiation in the dimorphic yeast *Yarrowia lipolytica*. *Eukaryot Cell* **1**, 526-537.

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 for new CYP52 family members.*Yeast* **16**, 1077-1087.

Ishige, T., Tani, A., Sakai, Y. and Kato, N. (2000). Long-chain aldehyde dehydrogenase that participates in *n*-alkane utilization and wax ester synthesis in *Acinetobacter* sp. strain M-1. *Appl Environ Micriobiol* **66**, 3481-3486.

Ishige, T., Tani, A., Takabe, K. Kawasaki, K., Sakai, Y. and Kato, N. (2002). Wax ester production from *n*-alkane by *Acinetobacter* sp. strain M-1: ultrastructure of cellular inclusions and role of acyl coenzyme A reductase. *Appl Environ Microbiol* **68**, 1192-1195.

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

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

Kelson, T., McVoy, J. and Rizzo, W. (1997). Human liver fatty aldehyde dehydrogenase: microsomal localization purification, and biochemical characterization. *Biochim Biophys Acta* **1335**, 99-110.

Kemp, G., Dickinson, F. and Ratledge, C. (1990). Light sensitivity of the *n*-alkane-induced fatty alcohol oxidase from *Candida tropicalis* and *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **32**, 461-464.

Kim, T-H, Lee, J-H, Oh, Y-S, Bae, K. and Kim S-J (1999). Identification and characterization of an oil-degrading yeast, *Yarrowia lipolytica* 180. *J Microbiol* **37**, 128-135.

Kim, T-H, Oh, Y-S and Kim, S-J (2000). The possible involvement of the cell surface in aliphatic hydrocarbon utilization by an oil-degrading yeast, *Yarrowia lipolytica* 180. *J Microbiol Biotechnol* **10**, 333-337.

Kok, M., Oldenhuis, R., van der Linden, M., Meulenberg, C., Kingma, J. and Witholt, B. (1989). The *Pseudomonas oleovorans alk*BAC operon encodes two structurally related rubredoxins and an aldehyde dehydrogenase. *J Biol Chem* **264**, 5442-5451.

Koller, A., Spong, A.P. Lüers, G.H. and Subramani, S. (1999). Analysis of the peroxisomal acyl-CoA oxidase gene product form Pichia pastoris and determination of its targeting signal. *Yeast* **15**, 1035-1044.

Kornfeld, S. and Mellman, I. (1989). The biogenesis of lysosomes. *Annu Rev Cell Biol* **5**, 483-525.

Kurtzman, C.P & Robnett, C.J. (1995). Molecular relationships among hyphal ascomycetous yeasts and yeastlike taxa. *Can J Bot* **73** (Suppl 1), S824-S830.

Kurtzman, C.P. & Fell, J.W. (1998). Definition, classification and nomenclature of the yeasts. In: The yeasts, A taxonomic study, C.P. Kurtzman and J.W. Fell (eds.) 4th edn. pp. 3-5. Elsevier Science B.V., Amsterdam.

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

Lin, Z., Carney, G. and Rizzo, W. (2000). Genomic organization, expression, and alternate splicing of the mouse fatty aldehyde dehydrogenase gene. *Mol Genet Metab* **71**, 496-505.

Liu, Z. (1996). Hetro-stagger cloning: efficient and rapid cloning of PCR products. *Nucleic Acids Res* **24**, 2458-2459.

Liu, Z.J., Hempel, J., Sun, J., Rose, J., Hsiao, D. Chang, W.R., Chung, Y.J., Kuo, I., Lindahl, R. and Wang, B.C. (1997b). Crystal structure of a class 3 aldehyde dehydrogenase at 2.6Å resolution. *Adv Exp Med Biol* **414**, 1-7.

Liu, Z.J., Sun, Y.J., Rose, J., Chung, Y.J., Hsiao, C.D., Chang, W.R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. Wang, B.C. (1997a). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Struct Biol* **4**, 317-326.

