

**Fatty Alcohol and Fatty Aldehyde  
Dehydrogenases of *Yarrowia lipolytica***



**Puleng Rose Matatiele**

**Fatty Alcohol and Fatty Aldehyde  
Dehydrogenases of *Yarrowia lipolytica***

**Submitted in fulfilment of the requirements for the degree**

**Philosophiae Doctor**

**in the**

**Department of Microbial, Biochemical and Food Biotechnology**

**Faculty of Natural and Agricultural Sciences**

**University of the Free State**

**Bloemfontein**

**Republic of South Africa**

**May 2005**

**by**

**Puleng Rose Matatiele**

**Promoter: Prof. MS Smit**

**Co-promoter: Dr. J Albertyn**

“The person who removes a mountain begins by carrying away small stones”.  
-- *Unknown*.

## **Acknowledgements**

My sincerest gratitude goes to my promoters Prof. Martie Smit and Dr. Jacobus Albertyn, for their invaluable guidance throughout the course of this study. I wish to thank them especially for having offered me the opportunity to advance my career up to this level. I am really grateful for every moment of this enriching experience of my life. I also thank Dr Jean-Marc Nicaud for all the plasmids and information from the Genolevures project. This study would have not been possible without all that information.

I also wish to thank my family, especially my old ailing mother for never doubting me once, even when things made sense no more. Your unfailing support and encouragement were a haven in times of pain and turmoil. This work, achieved through courage and perseverance, a character I cherish too much, is in memory of my late father.

Many thanks to my friends, colleagues and staff of this department for providing such a warm atmosphere. Your willingness to help at all times wherever possible enabled me to thrive. May God bless all of you!

Through the financial support from the National Research Foundation (NRF) and the Department of Labour (DoL) completion of this study was made possible. Both are greatly acknowledged.

And, finally through God's wish and will I have come this far. Praise be to Him always!

## Table of Contents

Acknowledgements.....	(i)
Table of contents.....	(ii)
List of abbreviations.....	(vii)
List of figures.....	(ix)
List of tables.....	(xi)
Prologue.....	(xii)
<b>Chapter One – Aldehyde- and alcohol dehydrogenases: A literature Review.....</b>	<b>1</b>
1.1 Aldehyde Dehydrogenases.....	1
1.1.1 Introduction.....	1
1.1.2 The aldehyde dehydrogenase structure.....	2
1.1.3 Aldehyde dehydrogenase reaction mechanism.....	4
1.1.4 Substrate specificity and coenzyme preference in ALDHs.....	5
1.1.5 Classification of aldehyde dehydrogenase enzymes.....	7
1.1.5(a) Reaction-Chemistry-Directed approach.....	7
1.1.5(b) Substrate specificity-based classification.....	9
1.1.6 Nomenclature of aldehyde dehydrogenases.....	14
1.1.7 The ALDH gene superfamily.....	17
1.1.8 Conserved residues and sequence motifs in ALDHs.....	21
1.1.8(a) Conserved residues.....	21
1.1.8(b) Conserved sequence motifs.....	22
1.1.9 Fatty aldehyde dehydrogenases.....	26
1.1.10 Fungal aldehyde dehydrogenase genes.....	29
1.2 Alcohol Dehydrogenases.....	32
1.2.1 Introduction.....	32
1.2.2 The different alcohol dehydrogenase classes.....	33
1.2.2(a) Zinc-dependent medium chain alcohol dehydrogenases.....	35
1.2.2(b) Short–chain alcohol dehydrogenases.....	36
1.2.2(c) Iron-activated alcohol dehydrogenases.....	37
1.2.2(d) Long chain alcohol dehydrogenases.....	38

1.2.3	The structure of alcohol dehydrogenase.....	38
1.2.4	Mechanism of action of alcohol dehydrogenase.....	42
1.2.5	Nomenclature of alcohol dehydrogenases.....	44
1.2.6	Fungal long chain alcohol dehydrogenases.....	49
1.2.7	Relationships among the different classes of ADH.....	50
1.3	Fatty alcohol and aldehyde dehydrogenases of <i>Yarrowia lipolytica</i> .....	54
	Motivation.....	57
	References.....	59
<b>Chapter Two – An exploratory study into the presence of long chain alcohol and aldehyde dehydrogenases in <i>Yarrowia lipolytica</i></b> .....		<b>78</b>
2.1	Introduction.....	78
2.2	Materials and Methods.....	79
2.2.1	Growth of organisms.....	79
	2.2.1 (a) <i>Pre-cultures</i> .....	79
	2.2.1(b) Growth of cells for FALDH and FADH enzyme assays.....	79
	2.2.1(c) Glucose derepressed cells for total RNA isolation.....	80
2.2.2	Preparation of enzyme extracts.....	80
	2.2.2 (a) Preparation of cell-free enzyme extracts.....	80
	2.2.2(b) Subcellular fractionation of enzyme activities.....	81
	2.2.2 (c) Y-PER™ treatment of cells.....	81
2.2.3	Enzyme assays.....	82
	2.2.3(a) Long chain alcohol dehydrogenase (FADH) and fatty aldehyde dehydrogenase (FALDH) assays.....	82
	2.2.3(b) NAD and NADP dependence of FALDH.....	83
	2.2.3(c) Fatty alcohol oxidase assay.....	83
2.2.4	BLAST searches for putative FAOD, FADH and FALDH genes.....	84
2.2.5	Identification of FAOD, FALDH and FADH genes in <i>Y. lipolytica</i> .....	84
	2.2.5(a) PCR amplification of FALDH and <i>b-actin</i> genes.....	84
	2.2.5(b) PCR amplification of FADH gene.....	86
	2.2.5(c) PCR amplification of FAOD gene in <i>C. tropicalis</i> OC3.....	87
	2.2.5(d) Southern Hybridization.....	87

2.2.6	Expression analysis of ALDH and FADH genes.....	87
2.2.6(a)	Northern Hybridization.....	87
2.2.6(b)	RT-PCR of ALDHs 1 and 2, and FALDHs 3 and 4.....	87
2.2.6(c)	Preparation of radiolabelled probes for Northern blot analyses...	88
2.3	Results and Discussion.....	89
2.3.1	FAOD, FADH and FALDH enzyme activity in <i>Y. lipolytica</i> .....	89
2.3.1.1	FAOD, FALDH and FALDH activity in hexadecane and glycerol grown cells.....	89
2.3.1.2	Subcellular fractionation of FADH and FALDH activity.....	91
2.3.1.3	NAD and NADP dependence of FADH and FALDH activity.....	93
2.3.2	Identification of FAOD, FALDH and FADH genes in <i>Y. lipolytica</i> .....	94
2.3.2.1	BLAST searches for putative FAOD, FADH and FALDH genes.....	94
	One putative FADH encoding gene.....	95
	Four putative FALDH encoding genes.....	98
2.3.2.2	Southern Hybridization to detect FAOD genes in <i>Y. lipolytica</i> .....	100
2.3.3	Expression of FADH and FALDH genes.....	100
2.3.3.1	Northern Hybridization Analysis of FALDH and FADH genes...	100
2.3.3.2	Analysis of ALDH/ FALDH expression by RT-PCR.....	103
2.4	Conclusions.....	104
	References.....	105
 <b>Chapter Three – Gene Disruption and Expression analysis of Fatty Aldehyde Dehydrogenases of <i>Yarrowia lipolytica</i></b> .....		
3.1	Introduction.....	110
3.2	Materials and Methods.....	111
3.2.1	Plasmids, strains and media.....	111
3.2.2	General molecular biology techniques.....	111
3.2.3	Construction of disruption cassettes.....	113
	Construction of the promoter-terminator (PT) cassette .....	113
	Construction of the final promoter-URA3-terminator (PUT) cassette.....	113
	Deletion of FALDH genes.....	113
3.2.4	Ura3 marker rescue by expression of Cre recombinase.....	114

3.2.5	Deletion Strategy.....	115
3.2.6	Growth analysis of FALDH disruption mutants.....	116
3.2.7	Dry weights determination.....	117
3.2.8	FALDH enzyme activity of disruption mutants.....	117
3.3	Results and Discussion.....	118
3.3.1	Construction of the <i>FALDH-PUT</i> cassettes.....	118
3.3.2	Verification of correct disruption of FALDH genes.....	120
3.3.3	Ura3 marker rescue by expression of Cre recombinase.....	122
3.3.4	Growth analysis of mutants.....	126
3.4	Conclusions.....	129
	References.....	131
<b>Chapter Four - Putative fatty aldehyde dehydrogenase-encoding genes</b>		
<b>from the sequenced fungal genomes.....</b>		
	Abstract.....	135
	Introduction.....	135
	Methods.....	139
	BLAST search of fungal genomes for putative FALDHs.....	139
	Multiple sequence alignment and phylogenetic analysis.....	139
	Identification of conserved ALDH regions from BLAST hits.....	139
	Domain analysis of identified fungal FALDHs.....	140
	Results and Discussion.....	141
	BLAST search of fungal genomes for putative fungal FALDHs.....	141
	Multiple sequence alignment and phylogenetic analysis.....	145
	Identification of conserved residues and motifs .....	152
	(a) Conserved Motifs.....	152
	(b) Conserved residues.....	154
	Domain analysis of the putative fungal FALDHs.....	156
	Conclusion.....	160
	References.....	161
<b>Chapter Five – General Discussion and Conclusions.....</b>		
		170

References.....	176
<b>SUMMARY</b> .....	179
<b>APPENDIX A</b> .....(Accession numbers for FALDH sequences used in construction of phylogenetic tree).....	181
<b>APPENDIX B</b> .....(?-DNA/EcoRI+HindIII molecular weight marker).....	185
<b>APPENDIX C</b> .....(website references used).....	185

## List of abbreviations

3-D structure – three dimensional structure

ABTS - 2,2' azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)