Luo, Y., Wang, H., Gopolan, K., Srivastava, D., Nicaud, J-M and Chardot, T. (2000). Purification and characterization of the recombinant form of acyl CoA oxidase 3 from the yeast *Y. lipolytica*. *Archives Biochem Biophys* **384**, 1-8.

Luo, Y., Nicaud, J-M, Van Veldhoven, P., and Chardot, T. (2001). The acyl-CoA oxidase from the yeast *Y. lipolytica*: characterization of Aox2p. *Archives Biochem Biophys* **407**, 32-38.

Madigan, M.T., Martinko, J.M. and Parker, J. (2000). Brock – Biology of microorgasnisms, 9th edn Prentice Hall, Inc. New Jersey

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

Madzak, C., Gaillardin, C. and Beckerich, J-M (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J Biotechnol* **109**, 63-81.

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.

Maftahi, M., Gaillardin, C. and Nicaud, J-M. (1996). Sticky-end polymerase chain reaction method for systematic gene disruption in *Saccharomyces cerevisiae*. *Yeast* **12**, 859-868.

Marchal, S. and Branlant, G. (1999). Evidence for the chemical activation of essential Cys-302 upon cofactor binding to nonphosphorylating glyceraldehydes 3-phosphate dehydrogenase from *Streptococcus mutans*. *Biochemistry* **38**, 12950-12958.

Marchal, S. and Branlant, G. (1999). Evidence for the chemical activation of essential Cys-302 upon cofactor binding to nonphosphorylating glyceraldehyde 3-phosphate dehydrogenase from *Streptococcus mutans*. *Biochemistry* **38**, 12950-12958.

Margesin, R., Gander, S., Zacke, G., Gounot, A.M. and Schinner, F. (2003). Hydrocarbon degradation and enzyme activities of cold-adapted bacteria and yeasts. *Extremophiles* **7**, 451-458.

Masaki, R., Kameyama, K. and Yamamoto, A. (2003). Post-translational targeting of a tail-anchored green fluorescent protein to the endoplasmic reticulum. *J Biochem* **134**, 415-426.

Masaki, R., Yamamoto, A. and Tashiro, Y. (1994). Microsomal aldehyde dehydrogenase is localized to the endoplasmic reticulum via its carboxy-terminal 35 amino acids. *J Cell Biol* **126**, 1407-1420.

Matatiele, P.R. (2005). Fatty alcohol and fatty aldehyde dehydrogenases of *Yarrowia lipolytica*. Ph.D thesis, University of the Free State, Bloemfontein, South Africa.

Mauersberger, S., Drechsler, H., Oehme, G. and Müller, H-G (1992). Substrate specificity and stereoselectivity of fatty alcohol oxidase from the yeast *Candida maltosa*. *Appl Microbiol Biotechnol* **37**, 66-73.

Mauersberger, S., Wang, H-J, Gaillardin, C., Barth, G. and Nicaud, J-M (2001). Insertional mutagenesis in the *n*-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. *J Bacteriol* **183**, 5102-5109.

Miyauchi, K. Masaki, R, Taketani, S, Yamamoto, A., Akayama, M. and Tashiro, Y. lipolytica (1991). *J Biol Chem* **266**, 19536-19542.

Müller, S., Sandal, T., Kamp-Hansen, P. and Dalboge, H. (1998). Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast* **14**, 1267-1283.

Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99-109.

Nakayasu, H., Mihara, K. and Sato, R. (1978). Purification and properties of a membrane-bound aldehyde dehydrogenase from rat liver microsomes. *Biochem Biophys Res Commun* **83**, 697-703.

Nazarko, T.Y., Nicaud, J-M. and Sibirny, A.A. (2005). Observation of the Yarrowia lipolytica peroxisome –vacuole dynamics by fluorescence microscopy with a single filter set. *Cell Biol Int* **29**, 65-70.

Nehlin, J.O., Carlberg, M. and Ronne, H. (1991). Control of yeast genes by MIG1 repressor: a transcriptional cascade in the glucose response. *EMBO J* **10**, 3373-3377.

Neuvéglise, C., Chalvet, F., Wincker, P., Gaillardin, C. and Casaregola, S. (2005). *Mutator*-like element in the yeast *Yarrowia lipolytica* displays multiple alternative splicings. *Eukaryot Cell* **4**, 615-624.

Nicaud, J-M., Fabre, E. and Gaillardin, C. (1989). Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. *Curr Genet* **16**, 253-260.

Nicaud, J-M., Le Clainche, A., Le Dall, M-T, Wang, H and Gaillardin, C. (1998). *Yarrowia lipolytica*, a yeast model for the genetic studies of hydroxyl fatty acids biotransformation into lactones. *J Mol Catal, B Enzym* **5**, 175-181.

Nilsson, T., Jackson, M. and Peterson, P.A. (1989). Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* **58**, 707-718.

Nishizawa, M., Araki, R. and Teranishi, Y. (1989). Identification of an upstream activating sequence and an upstream repressible sequence of the pyruvate kinase gene of the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **9**, 442-451.

Nthangeni, M.B., Urban, P., Pompon, D., Smit, M.S. and Nicaud, J-M (2004). The use of *Yarrowia lipolytica* for the expression of human cytochrome P450 CYP1A1. *Yeast* **21**, 583-592.

Ohkuma, M., Zimmer, T., Iida, T., Schunck, W-H., Ohta, A. and Takagi, M. 1998. Isozyme function of *n*-alkane –inducible cytochromes P450 in *Candida maltosa* revealed by sequential gene disruption. *J Biol Chem* **273**, 3948-3953.

Olesen, J.T. and Guarente, L. (1990). The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. *Genes Dev* **4**, 1714-1729.

Panke, S., Meyer, A., Huber, C., Witholt, B. and Wubbolst, M. (1999). An alkane-responsive expression system for the production of fine chemicals. *Appl Environ Microbiol* **65**, 2324-2332.

Parshyna, I. and Barth, G. (2003). Detection of biosynthesis and degradation of peroxisomes by different carbon sources with a combined assay using a β-galactosidase-GFP-SKL chimaeric protein for labeling of peroxisomes and the peroxisomal isocitrate lyase as control. *Non-conventional yeasts in genetics, biochemistry and biotechnology*, K. Wolf, K. Breunig and G. Barth (Eds.), 385-392.

Perozich, J., Kuo, I., Lindahl, R., Hempel, J. (2001). Coenzyme specificity in aldehyde dehydrogenase. *Chemico-Biological Interactions* **130-132**, 115-124.

Perozich, J., Kuo, I., Wang, B., Boesch, J. Lindahl, R. and Hempel, J. (2000). Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. *Eur J Biochem* **267**, 6197-6203.

Perozich, J., Nicholas, H., Wang, B-C, Lindahl, R. and Hempel J. (1999). Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* **8**, 337-146.

Pignède, G., Wang, H-J, Fudalej, F., Seman, M., Gaillardin, C. and Nicaud, J-M 2000. Autocloning and amplification of *LIP2* in *Yarrowia lipolytica*. *Appl Environ Microbiol* **66**, 3283-3289.

Prasher, D.C., Eckenrode, V.K. Ward, W.W., Prendergast, F.G. and Cormier, M.J. (1992). Primary structure of the *Aequorea victoria* green fluorescent protein. *Gene* **111**, 229-233.

Rapieko, P.J. and Gilmore, R. (1992). Protein translocation across the ER requires a functional GTP binding site in the a subunit of the signal recognition particle receptor. *J Cell Biol* **92**, 1-22.

Ratledge, C. (1984). Microbial conversions of alkanes and fatty acids. *J Am Oil Chem Soc* **61**, 447-453.