ADH – alcohol dehydrogenase

ALDH – aldehyde dehydrogenase

cDNA - complementary DNA

CoA – coenzyme A

C-terminal – carboxyl terminal

DADH - *Drosophila* ADH

DMSO – dimethyl sulphonyl oxide

DNA – deoxyribonucleic acid

dNTP- deoxyribonucleotide triphosphate

EC number - enzyme commission number

FADH – fatty (long chain) alcohol dehydrogenase

FALDH – fatty (long chain) aldehyde dehydrogenase

FAOD - fatty (long chain) alcohol oxidase

G-3-PDH - glyceraldehyde-3-phosphate dehydrogenase

GGsALDH -  $\gamma$ -glutamyl semialdehyde dehydrogenase

HGNC - Human Genome Nomenclature Committee

HUGO – The Human Genome Organisation

ICGSN – International Committee on Genetic Symbols and Nomenclature

IUB - International Union of Biochemistry

IUPAC - The International Union of Pure and Applied Chemistry

JCBN - Joint Commission on Biochemical Nomenclature

kDa – kilodalton

LADH - horse liver ADH

MDR - medium chain dehydrogenases/reductases

mRNA – messenger RNA

NADP<sup>+</sup> - Nicotinamide Adenine Dinucleotide Phosphate

NAD<sup>+</sup> - Nicotinamide Adenine Dinucleotide

N-terminal – amino terminal

NC-IUBMB - Nomenclature Committee of the international Union of Biochemistry  
and Molecular Biology

OD<sub>620nm</sub> – optical density at a wavelength of 620nm

PCR – polymerase chain reaction

PF No - product family number

RNA – ribonucleic acid

RT-PCR- reverse transcription-polymerase chain reaction

SDS –sodium dodecyl sulphate

SDR - short-chain dehydrogenases/reductases

SLS - Sjögren-Larsson syndrome

SSC- sodium chloride-sodium citrate

Tris-HCl – Tris (hydroxymethyl) aminomethane Hydrochloride

YNB – yeast nitrogen base

## List of figures

### Chapter one - Aldehyde dehydrogenases

<b>Figure 1.1.1:</b> Reaction catalyzed by aldehyde dehydrogenases.....	1
<b>Figure 1.1.2:</b> Mechanism action of aldehyde dehydrogenase.....	4
<b>Figure 1.1.3:</b> The phylogenetic relationships among ALDH subfamilies.....	18
<b>Figure 1.1.4:</b> The 3-dimensional structure of rat ALDH3.....	25
<b>Figure 1.1.5:</b> C-terminal sequences of mouse and human major FALDHs.....	27

### Chapter one - Alcohol dehydrogenases

<b>Figure 1.2.1:</b> The overall reaction catalyzed by alcohol dehydrogenases.....	32
<b>Figure 1.2.2:</b> A ribbon diagram of horse liver alcohol dehydrogenase .....	39
<b>Figure 1.2.3:</b> A ribbon diagram of horse liver ADH dimer illustrating the nature of interaction surfaces of the monomers prior to dimerization.....	40
<b>Figure 1.2.4:</b> The 3-D structure of horse liver ADH showing coenzymes and cofactors important to functioning of the enzyme.....	43
<b>Figure 1.2.5:</b> Mechanism of action of alcohol dehydrogenase.....	44
<b>Figure 1.2.6:</b> Schematic representation of mechanism of action of alcohol dehydrogenase.....	44
<b>Figure 1.2.7:</b> Phylogenetic relationship of human and mouse ADH genes .....	47

### Chapter two

<b>Figure 2.1:</b> Induction of FALDH and FADH activities in <i>Y. lipolytica</i> H222 cells during batch culture. ....	90
<b>Figure 2.2:</b> A comparison of FAOD induction in relation to wet cell mass in <i>Y. lipolytica</i> H222 cells growing in hexadecane. ....	91
<b>Figure 2.3:</b> Subcellular fractionation of FADH, FALDH and FAOD.....	92
<b>Figure 2.4:</b> Comparison of the SFA protein sequence with the partial FADH amino acid sequence of <i>Y. lipolytica</i> .....	96
<b>Figure 2.5:</b> PCR product of partial FADH gene sequence of <i>Y. lipolytica</i> .....	97
<b>Figure 2.6:</b> Alignment of full FADH protein sequence of <i>Y. lipolytica</i> against FADHs of other yeasts.....	98

<b>Figure 2.7:</b> Southern blot analysis of FAOD gene(s) of <i>Y. lipolytica</i> .....	100
<b>Figure 2.8:</b> Northern Hybridization Analysis of FALDH and FADH genes.....	102
<b>Figure 2.9:</b> Expression Analysis of ALDH, FALDH and FADH genes by RT-PCR.....	103

### Chapter three

<b>Figure 3.1:</b> The strategy for construction of the FALDH deletion cassettes.....	115
<b>Figure 3.2:</b> Integration of the deletion cassettes into the yeast genome.....	116
<b>Figure 3.3:</b> Ethidium bromide stained gels depicting the construction of <i>FALDH</i> 3 and 4 deletion cassettes.....	119
<b>Figure 3.4:</b> The PCR products for construction of PUT cassettes for FALDH1 and FALDH2.....	120
<b>Figure 3.5:</b> PCR products obtained for verification of correct disruption of FALDHs using primers FALDH- FIM/Ura3-RIM.....	121
<b>Figure 3.6:</b> EcoRI digest of PCR products of the deletion cassettes from positive clones.....	121
<b>Figure 3.7:</b> Recovery of the <i>URA3</i> and <i>LEU2</i> markers. ....	123
<b>Figure 3.8:</b> A scheme of sequential gene disruption of the <i>Y. lipolytica</i> <i>FALDH</i> genes.....	124
<b>Figure 3.9:</b> Growth analysis of <i>Y. lipolytica</i> FALDH deletion mutants on YPD and YNB-hexadecane agar plates.....	125
<b>Figure 3.10:</b> A comparison of growth of FALDH deletion mutants on glucose and octadecane .....	126
<b>Figure 3.11:</b> Growth of FALDH deletion mutants in alkanes.....	127
<b>Figure 3.12:</b> Cellular protein content of FALDH mutants.....	128

### Chapter four

<b>Figure 4.1:</b> A phylogenetic tree showing the position of the 25 putative fungal FALDH sequences in relation to other ALDHs.....	146
<b>Figure 4.2:</b> DNAssist multiple sequence alignment of the 25 FALDH/FALDH-like protein sequences and other class 3 ALDH proteins .....	151

## List of tables

### Chapter one - Aldehyde dehydrogenases

<b>Table 1.1.1:</b> Reactions catalyzed by aldehyde dehydrogenases.....	9
<b>Table 1.1.2:</b> The four categories of the ALDH gene superfamily .....	11
<b>Table 1.1.3:</b> Summary table of the different ALDH families.....	19
<b>Table 1.1.4:</b> Summary table of the ALDH gene superfamily.....	20
<b>Table 1.1.5:</b> The ten most conserved sequence motifs in ALDHs.....	24
<b>Table 1.1.6:</b> A summary of fungal ALDH genes.....	30

### Chapter one - Alcohol dehydrogenases

<b>Table 1.2.1:</b> The four classes of alcohol dehydrogenases.....	34
<b>Table 1.2.2:</b> The new ADH nomenclature for human and mouse.....	46

### Chapter two

<b>Table 2.1:</b> Oligonucleotide sequence of primers used in chapter 2.....	85
<b>Table 2.2:</b> FALDH activity levels in hexadecane and glucose grown cells using NAD and NADP as cofactors.....	94
<b>Table 2.3:</b> A comparison of the four putative ALDHs of <i>Y. lipolytica</i> with other ALDHs.....	99

### Chapter three

<b>Table 3.1:</b> Organisms, strains and plasmids used in chapter 3.....	112
<b>Table 3.2:</b> Oligonucleotide sequence of primers used in chapter 3.....	114

### Chapter four

<b>Table 4.1:</b> Fungi genome projects report.....	141
<b>Table 4.2:</b> BLAST search of fungal genome databases.....	143
<b>Table 4.3:</b> The ten most conserved sequence motifs in the 25 putative fungal FALDH sequences.....	153
<b>Table 4.4:</b> Table summary of the position and number of transmembrane segments (TMS) in the 25 fungal FALDH/FALH-like proteins.....	158

## PROLOGUE

Many yeasts have the ability to utilize long-chain *n*-alkanes. Most studies on the biochemistry and genetics of the alkane-degradation pathway in yeasts have focused on cytochrome P450 monooxygenases. Consequently, information on the fatty alcohol oxidases (FAOD) and fatty alcohol dehydrogenases (FADH) that should be responsible for oxidation of alcohols to aldehydes and of the fatty aldehyde dehydrogenases (FALDHs) that should be responsible for oxidation of aldehydes to fatty acids is very limited.

Our group is interested in genetic engineering of the alkane degradation pathway in *Yarrowia lipolytica* that might lead to accumulation of long chain alcohols from alkanes. Through collaboration with Dr. Jean-Marc Nicaud of the INRA-CNRS at Grignon, France, we have access to an array of cloning systems which they have developed. We also had since the beginning of 2000 access to new sequence information as it became available from the *Y. lipolytica* sequencing project, which was completed at the beginning of 2004.

This project, which aimed at the identification of fatty alcohol and aldehyde dehydrogenases in *Y. lipolytica*, developed in three stages. Initially (Chapter 2) we tried to establish through enzyme activity assays, BLAST searches, Southern and Northern blot analyses and RT-PCR the presence and induction during growth on alkanes of FAODs, FADHs and FALDHs in *Y. lipolytica*. Only in the case of the FALDHs could we establish through activity assays, BLAST searches, Northern blot analyses and RT-PCR presence and involvement in alkane assimilation. BLAST searches of the completed genome sequence database eventually revealed the presence of four putative FALDH encoding genes. In the second stage of the project (Chapter 3) all four of these FALDH genes were deleted using the Cre-Lox P recyclable tools, which enable marker rescue allowing simultaneous deletion of several members of a gene family by use of a recyclable single marker (Fickers *et al.*, 2003). Fifteen mutants with the four FALDHs deleted in all possible combinations were constructed. Time did not

allow thorough phenotypic characterization of these FALDH deleted mutants, but a few growth studies were carried out on the quadruple deletion and triple deletion strains. Finally BLAST searches for the FALDH encoding genes in *Y. lipolytica* lead to the realization that there are available from the NCBI databases putative FALDH or class 3 ALDH encoding genes that represent new families of fungal class 3 ALDHs. A comparative study of these sequences was written up as a paper that is ready to be submitted for publication (Chapter 4).

# CHAPTER 1

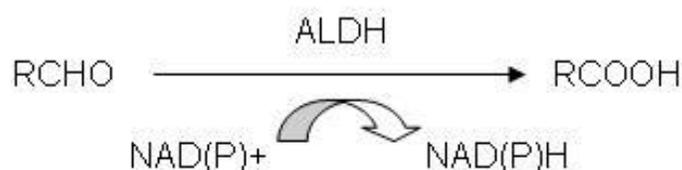
## Aldehyde- and alcohol dehydrogenases - a literature review

### 1.1 ALDEHYDE DEHYDROGENASES

#### Introduction

Aldehydes are found in living cells as physiologically derived intermediates in the metabolism of other compounds. Some aldehyde-mediated effects are beneficial, such as that from retinaldehyde in vision, but many other effects such as cytotoxicity, mutagenicity, genotoxicity and carcinogenicity are harmful. Thus selective elimination of aldehydes from biological systems is important for maintenance of healthy living cells, because aldehydes are chemically reactive molecules. A variety of enzymes have evolved whose function is the detoxification of aldehydes, and these enzymes are called aldehyde dehydrogenases (E.C. 1.2.1.3, ALDH).

Aldehyde dehydrogenases (ALDHs) comprise a diverse set of enzymes, which catalyze the NAD(P)<sup>+</sup>-dependent oxidation of aldehydes in a virtually irreversible reaction as shown in figure 1.1.1 (Liu *et al.*, 1997) .



**Figure 1.1.1: Reaction catalyzed by aldehyde dehydrogenases.**

The physiological importance of ALDHs is manifest in several autosomally inherited diseases resulting from ALDH deficiencies. The Sjörgen-Larsson syndrome, an inborn neurologic impairment, characterized by mental retardation and spasticity, results from a mutation in the human fatty aldehyde dehydrogenase. Due to the defective enzyme, there is accumulation of long

chain aldehydes, which subsequently react with other compounds to give the previously described symptoms. Another autosomal recessive disorder known as type II hyperprolinemia, results from the loss of  $\gamma$ -glutamyl semialdehyde dehydrogenase function. Patients suffering from this disorder exhibit high plasma levels of proline and  $\Delta^1$ -pyrroline-5-carboxylate and may suffer from mental retardation and seizures. Succinic semialdehyde dehydrogenase deficiency results in intracellular accumulation of succinic semialdehyde and an increase in 4-hydroxybutyrate in physiological fluids, which in turn affects the central nervous system, causing altered motor activity and speech delay. In plants, a putative ALDH similar to mammalian ALDH has been shown to be the nuclear restorer protein of male-sterile T-cytoplasm maize, Rf2. ALDH has also been shown to play an important role in the metabolism and development of pheromones in insects (Tasayco and Prestwich, 1990). It is thus clear and obvious that ALDHs are necessary and essential components of cells for a variety of reasons.

## **THE ALDEHYDE DEHYDROGENASE STRUCTURE**

The first crystal structure of an ALDH (rat class 3 ALDH) was described by Liu *et al.*, (1997a) at a resolution of 2.6Å. The active form of the enzyme is a dimer, consisting of two identical subunits. Each subunit consists of an NAD(P)-binding domain, a catalytic domain and an “arm-like” bridging domain. At the interface of these domains is a long funnel-shaped passage with an opening leading to a putative catalytic pocket.

The core of the coenzyme-binding domain in dehydrogenase enzymes is made up of a common structural motif comprising a parallel left-handedly twisted  $\beta$ -sheet (Liu *et al.*, (1997b). The  $\beta$ -sheet includes two supersecondary structural elements each made up of three parallel  $\beta$ -strands connected by  $\alpha$ -helices,  $\beta$ -strands or irregular loops, the so-called Rossmann fold (Kutzenko *et al.*, 1998). The ALDH structure exhibits a totally new and unexpected mode of NAD interaction, which is different from the classic mode of NAD-binding Rossmann

folds, observed in all other dehydrogenase families (Liu *et al.*, 1997a; Hempel *et al.*, 1999). In ALDH the Rossmann fold contains five  $\beta$ -strands connected by four  $\alpha$ -helices, instead of the usual six  $\beta$ -strands found in other NAD-dependent dehydrogenases. Also, in contrast to the classic  $\beta$ - $\alpha$ - $\beta$  motif found in other NAD-dependent dehydrogenases a novel mode of NAD-binding,  $\beta$ - $\alpha$ , $\beta$  motif is observed in ALDH. Moreover, the GxGxxG motif (where x denotes any amino acid) usually found associated with the  $\beta$ 1- $\alpha$ A loop is found in ALDH to be in the  $\beta$ 4- $\alpha$ D loop, as GxTxxG and associated with the nicotinamide ring. However, the role of  $\beta$ 2 in NAD-binding specificity is still retained, resulting in the unique ALDH  $\beta$ - $\alpha$ , $\beta$  binding motif.

In the ALDH family of enzymes Cys<sup>243</sup> in class 3 (Cys<sup>302</sup> in class 1 and 2) ALDHs, is strictly conserved and has generally been accepted as the catalytic thiol (Liu *et al.*, 1997). (The residue numbering refers to rat liver microsomal (class 3), human liver mitochondrial (class 2) and cytosolic (class 1) ALDHs). The orientation of the catalytic thiol within the funnel-shaped passage agrees well with the common chemical mechanism of hydride transfer from aldehyde to NAD, and thus this funnel is interpreted as the catalytic pocket (Liu *et al.*, 1997b). The catalytic pocket, which is lined with several highly conserved amino acid residues lies at the bottom of the passage between the catalytic thiol and the nicotinamide ring. The bridging domain forms part of the mouth of the catalytic passage and it plays an important role in the formation and stabilization of the ALDH dimer. Despite some large differences in sequences, mode of aggregation and substrate specificity, conserved residues are still found at key locations within the ALDH structure (Liu *et al.*, 1997b). This therefore implies that in all classes of ALDH the overall structural fold, the novel NAD-binding motif as well as the catalytic site environment are still maintained.

Even though ALDH is a classic  $\beta$ - $\alpha$ - $\beta$  protein, it does not contain an iron-sulphur cluster, uses NAD(P) as a substrate and does not require a molybdoterin-based cofactor (Liu *et al.*, 1997b). As such its biochemical properties and structural fold

are different from other  $\beta$ - $\alpha$ - $\beta$  group of enzymes, making the ALDH family a unique class of molecules, designed to control accumulation of aldehydes in possibly all tissues in the biological system (Liu *et al.*, 1997b).

### Aldehyde dehydrogenase reaction mechanism

ALDH oxidizes aldehydes to their corresponding carboxylic acids in the presence of NAD(P). The catalytic cycle involves several individual reactions, namely nucleophilic attack, hydride transfer and (de)acylation (Wymore *et al.*, 2001, 2004, <http://www.psc.edu/biomed>). Quantum/molecular mechanistic studies of class 3 ALDH reaction mechanism by Wymore *et al.*, (2004; available at [http://www.psc.edu/science/2002/wymore/what\\_happens\\_at\\_the\\_active\\_site.html](http://www.psc.edu/science/2002/wymore/what_happens_at_the_active_site.html)) show that the substrate binds in an orientation that produces a protonated thiohemiacetal in the R-configuration.

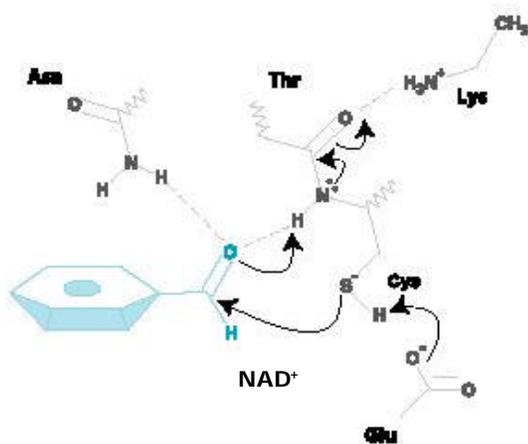


Figure 1.1.2: A schematic drawing for the proposed mechanism of the action of aldehyde dehydrogenase. Benzaldehyde is the substrate in this example (Wymore *et al.*, 2002, [http://www.psc.edu/science/2002/wymore/what\\_happens\\_at\\_the\\_active\\_site.html](http://www.psc.edu/science/2002/wymore/what_happens_at_the_active_site.html)).

The sulfhydryl proton on Cys<sup>243</sup> is important for initially recognizing the substrate. Substrate binding then causes a shift in the hydrogen bonding partner for Lys<sup>235</sup> in which it then primarily interacts with the carbonyl oxygen of Thr<sup>242</sup>, an interaction which is important for formation of the thiohemiacetal intermediate.

The first intermediate forms when an active site cysteine bonded to hydrogen binds with the aldehyde. The cysteine's sulphur binds with the aldehyde carbon upon which it is protonated and later deprotonated through interaction with a nearby amino acid group. The enzyme donates a proton to the intermediate state simultaneously with the cysteine's sulphur binding to the aldehyde. The donor protein actually comes from the nitrogen atom of the enzyme next to cysteine. In other words, formation of the hemiacetal intermediate is a concerted proton transfer from the Cys<sup>243</sup> amide backbone supported by interactions with Lys<sup>235</sup>. The mammalian (rat) ALDH3 reaction mechanism is a novel enzyme mechanism, where the enzyme backbone rather than the side chains do all the chemistry. The ALDH3 reaction mechanism during oxidation of benzaldehyde is shown in figure 1.1.2.