Richard, M., Quijano, R., Bezzate, S., Bordon-Pallier, F. and Gaillardin, C. (2001). Tagging morphogenetic genes by insertional mutagenesis in the yeast *Yarrowia lipolytica*. *J Bacteriol* **183**, 3098-3107.

Rizzo, W., Lin, Z. and Carney, G. (2001). Fatty aldehyde dehydrogenase: genomic structure, expression and mutation analysis in Sjögren-Larsson syndrome. *Chem Biol Interact* **130**, 297-307.

Rogers, G., Markova, N., de Laurenzi, V., Rizzo, W. and Compton, J. (1997). Genomic organization and expression of the human fatty aldehyde dehydrogenase gene (FALDH). *Genomics* **39**, 127-135.

Ross-Macdonald, P., Sheehan, A., Shirleen Roeder, G. and Snyder, M. (1997). A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **94**, 190-195.

Ruiz-Herrera, J., García-Maceira, P., Castillo-Barahona, L, Valentín, E. and Sentandreu, R. (2003). Cell wall composition and structure of *Yarrowia lipolytica* transposon mutants affected in calcofluor sensitivity. *Antonie van Leeuwenhoek* **84**, 229-238.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning. A laboratory manual Vol. 1 (2nd ed.). Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Sauer, B. (1987). Functional expression of the *cre-lox* site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**, 2087-2096.

Scheller, U., Zimmer, T., Becher, D., Schauer, F. and Schunck, W.H. (1998). Oxygenation cascade in conversion of *n*-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A2. *J Biol Chem* **273**, 32528-32534.

Schlegel, H. (1995). General microbiology 7th edition. Cambridge University Press, Cambridge.

Schmid-Berger, M., Schmid, B. and Barth, G. (1994). Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. *J Bacteriol* **176**, 2477-2482.

Senam S. and Barth G. (2003). Chapter 57. Detection and activation of the Ylt1 retrotransposon in the yeast *Yarrowia lipolytica*. In: Wolf K., Breunig K., Barth G. (Eds). Non-conventional yeasts in genetics, biochemistry and biotechnology. Springer-Verlag, Berlin, Heidelberg, New York, pp 357-364.

Singer, M. and Finnerty, W. (1985). Fatty aldehyde dehydrogenase in *Acinetobacter* sp. strain HO1-N: role in hexadecane and hexadecanol metabolism. *J Bacteriol* **164**, 1011-1016.

Sophos, N, Pappa, A., Ziegler, T. and Vasiliou V. (2001). Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chem Biol Interact* **130-132**, 323-337.

Sophos, N. & Vasiliou, V. (2003). Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact* **143-144**, 5-22.

Sumita, T., Iida, T., Hirata, A., Horiuchi, H., Takagi, M. and Ohta, A. (2002). Peroxisome deficiency represses the expression of *n*-alkane-inducible *YlALK1* encoding cytochrome P450ALK1 in *Yarrowia lipolytica*. *FEMS Microbiol Lett* **214**, 31-38.

Tillet, D. and Neilan, B. (1999). Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Res* **27**, e26 (i)-(iii).

Urlacher, V., Lutz-Wahl, S. and Schmid R. (2004). Microbial P450 enzymes in biotechnology. *Appl Microbiol Biotechnol* **64**, 317-325.

van Beilen, J., Eggink, G., Enequist, H. Bos, R. and Witholt, B. (1992). DNA sequence determination and functional characterization of the OCT-plasmid-encoded *alk*JKL genes of *Pseudomonas oleovorans*. *Mol Microbiol* **6**, 3121-3136.

van Beilen, J., Panke, S., Lucchini, S., Franchini, A., Rothlisberger M. and Witholt, B. (2001). Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the *alk* genes. *Microbiology* **147**, 1621-1630.

van Beilin, Z.L., Duetz, W.A., Smits, T.H.M. and Witholt, B. (2003). Diversity of alkane hydroxylase systems in the environment. *Oil & Gas Science and Technology – Rev IFP* **58**, 427-440.