### **Substrate specificity and coenzyme preference in ALDHs**

In any protein molecule the role of the catalytic domain is to provide residues essential for catalysis and substrate coordination. The rat class 3 ALDH catalytic domain is described by Liu *et al.*, (1997) as a funnel-shaped pocket. The upper portion of the funnel is formed by 14 residues from the coenzyme-binding domain, 14 residues from the catalytic domain and 7 residues from the bridging domain. It is this upper portion of the funnel, which provides the required ALDH specificity towards a particular aldehyde. A change in protein structure resulting from substituting one residue with another alters the shape of the substrate binding pocket and substrate specificity thereupon (Perozich *et al.*, 2000). For example, a mutation of Lys<sup>192</sup> to Glu in class 2 ALDHs was found to result in change of substrate specificity from aliphatic to aromatic aldehydes.

Of all the 35 amino acid residues lining the surface of the upper funnel only one (Phe<sup>401</sup>) is highly conserved. The lower portion of the funnel, which is positioned between the catalytic thiol and the nicotinamide NC4, is lined with 10 highly conserved residues (Asn<sup>114</sup>, Thr<sup>186</sup>, Glu<sup>209</sup> and Leu<sup>210</sup> are highly conserved; Gly<sup>187</sup>, Gly<sup>211</sup>, Cys<sup>243</sup>, Glu<sup>333</sup> and Phe<sup>335</sup> are strictly conserved). This is the site

where catalytic hydride transfer from aldehyde to coenzyme takes place. Shifts in conformation of the active site resulting from aldehyde binding result in pro-R specificity, which in fact has been shown to be stereospecific for class 1, 2 and 3 aldehydes (Liu *et al.*, 1997).

The pyridine-nucleotide dependent enzyme families, of which ALDH is a member, are generally known for their strict specificity for either NAD or NADP. NAD-specific enzymes generally catalyze oxidative, anabolic reactions while NADP-specific ones are involved in reductive, anabolic roles (Perozich *et al.*, 2000). The extended family of ALDHs [aldehyde: NAD(P)<sup>+</sup> oxidoreductases, EC 1.2.1] is mostly composed of ALDHs that are specific for NAD (EC 1.2.1.3). Only class 3 ALDHs exhibit a well established dual coenzyme preference, whereas the coenzyme preference of a few others has yet to be sufficiently characterized (Perozich *et al.*, 2000).

As already mentioned ALDHs exhibit an altered mode of coenzyme binding to a Rossmann fold not observed in other NAD-dependent dehydrogenases (Liu *et al.*, 1997). In ALDHs the coenzyme binding domain is a 5-stranded open  $\alpha/\beta$  domain with an extended loop between  $\beta$ -1 and  $\alpha$ -A, and the coenzyme binds between 2 helices,  $\alpha$ -C and  $\alpha$ -D (Perozich *et al.*, 2000). Unlike in other classical NAD-binding proteins, in ALDHs the pyrophosphate moiety of the NAD binding domain is away from and does not interact with the dinucleotide binding helix ( $\alpha$ -A). Instead the NAD nicotinamide ring is oriented and stabilized by stacking between Gly<sup>187</sup> and Phe<sup>335</sup>, as well as by hydrogen bonding with Arg<sup>292</sup>, from  $\alpha$ -11 of the catalytic domain, to NO<sub>3</sub><sup>\*</sup> of the nicotinamide ribose (Liu *et al.*, 1997). The NAD binding motif in ALDHs involves the loop region between  $\beta$ 4,  $\alpha$ D and  $\beta$ 2 thus giving it the name  $\beta$ 4- $\alpha$ D- $\beta$ 2 or  $\beta$ - $\alpha$ , $\beta$ ) motif instead of the classic  $\beta$ 1- $\alpha$ A- $\beta$ 2 (or  $\beta$ - $\alpha$ - $\beta$ ) motif observed in other NAD-dependent oxidoreductases.

The glycine-rich sequence, G<sub>1</sub>-xxx-G<sub>2</sub> (residues 187-192) equivalent to the NAD-binding motif G<sub>1</sub>xG<sub>2</sub>xx in other NAD-dependent dehydrogenases is found at the

end of  $\alpha$ A. About 21 positions downstream of this sequence motif lies an acidic residue Glu<sup>140</sup>, which is the NAD recognition residue in NAD-specific ALDHs (Perozich *et al.*, 2001). This residue coordinates the 2'- and 3'- adenosine hydroxyls of NAD and repels the 2'-phosphate of NADP. In NADP-specific ALDHs the acidic residue is replaced by a polar residue, which hydrogen bonds to the 2'-phosphate of NADP, thus conferring NADP-dependence. For example, in the NADP-specific FALDH of *Vibrio harveyi* the polar residues are Thr<sup>175</sup> (Zhang *et al.*, 2001) and Thr<sup>180</sup> in GAPDH (Cobessi *et al.*, 1999).

Class 3 ALDHs however, are able to bind both NAD and NADP though Glu<sup>140</sup> is still present. This may therefore imply that the presence of Glu<sup>140</sup> in ALDHs does not guarantee NAD specificity as in other oxidoreductases. No clear explanation is yet available about this behaviour of class 3 ALDHs but it has been observed that these ALDHs bind NADP in a conformation different to NAD, which is not found in other ALDHs (Perozich *et al.*, 2000). Perozich *et al.*, (2001) further hypothesize that the larger space of the adenine-binding cleft in class 3 ALDHs, accompanied by a conformational change in the loop between  $\beta$ -3 and  $\alpha$ -C upon binding NADP help accommodate the 2'-phosphate of NADP in class 3 ALDHs.

## **CLASSIFICATION OF ALDEHYDE DEHYDROGENASE ENZYMES**

### **Reaction-Chemistry-Directed approach**

ALDHs have been identified in virtually every living organism studied, and have been found to exist in multiple forms, which differ in physical and/or functional properties (Weiner, 1979). For example, mammalian ALDHs found in the liver, stomach, kidney, eye and brain exist as distinct enzymes differing not only in locations but also with substrate specificities (Liu *et al.*, 1997). Some ALDHs are constitutive while others are inducible. Some forms have broad substrate specificity and oxidize a variety of aliphatic and aromatic aldehydes, while some forms display a narrow substrate preference, and utilize small aliphatic aldehydes only (Lindahl and Hempel, 1991).

Due to limited data on structural and functional relationships among the various aldehyde dehydrogenases it was not possible in the past to classify all known ALDHs into distinct classes. Traditionally, only three groups, namely class 1, 2 and 3 ALDHs, have received most attention while many other ALDH families with various metabolic roles remained unattended.

The ALDH family consists of many and diverse enzyme families. In the past several methods of classification of ALDHs have been suggested but none of them ever held for long as a result of continuous discovery of many additional enzymes in almost every organism studied. In addition, the physiological substrates for most ALDHs being discovered could not be identified, and thus their primary biological role was not clear. The first ALDH classification called "Workshop on Aldehyde Dehydrogenase, 1988" excluded several enzymes now fully known to be members of this class of enzymes; namely semialdehyde, steroid aldehyde (aldosterone), vitamin aldehyde (retinaldehyde), and formaldehyde dehydrogenases (Shah and Pietruszko, 1999). The next classification attempt, which was based on primary structure (Vasiliou, *et al.*, 1995), also excluded some enzymes such as aspartic semialdehyde dehydrogenase, phosphorylating glyceraldehyde-3- dehydrogenase and formaldehyde dehydrogenase.

Following this, Shah and Pietruszko (1999) attempted a chemistry oriented approach, in which they thought that one way of grouping such a diverse variety of enzymes was to look at the reactions they catalyzed. The motivation being that previous attempts to classify these enzymes based on such models as primary structure, positional identity and analysis of the sequence alignments tended to exclude some members in a group as already seen. It was therefore hoped that attention to the reaction catalyzed was an easier classification process comparable to the classification of alcohol dehydrogenases and other similar enzymes. It was also hoped that this method of classification would also

be informative and thus very helpful in elucidation of the enzymes' mechanism of action.

According to the reaction-chemistry-directed approach, five distinct groups of reactions catalyzed by ALDHs were recognizable as shown in Table 1.1.1. All these reactions are NAD(P)-dependent. Generally speaking, Group 1 enzymes catalyze a unidirectional and irreversible reaction in the presence of water and NAD(P), reaction 5 though reversible also requires water, while all the other enzymes do not need water and catalyze a reversible reaction for which additional co-factors such as co-enzyme A, glutathione or adenosine monophosphate are needed.

**Table 1.1.1: Reactions catalyzed by aldehyde dehydrogenases**

Reaction No.	Reaction catalyzed
1	Aldehyde + NAD(P) + H <sub>2</sub> O $\longrightarrow$ Acid + NAD(P)H + H <sup>+</sup>
2	Aldehyde + NAD(P) + CoA $\longrightarrow$ Acyl-CoA + NAD(P) + H <sup>+</sup>
3	Aldehyde + orthophosphate + NAD (P) $\longrightarrow$ Acyl phosphate + NAD(P)H + H <sup>+</sup>
4	Formaldehyde + Glutathione + NAD $\longrightarrow$ S-formylglutathione + NADH + H <sup>+</sup>
5	Aryl aldehyde + NADP + AMP + pyrophosphate + H <sub>2</sub> O $\longrightarrow$ Aromatic acid + NADPH + ATP + H <sup>+</sup>

### Substrate specificity-based classification

While no one has disputed the reaction-chemistry-directed ALDH classification approach, it seems to have gone unnoticed by those involved in this line of research, since very few papers if any at all refer to this work. According to Shah and Pietruszko (1999) substrate specificity of ALDHs could not be used as a classification criterion because of the fact that different ALDH enzymes utilize the same substrates. For example, glyceraldehyde-3-phosphate in reactions 1 and 3

in Table 1.1.1 above. However, during the same period the reaction-chemistry-directed ALDH classification approach was proposed, a sequence alignment of all the then available ALDH protein sequences was done so as to aid in determining ALDH structure and establishing relationships between various ALDH families (Perozich *et al.*, 1999). Due to the results from the alignment of the 145 ALDHs any previous perceptions on ALDH classification again changed.

Consequently, the widely accepted method of classification now is based on substrate specificity. Indisputably, substrate specificity of the ALDH superfamily is very broad with members oxidizing a variety of both aliphatic and aromatic aldehydes, whereas other forms exhibit narrower substrate preferences, hence why it was not so obvious in the past that substrate preference could be a useful character for classification of the ALDH superfamily. ALDH enzymes are now broadly grouped into four categories, as shown in table 1.1.2. The four main ALDH enzymatic functions are presently recognized as 1) detoxification, 2) intermediary metabolism, 3) osmotic protection and 4) NADPH generation, as well as some function in structural capacity (Perozich *et al.*, 1999).

At least two main classes of non-specific, variable substrate ALDHs can be distinguished; namely Class 1/2 and Class 3 ALDHs (Yoshida *et al.*, 1998). Class 1/2 ALDHs are cytosolic or mitochondrial tetrameric enzymes involved in detoxification and metabolism of acetaldehyde and other dietary aldehydes, xenobiotics, lipid peroxidation products and certain anti-cancer drugs. Malfunction of these enzymes is associated with susceptibility to ethanol-related diseases. The Class 3 ALDHs are cytosolic or microsomal dimeric enzymes associated with oxidation of aromatic aldehydes and fatty aldehydes (medium-chain aliphatic aldehydes) and are also associated with carcinogenesis and severe genetic disorders.

**Table 1.1.2: The four categories of the ALDH gene superfamily based on substrate specificity.**

Semialdehyde dehydrogenases	Nonspecific ALDHs	Other ALDHs	ALDH-like proteins
	Commonly known as		
➤ Semialdehyde dehydrogenases (E.C. 1.2.1.32)	➤ Class 1	➤ Betaine dehydrogenase (E.C.1.2.1.8)	➤ 10-formyltetrahydrofolate dehydrogenase (E.C.1.2.1.6)
➤ Escherichia coli (E.C. 1.2.1.16) and mammalian (E.C. 1.2.1.24) succinate-semialdehyde dehydrogenase	➤ Class 2	➤ Non phosphorylating glyceralddehydes 3-phosphate dehydrogenase (E.C. 1.2.1.9)	➤ $\Delta^1$ -pyrroline-5-carboxylate (E.C.1.2.1.12)
➤ Glutamate semialdehyde dehydrogenase (E.C. 1.2.1.41)	➤ Class 3	➤ Phenylacetaldehyde dehydrogenase (E.C.1.2.1.39)	➤ Antiquitin
➤ Aspartate semialdehyde dehydrogenase (E.C. 1.2.1.11)		➤ Methylmalonate semialdehyde dehydrogenase (E.C. 1.2.1.27)	➤ Human 56-kDa androgen-binding protein
➤ 2-amino-adipate-6-semialdehyde dehydrogenase (E.C. 1.2.1.31)			➤ Crystallins
➤ Methylmalonate- semialdehyde dehydrogenase (E.C. 1.2.1.27)			

The substrate specific semialdehyde dehydrogenases are involved in the majority of basic metabolic pathways, especially biosynthesis of various amino acids. For example, aspartate-semialdehyde dehydrogenases are involved in the biosynthesis of various amino acids from aspartate, and glutamate semialdehyde dehydrogenases are required for the biosynthesis of arginine. The glutamate semialdehyde dehydrogenases and methylmalonate semialdehyde

dehydrogenases are also both involved in amino acid metabolism (Perozich *et al.*, 1999).

Expression of ALDH-related genes in response to osmotic stress, including dehydration and high salinity, has been reported in many organisms. Under osmotic stress, betaine aldehyde dehydrogenase (BADH) oxidizes glycine betaine aldehyde into the osmoprotectant glycine betaine. Several BADH genes have been characterized from several plant species (Kirch *et al.*, 2001) and bacteria (Rosenstein *et al.*, 1999), and BADH proteins show about 40% homology to variable substrate ALDHs at amino acid level. In response to osmotic turgor and during fruit ripening, another group of ALDH-like genes, the turgor-responsive genes, is induced in plants (Yamada *et al.*, 1999). Turgor-responsive ALDHs are remarkably similar to the human antiquitin, and display about 30% identity to various ALDHs, but their function is unknown (Lee *et al.*, 1994). It is proposed that since turgor-responsive ALDHs are found in the “Amino Acid Intermediate” sub-branch of the “Class 3” trunk it is possible that they may function in similar metabolic pathways (Perozich *et al.*, 1999).

The cytosolic non-reversible non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (G3PDH) is the key contributor of the NADPH required for photosynthetic reactions (Gao and Loescher, 2000). Occurrence of G3PDH is a specific feature of those organisms with chloroplasts or cyanelles, and sequence comparisons indicate that G3PDH is a member of the ALDH superfamily but with no relationship to the phosphorylating G3PDH found in the chloroplast and cytosol. It has been proposed that in green leaf tissues G3PDH is a component of a photosynthetic shuttle transferring reducing equivalents from the chloroplast to the cytosol so that the reductant generated as such may be used to meet several biosynthetic requirements including mannitol biosynthesis (Gao and Loescher, 2000).

It has increasingly become evident that some ALDH proteins exhibit other functions in addition to their catalytic properties (Sophos *et al.*, 2003). These proteins are grouped together in the category ALDH-like proteins, and they are found to be involved in structure and development, as well as in protein binding. Members of this group include fusion proteins that have an ALDH domain. For example, the mammalian formyltetrahydrofolate dehydrogenase is a large cytosolic fusion protein of about 900 residues with three domains; the carboxy-terminal (480 amino acids) of which is structurally and functionally related to class 1 and 2 ALDHs (Perozich *et al.*, 1999). Another member of this group is the bacterial multifunctional putA proteins, which have a  $\gamma$ -glutamyl semialdehyde dehydrogenase domain fused to a proline residue. The yeast *PUT2* gene, encoding  $\Delta^1$ -pyrroline-5-carboxylate dehydrogenase, which converts proline to glutamate, is also a member of this group. Protein-binding members of this group include the androgen-binding protein in human genital fibroblasts, thyroid hormone-binding protein in *Xenopus* liver, sterol-binding protein in bovine lens epithelial cells, flavopiridol-binding protein in non-small cell lung carcinomas, dianorubicin-binding protein in rat liver and benzopyrene-binding protein in mouse liver (Sophos *et al.*, 2003). Other members of the ALDH-like group include the maize *r2* gene known as the nuclear restorer and antiquitin, which may function in regulation of turgor pressure and/or general stress response.

Finally,  $\Omega$ ,  $\eta$  and  $\omega$ -crystallins have been identified as minor structural components of cephalopods and shrews eye lens (Zinovieva *et al.*, 1993). Also in vertebrates crystallins are found as major cornea and lens proteins responsible for the structural integrity and functional utility of these visual tissues (Cooper *et al.*, 1993). Crystallins are evolutionarily related to ALDHs but have lost their catalytic activity (Jornvall *et al.*, 1997) which may suggest that these proteins evolved by duplication of an ancestral gene encoding ALDH and subsequently specialized for light refraction while losing ALDH activity and expression in other tissues (Zinovieva *et al.*, 1993).

## NOMENCLATURE OF ALDEHYDE DEHYDROGENASES

When the fields of biochemistry and enzymology exploded in the 1950s and 60s, the International Union of Biochemistry (IUB) deemed it necessary to establish a uniform numbering system for all enzymes called the Enzyme Commission number, "EC number" (IUB, 1973). The EC number for all enzymes, independent of species, contains four numbers separated by periods (e.g. 1.1.1.1 for alcohol dehydrogenase; 1.2.1.3 for aldehyde dehydrogenase), classifying the enzyme by class, subclass and sub-subclass. This is today a commonly accepted enzyme naming system.

In 1993 the Nomenclature Committee of the international Union of Biochemistry and Molecular Biology (NC-IUBMB), the International Union of Pure and Applied Chemistry (IUPAC), and the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) came up with the systematic naming of genes across all species, based on divergent evolution of gene superfamilies (<http://www.uchsc.edu/sp/sp/alcdbase/aldhcov.html>; Nelson *et al.*, 1993, 1996; Jez *et al.* 1997; Nebert *et al.*, 1989). This system has been successfully applied across a number of gene superfamilies (e.g. cytochrome P450 (CYP), glutathione-S-transferases (GST), glycosyltransferase (UGT) and sulfotransferases (ST)).

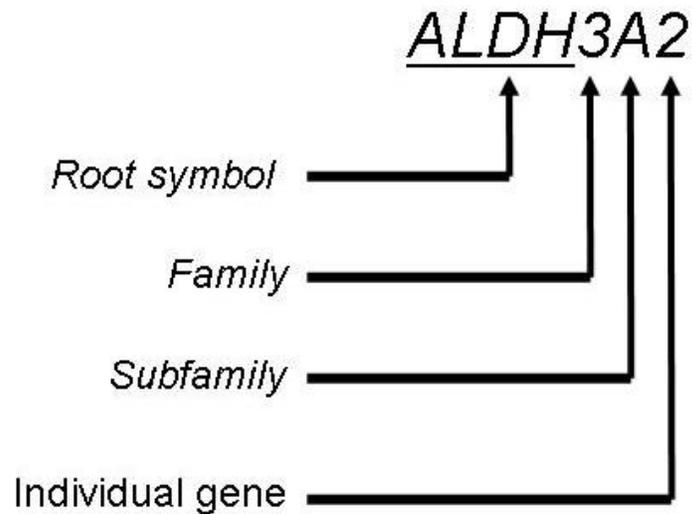
According to divergent evolution, each gene in a superfamily has originated from an ancestral gene, present usually more than two billion years ago, and exhibits more than 15-20% similarity to every other gene in that superfamily (Dayhoff, 1976). The following main differences among species that have diverged during the last several hundred million years are reflected within each superfamily; (a) gene duplication events, (b) the appearance of new gene functions resulting from genetic drift, unequal crossing-over and then (c) additional gene duplication events (Gupta, 1997; Todd *et al.*, 2001; Davis, 2002; Matsuda *et al.*, 2003). In future when all genes in more than 200 genomes will have been isolated and characterized there will be a need to name about 3 to 5 million genes. It is thus

proposed that in this decade of genomics the most rational and efficient method for gene nomenclature is to name all genes within each superfamily, across all species, on the basis of divergent evolution (Nebert, <http://www.uchsc.edu/sp/sp/alcdbase/aldhcov.html>).