Vanhanen, S., West, M., Kroon, T.M., Linder, N., Casey, Q.C., Elborough, M. and Slabas, A.R. 2000. A consensus sequence for Long-chain fatty-acid alcohol oxidases from *Candida* Identifies a family of genes involved in lipid ω-oxidation in yeast with homologues in plants and bacteria. *J Biol Chem* **275**, 4445-4452.

Vasiliou, V., Kozak, C.A., Lindahl, R. and Nebert, D.W. (1996). Mouse microsomal Class 3 aldehyde dehydrogenase: ADH3 cDNA sequence, inducibility by dioxin and clofibrate, and genetic mapping. *DNA Cell Biol* **15**, 235-245.

Vasiliou, V., Pappa, A. and Petersen, D. (2000). Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem Biol Interact***129**, 1-19.

Vedadi, M. Szittner, R., Smillie, L, an Meighen, E. (1995). Involvement of cysteine 289 in the catalytic activity of an NAP^{+}-specific fatty aldehyde dehydrogenase from *Vibrio harveyi*. *Biochemistry* **34**, 16725-16732.

Vedadi, M., Croteau, N., DeLarge, M., Vrielink, A. and Meighen, E. (1996). Structural and funtional studies of a $NADP^+$-specific aldehyde dehydrogenase from the luminescent marine bacterium *Vibrio harveyi*. *Adv Exp Med Biol* **414**, 269-275.

von Arnim, A.G., Deng, X-W. and Stacey, M.G. (1998). Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. *Gene* **221**, 35-43.

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

Watkinson, R.J. and Morgan, P. 1990. Physiology of aliphatic hydrocarbon-degrading microorganisms. *Biodegradation* **1**, 79-92.

Watson, J.D., Gilman, M., Witkowski, J. and Zoller, M. (1992). Recombinant DNA, 2nd edn. Scientific American Books, New York.

White, J.S. & White, D.C. (1997). Source book of enzymes, CRC Press LLC, Florida USA.

Willemsen, M.A.A.P., IJlst, L., Steijlen, P.M. Rotteveel, J.J., de Jong, J.G.N., van Domburg, P.H.M.F., Mayatepek, E., Gabreëls, J.M. and Wanders, R.J.A. (2001). Clinical, biochemical and molecular genetic characteristics of 19 patients with the Sjögren-Larsson syndrome. *Brain* **124**, 1426-1437.

Wymore, T., Field, M.J., Deerfield, D.W., Nicholas, H.B. and Hempel, J. (2004). Mechanistic details of aldehyde dehydrogenase chemistry from MM and QM/MM simulations. http://www.psc.edu/biomed.

Xie, Y.Q., Takimoto, K., Pitot, H., Miskimins, W.K. and Lindahl, R. (1996). Characterization of the rat Class 3 aldehyde dehydrogenase gene promoter. *Nucleic Acid Res* **24**, 4185-4191.

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

Yamagami, S., Morioka, D., Fukuda, R. and Ohta, A. (2004). A basic helix-loop-helix transcription factor essential for cytochrome P450 induction in response to alkanes in yeast *Yarrowia lipolytica*. *J Biol Chem* **176**, 2477-2482.

Yang, T-T., Sinai, P., Green, G. Kitts, P.A., Chen Y-T., Lybarger, L., Chervenak, R., Patterson, G.H., Piston, D.W. and Kain, S.R. (1998). Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein. *J Biol Chem* **273**, 8212-8216.

Yoshida, A., Rzhetsky, A., Hsu, L.C. and Chang, C. (1998). Human aldehyde dehydrogenase gene family. *Eur J Biochem* **251**, 549-557.

Zhang, L., Ahvazi, B., Szittner, R. Vrielink, A. and Meighen, E. (2001). Difference in nucleotide specificity and catalytic mechanism between *Vibrio harveyi* aldehyde dehydrogenase and other members of the aldehyde dehydrogenase superfamily. *Chem Biol Interact* **130-132**, 29-38.