With aldehyde dehydrogenases emerging as crucial enzymes in metabolic and detoxifying systems in a wide range of organisms more than 300 ALDH genes have been characterized to date, and the correlation of biochemical activities and gene sequences has been very confusing (Navarro-Avino *et al.*, 1999) hence the need for a uniform nomenclature system. Since 1998 a standardized ALDH nomenclature system has been established based on divergent evolution (Vasiliou *et al.*, 1995). Details of the system and names of members of the nomenclature committee are available on the ALDH website (<http://www.uchsc.edu/sp/sp/alcdbase/aldh-nomencl.html>).

In brief, the ALDH naming system used is based on the definitions of the following evolutionary terms;

- ❖ **Gene family:** An ALDH protein from one gene family is defined as having about  $\leq 40\%$  amino acid identity to that from another family.
- ❖ **Subfamily:** Two members of the same subfamily exhibit approximately  $\geq 60\%$  amino acid identity and are expected to be located at the same subchromosomal site.
- ❖ For naming each gene, root symbol "**ALDH**" denoting "aldehyde dehydrogenase" 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 as shown below:



- ❖ All letters are capitalized in all mammals except mouse and fruit fly, e.g. “human *ALDH1A1* (mouse, *Drosophila Aldh1a1*)”. The gene is italicized, whereas the corresponding cDNA, mRNA, protein or enzyme activity is written with uppercase letters and without italics, e.g.” human, mouse or *Drosophila* ALDH1A1 cDNA, mRNA, or activity” as shown below.

	<i>Mouse, Drosophila</i>	<i>All other species</i>
Gene	<b><i>Aldh1a1</i></b>	<i>ALDH1A1</i>
cDNA, mRNA,	ALDH1A1	ALDH1A1

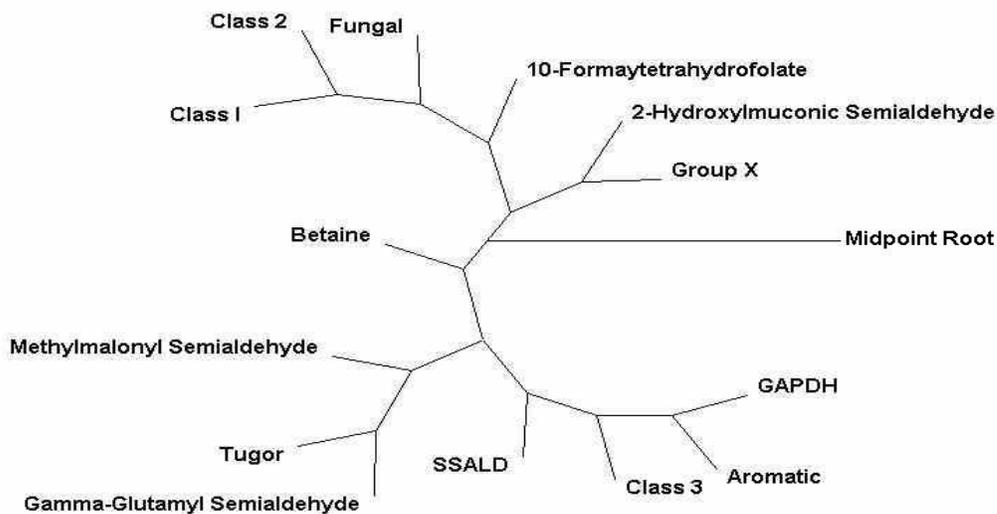
- ❖ It is also recommended that **human *ALDH* variant alleles** be given numbers (or number plus a capital letter) following an asterisk (e.g. “*ALDH3A2\*2*, *ALDH2\*4C*”).
- ❖ **Naming an *ALDH* gene and *ALDH* enzyme:** If an orthologous gene between species cannot be identified with certainty, it is recommended that one sends, at least the deduced amino acid sequence of the newly discovered *ALDH* gene or cDNA to the ***ALDH* Gene Nomenclature Committee**, in which case

sequential naming of the genes will be carried out in chronological order as they are reported to the committee. .

### **THE ALDH GENE SUPERFAMILY**

The genomic era, which began earnestly during the late 1980s, is now at a very advanced stage, with the sequences of more and more whole genomes becoming available. More than 50 complete genome sequences are now publicly available (NCBI's Entrez Genomic site) resulting in discovery of many unexpected additional genes in almost every superfamily. The next enormous task following these labour intensive sequencing projects is the systematic, large-scale study of the genomes in order to make sense out of these massive amounts of data and to be able to assign roles to uncharacterized open reading frames.

In phylogenetic terms a gene superfamily is “a cluster of evolutionarily related sequences” (Dayhoff, 1976), and consists of families, that are clusters of genes from different genomes that include both orthologs and paralogs (Tatusov *et al.*, 1997). Orthologs are genes in different species that evolved from a common ancestor by separation, whereas paralogs are gene products of gene duplication events within the same genome. In 1999, Perozich and co-workers carried out a sequence alignment of the then available 145 ALDH protein sequences to aid in determining ALDH structure and establishing relationships between various ALDH families. The protein sequence alignment of the 145 ALDHs showed that there are at least 13 ALDH families (Table 1.1.3), which can be split further into two main trunks of the phylogenetic tree (figure 1.1.3) (Hempel *et al.*, 1993). These trunks are the “Class 3” (class 3 ALDH down to betaine-ALDH (BALDH)) and “Class 1/2” (class 1 down to Group X). Each branch represents a point of divergence, where a gene duplicates and evolves to a new function. Distance between branches corresponds to evolutionary distance, which is measured by how much the sequences differ.



**Figure 1.1.3: The phylogenetic tree showing evolutionary relationships among ALDH subfamilies (Perozich *et al.*, 1999, also available at <http://www.psc.edu/biomed/pages/research/HBN/ALDH145/gtreer.jpeg>). SSD = succinic semialdehyde dehydrogenase, GAPDH = Glyceraldehyde-3-phosphate**

Divergence into any of the two trunks shows no correlation with subcellular localization, quaternary structure, or co-enzyme preference. However some families appear to be specific to a certain kingdom. For instance, class1 ALDHs have only been found in animals, Fungal ALDHs in fungi, and Aromatic ALDHs in bacteria (Perozich, 1999). A limited correlation in terms of substrate specificity has been observed among ALDHs belonging to one family. For example, the “Class 3” trunk consists of ALDH families of substrate specific enzymes (with the exception of one, namely class 3 ALDHs). On the other hand, “Class 1/2” trunk consists mostly of variable substrate enzymes, with the exception of a few.

**Table 1.1.3: A summary table of different ALDH families with information for each extended ALDH family (Adapted from Perozich *et al.*, 1999).**

Family	Examples	Pathway(s)	Substrate specificity	Coenzyme preference	Reaction
Class 3 ALDH	ALDH3A1 ALDH3A2	Metabolism of lipid peroxidation products, long chain "fatty" aldehydes and certain anti-cancer drugs	Variable, aromatics preferred	NAD or NADP	
Non-phosphorylating Glyceraldehyde-3-phosphate Dehydrogenase (GAPDH)	ALDH11A3	Glycolysis in the dark to generate NADPH for photosynthetic reactions	Highly specific	NADP	Glyceraldehyde-3-phosphate to 3-phosphoglycerate
Aromatic ALDH	FeaB (phenyl acetaldehyde dehydrogenase)	Catabolism of aromatic aldehydes by microbes	Each aromatic ALDH oxidizes a specific aromatic aldehyde	NAD	
Succinic Semialdehyde Dehydrogenase (SSALDH)	ALDH5A1	Metabolism of $\gamma$ -aminobutyraldehyde (GABA)	Specific	NAD in animals, NADP in bacteria	Succinic semialdehyde to succinate
Turgor ALDH	ALDH3I1 ALDH7B4 ALDH7B6 ALDH21A1	Response to dehydration and osmotic turgor		Not yet determined	Not yet determined
$\gamma$ -Glutamyl Semialdehyde Dehydrogenase (GGSALDH)	ALDH4A1 ALDH4B1	Proline metabolism	Specific	NAD	$\gamma$ -Glutamyl Semialdehyde to glutamate
Methylmalonyl Semialdehyde Dehydrogenase (MMSALDH)	ALDH6A1 ALDH6B1	Valine and pyrimidine metabolism	Specific	NAD	Malonyl- and Methylmalonyl semialdehyde to acetyl- and propionyl-CoA
Betaine ALDH (BALDH)	ALDH1H1 ALDH10A1	Resistance to dehydration and osmotic turgor	Specific	NAD	Betaine aldehyde to betaine
Class 1 ALDH	ALDH1A1	Metabolism of ethanol, retinaldehyde, 11-hydroxy-thromboxane B <sub>2</sub> and certain anti-cancer drugs, structural crystallins in shrews and cephalopods	Variable, prefer aliphatic aldehydes	NAD	
Class 2 ALDH	ALDH2B1	Metabolism of ethanol, pollen maturation	Variable, prefers aliphatic aldehydes	NAD	
Fungal ALDH	ALDH1D3 ALDH1E3 ALDH17		Variable	NAD	
10-Formyl-tetrahydrofolate Dehydrogenase (FTDH)	ALDH1L1	Folate metabolism	Specific	NADP	10-Formyltetrahydrofolate to tetrahydrofolate and CO <sub>2</sub>
2-Hydroxy-muconic Semialdehyde Dehydrogenase (HMSALDH)	ALDH12A1	Meta-fission pathway for catechols	Specific for substituted 2-hydroxy-muconic semi-aldehydes	NAD	2-Hydroxymuconic semi-aldehyde to 2-hydroxyhexa-2,4-diene-1,6-dioate
Group X			Specific for different aliphatic aldehydes	NAD	

The ALDH superfamily is very large as already said, and its members constitute a variety of isozymes that can generally be categorized into four groups that can be split further into 13 families (tables 1.1.3 and 1.1.4).

In 2002 the ALDH gene superfamily was updated to include 555 distinct cDNAs/genes (Sophos *et al.*, 2003) whose protein products contain the “ALDH signature sequence” (Hempel *et al.*, 1993), and are thus regarded as members of this superfamily. These gene sequences include 32 in archaea, 351 in eubacteria, and 172 in eukaryotes. A summary of different ALDH families with information for each extended ALDH family is shown in table 1.1.3.

**Table 1.1.4: Summary of the ALDH gene superfamily, as of June 2002** (Sophos and Vasiliou, 2003).

Superkingdom	Taxon	Number of genes	Total number of genes
Archaea	Crenarchaea	9	26
	Euryarchaeota	17	
Bacteria	Aquificales	2	331
	Cyanobacteria	5	
	Firmicutes	110	
	Proteobacteria	201	
	Spirochaetales	1	
	Thermotogales	1	
	Thermus/Deinococcus group	11	
Eukaryota	Diplomonadida	1	172
	Euglenozoa	2	
	Entamoebidae	2	
	Fungi	32	
	Metazoa	90	
	Viridiplantae	45	
Total number			529

The 2002 study of the individual genomes revealed that the number of ALDH genes found per organism ranged from 1-5 in archaeal species, 1-26 in eubacteria, and 8-17 genes in eukaryotic species (Sophos *et al.*, 2003). Out of a total of 172 genes identified in the eukaryotic ALDH gene superfamily, 5 were found in animals and 32 in fungi including yeast whereas the remaining belong to plants and other eukaryotes (Sophos *et al.*, 2003). A summary of the distribution of the 2002 ALDH gene superfamily is shown in table 1.1.4. The 2004 update of the aldehyde dehydrogenase superfamily is expected in 2005.

## **CONSERVED RESIDUES AND SEQUENCE MOTIFS IN ALDHS**

### **Conserved residues**

As is to be expected, catalytically important residues and segments in the ALDH structure are highly conserved so as to preserve tertiary structure, which in turn results in functional conservation. These conserved residues are essential for maintaining critical turns and loops in the tertiary structure of the ALDH protein, which in turn has direct bearing on functional definition of the protein.

In 1993 Hempel and co-workers described identification of 23 invariant ALDH residues while working on a group of 16 ALDH sequences representing the diversity of the ALDH family diversity at that time. These residues were found to be at least 95% conserved in all ALDHs studied then. Thereafter 25 conserved ALDH residues were found to play a direct role in catalysis as stated by Hurley and Weiner (1999). However, in 1999 Perozich and co-workers described alignment of 145 ALDH sequences and in addition performed mutational analyses to confirm their observations. Of the 16 conserved residues identified in the previous alignment by Hempel and co-workers only the following four (also among Hurley and Weiner's list) have been identified as invariably conserved among all known ALDHs to date (rat cytosolic class 3 ALDH numbering);

Gly<sup>187</sup> and Phe<sup>335</sup> – these two residues form an integral part of the coenzyme binding Rossmann fold by interacting with the nicotinamide portion of NAD(P).

Gly<sup>240</sup> – maintains tertiary ALDH structure by allowing the main chain to twist back on itself so as to be able to position the catalytic nucleophile for catalysis.

Glu<sup>333</sup> – Due to its close proximity to the catalytic thiol, the residue may serve to activate the thiol through a water molecule. Involvement of this residue in cofactor binding is also indicated.

In addition to above four invariant residues, another 12 residues were found conserved in more than 95% of the 145 ALDH sequences examined. These are Arg<sup>25</sup>, Gly<sup>105</sup>, Asn<sup>114</sup>, Pro<sup>116</sup>, Gly<sup>131</sup>, Lys<sup>137</sup>, Gly<sup>211</sup>, Cys<sup>243</sup>, Pro<sup>337</sup>, Gly<sup>383</sup>, Asn<sup>388</sup>, and Gly<sup>403</sup>. These 12 residues plus 9 others, which are excluded by mutational analysis studies, add up to Hurley and Weiner's list of 25 conserved residues.

Although no specific roles can be assigned to each of the 12 conserved residues most of them are found to lie at critical turns and loops in the ALDH structure (Perozich *et al.*, 1999). For example, Gly<sup>211</sup> is the first residue in the Gly-Gly dipeptide part of the boundary between coenzyme binding and catalytic domains, whereas Gly<sup>403</sup> is part of the "U-turn" region. Glu<sup>209</sup>, though less than 95% conserved, this residue may possibly be acting as a general base for the catalytic mechanism, serving to deprotonate the catalytic thiol or aiding in expulsion of the free acid product. Cys<sup>243</sup> is the catalytic thiol, and together with Arg<sup>25</sup> are present in all ALDHs with catalytic activity. ALDHs lacking enzymatic activity such as the  $\Omega$ -crystallins have other residues in these positions. Two mold ALDHs (*Alternaria alternata*-aldh and *Cladosporium hebarum*-aldh) identified simply as allergens without any ALDH activity also lack the Asn<sup>114</sup> residue and instead have Glu at this position (Perozich *et al.*, 1999).

### **Conserved sequence motifs**

Sequence comparisons among ALDH genes from bacteria, plants and animals demonstrating ALDH enzymatic activity have shown at a glance, three diagnostic amino acid motifs; (i) the ALDH glutamic acid active site signature sequence MELGGNA (LELGGKS for mammalian class 3 ALDHs), (ii) the Rossmann fold

GxGxxG (or GxTxxG) coenzyme binding site and (iii) the catalytic thiol (Kirch *et al.*, 2004). In addition, seven other amino acid sequence motifs are observed. Generally there are 10 conserved sequence motifs among the ALDH extended family (Perozich *et al.*, 1999). The 10 most conserved sequence motifs among ALDHs are described in Table 1.1.5. These motifs are stretches of sequences ranging from five up to 14 or 15 amino acids. They are spread along the entire ALDH sequence, but fold back together and come into contact with each other in the 3- D structure.

Overall the 10 motifs reside at or near the active site of the ALDH molecule, and appear to effect essential ALDH structure/function elements. Most of these motifs contain a conserved turn or loop (Perozich *et al.*, 1999) with a highly conserved and hydrophobic small amino acid such as glycine, proline, aspartic acid or asparagines, which does not take part in enzyme function. Table 1.1.5 provides a description of all the 10 most conserved sequence motifs among ALDHs. The three dimensional structure of the 10 most conserved ALDH sequence motifs are shown in figure 1.1.4.

**Table 1.1.5: The ten most conserved sequence motifs in ALDHs (Perozich *et al.*, 1999)**

Motif number <sup>a</sup>	Length	Motif <sup>c</sup>
1	5	[Past]-[WFy]-[ <b>Ne</b> ]-[FYgalv]-[PtI]
2	14	[Apnci]-[Liamv]-[Avslcimg]-[ACtlmgf]- <b>G</b> -[Ncdi]-[Tavcspg]- [Vaimfcltgy]-[Vil]-[Lvmiwafhcy]-[ <b>Kh</b> ]-[Ptvghms]-[ASdhp]- [Epsadqgilt]
3	10	[Grkpwhsay]-[FLeivqnarmhk]-[Pg]-[Plakdievsrf]-[Gnde]-[Vliat]- [VLifyac]-[Nglqshat]-[Vllyaqqfst]-[IVlms]
4	10	[IVlgy]-[SATmnlfhq]-[Fyla]-[Tvil]- <b>G</b> -[Sgen]-[Tsvrindepaqk]- [EAprqgktnldh]-[VTiasgm]-[Gafi]
5	16	[Lamfgs]-[Enlqf]-[Ltmcagi]-[ <b>Gs</b> ]-[Ga]-[Knlmqshiv]-[SNade]- [Pahftswv]-[cnlmgivahst]-[lvlyfa]-[Viamt]-[Fdlmhcanyv]- [Daeskpmt]-[Dsntaev]-[Acvistey]-[Dnlera]
6	8	[FyvIma]-[FgyIrmdaqetwsvikp]-[Nhstyfaci]-[QAshtcmg]- <b>G</b> - [Qe]-[crvitksand]-[Cr]
7	9	[Gdtskae]-[Yfnarthelswv]-[FYlwvis]-[IVlfym]- [Qeapkgmynhlswyv]-[Pa]-[Tachlmy]-[VII]-[FLivwn]
8	7	[Ektdrqs]- <b>E</b> -[lvtnfsp]- <b>F</b> -[Ga]-[ <b>Ps</b> ]-[Vilcf]
9	15	[Nrst]-[Dnaseqtkregi]-[TSrvnacqgik]-[EpdtgqikvrfshyncI]- [Yfkqvm]-[Gpa]-[Lnmv]-[Astgvqcf]-[Agsltfe]-[AGysct]-[Vllfams]- [Fhwyivlem]-[TSag]-[KRnsqteahdp]-[DNsileakt]
10	12	[Pasw]-[Fwyahv]-[Gtqs]- <b>G</b> -[Fvyesnimtawrq]-[Kgm]- [mqarelnskghdpt]-[Stm]-[Gfls]-[lfntlmygshrvq]-[Gdnhrsy]- [Rdpsagkte]

<sup>a</sup>Motifs are numbered consecutively in order of appearance in the ALDH sequences.

<sup>b</sup>Motifs are given as ProSite patterns. Capitalized letters represent residues that are predominant at each bracketed position, and residues highlighted in bold are conserved in at least 95% of known ALDHs.

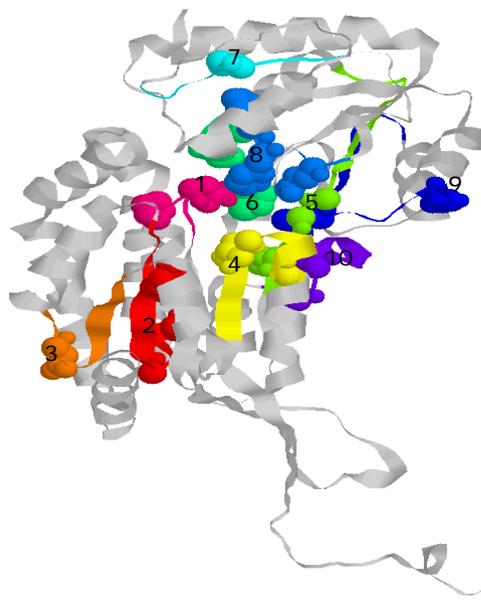


Figure 1.1.4: A ribbon diagram of the 3-D structure of rat ALDH 3 showing the 10 conserved motifs in colour (Perozich *et al.*, 1999; available at <http://www.uchsc.edu/sp/sp/alcdbase/aldh-nomencl.html>). The numbers 1-10 in bold, indicate position of each of the ten motifs.

Though no specific roles can be assigned to each of these motifs the following characteristics about some of them are observed:

**Motif 1** is the most conserved motif in ALDHs and it bears at its centre Asn 114, the residue nearest to the catalytic thiol.

**Motif 4** covers the essential NAD-binding turn of the Rossmann fold, between  $\beta$ -4 and  $\alpha$ -D in the class 3 ALDH structure. This motif is resident to the conserved Gly 187 found as the first glycine in the NAD-binding turn in all ALDHs as well as in the Rossmann fold of several other dehydrogenase families.

In **Motif 5** reside both Glu 209, proposed to act as a general base as well as Gly-Gly 400-401, the dipeptide forming the boundary between coenzyme and catalytic domains.

**Motif 6** bears the invariant Gly 240 and the catalytic thiol, Cys 243.

**Motif 8** is the only motif bearing several of the invariant residues, namely Glu 338 and Phe 335.

Encoded in **Motif 10** is the intriguing “U-turn” spanning  $\beta$ -12 and  $\alpha$ -14

## FATTY ALDEHYDE DEHYDROGENASES

Fatty aldehyde dehydrogenase (FALDH) is a microsomal ALDH enzyme that catalyses the oxidation of a wide range of aliphatic aldehydes ranging from 2 to 24 carbons in length, but it prefers medium to long chain aldehydes (fatty aldehydes), including saturated and unsaturated aldehydes. Similar to the majority of ALDH enzymes this enzyme also prefers NAD<sup>+</sup> to NADP<sup>+</sup> as nucleotide cofactor (Kelson *et al.*, 1997; Rizzo *et al.*, 2001). However, bacterial FALDHs from *Vibrio harveyi* and *Acinetobacter spp.* were found to have a higher affinity for NADP<sup>+</sup> than NAD<sup>+</sup> (Singer and Finnerty, 1985; Zhang *et al.*, 2001).

The mammalian FALDH enzyme (class 3 ALDH, ALDH3A2) has been purified from rats, rabbits and humans and FALDH activity has also been detected in several species of alkane-metabolizing yeasts and bacteria such as *Candida*, *Pseudomonas* and *Acinetobacter* (Ueda and Tanaka, 1990; Singer and Finnerty, 1985; Fox *et al.*, 1992; Zhang *et al.*, 2001). The mammalian FALDH protein has a subunit molecular weight of about 54kDa, similar to most ALDHs (Kelson *et al.*, 1997; Rizzo *et al.*, 2001). The enzyme is synthesized on free polysomes and then inserted post-translationally into the endoplasmic reticulum. This protein is not known to undergo any post-translational modification. The mammalian enzyme serves in detoxification of aldehydes resulting from metabolism of such compounds as fatty alcohol, phytanic acid, ether glycerolipids and leukotriene-B<sub>4</sub>. In bacteria and yeast the enzyme is found in organisms growing on alkanes and related compounds, where it participates in the carbon flow from n-alkanes to cell constituent synthesis and energy production through  $\beta$ -oxidation.

The FALDH cDNAs from rat, mouse and human have been cloned, and the organization of the human and mouse FALDH genes have been described (Miyachi *et al.*, 1991; Vasiliou *et al.*, 1996; Rogers *et al.*, 1997; Chang and Yoshida, 1997; Lin *et al.*, 2000). All three FALDHs (rat, mouse and human) are microsomal and highly homologous to each other. The amino acid identity

between human and rat FALDH protein is 84% and 95% between the two rodent proteins.

A distinguishing feature of FALDH as opposed to other ALDHs is the presence of a distinct hydrophobic domain at the carboxy-terminal, which is made up of 35 amino acids that help to anchor the protein to the microsomal membrane (Masaki *et al.*, 1994). Both human and mouse FALDHs consist of 11 exons and 10 introns, and are subject to alternative splicing (Chang and Yoshida, 1997; Lin *et al.*, 2000) (see figure 1.1.5).

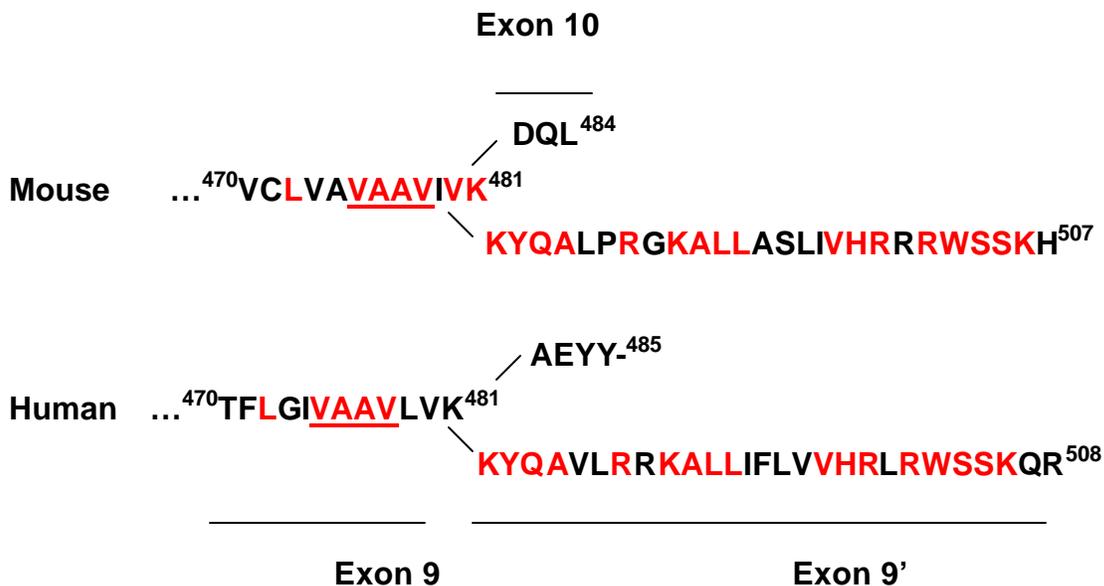


Figure 1.1.5: Carboxy-terminal sequences of mouse and human major FALDH protein (exons 9-10) and minor FALDH? protein (exons 9'-10). Red coloured residues are identical in human and mouse (Chang and Yoshida, 1997; Lin *et al.*, 2000).

In both species, alternative splicing inserts an additional exon (exon9') between exons 9 and 10, replacing the carboxy-terminal amino acids with others. This results in production of a second minor protein, FALDH?, with a variant carboxy-terminal, whose function is unknown. The major protein species, FALDH, (484 and 485 amino acids in mouse and human respectively) consists of exons 1-10, whereas FALDH? comprising about less than 10% of the total transcripts has exon 9' spliced between exons 9 and 10. Exon 10 is not translated in FALDH?

due to the presence of a stop codon in exon 9'. The mouse FALDH is slightly smaller (25kb) than human (31kb) due to decreased intron size, however the intron-exon boundaries still remain identical in both species (Lin *et al.*, 2000).

On the other hand, the mouse FALDH<sup>?</sup> has a carboxy-terminal amino acid, SKH, which resembles the consensus peroxisomal targeting signal-1 sequence, SKL, but studies have not yet been done to determine whether mouse FALDH<sup>?</sup> is targeted to peroxisomes. No peroxisomal FALDH enzyme activity has been detected in humans. Due to its function plus the fact that the enzyme is found in a diverse variety of all tissues so far tested, FALDH is considered a housekeeping gene (Chang and Yoshida, 1997).

Based on its role in lipid metabolism and microsomal localization, FALDH is also classified as a class 3 ALDH (Kelson *et al.*, 1997). Both ALDH3 and FALDH (ALDH3A1 and ALDH3A2 respectively) are class 3 enzymes whose major substrates are fatty and aromatic aldehydes. In addition, both enzymes exhibit genes that are virtually identical in terms of exon/intron structure. However, ALDH3 has a shorter exon 9, which does not produce the COOH-terminal transmembrane domain encoding a microsomal enzyme, and instead produces a cytosolic enzyme. The FALDH gene is located about 50-85kb away from ALDH3. This can only imply that by such high sequence similarity, conservation of structure, and physical proximity the two genes are likely to have arisen by duplication from a common ancestral gene (Rogers *et al.*, 1997). Though exhibiting such extensive similarity, the few differences observed between the two enzymes are so crucial that ALDH3 cannot supplement or substitute the role of FALDH in the synthesis of membrane lipid.

The importance of human FALDH in alcohol metabolism is emphasized by its genetic deficiency in Sjögren-Larsson syndrome (SLS), a disease characterized by presence of congenital ichthyosis, mental retardation, spasticity and di- or tetraplegia (Rizzo *et al.*, 2001). In affected patients FALDH activity is diminished

resulting in an impaired ability to oxidize a broad range of alcohols and aldehydes. As a result there is accumulation of fatty alcohols in plasma, which is a cause of the described symptoms. Shorter chain alcohols do not accumulate probably due to presence of other ALDHs. Free fatty aldehydes have never been observed in SLS patients, probably due to their being highly reactive and forming stable covalent derivatives thus manage to elude detection (Kelson *et al.*, 1997). The FALDH seen to be deficient in SLS patients appears to act on long chain aliphatic aldehydes and is particulate in nature, but its precise subcellular localization is unknown and it has never been purified.

Molecular studies of mutations in SLS have aided in understanding structure-function correlations in FALDH and other ALDH proteins. More than 60 mutations have been identified in SLS patients, and these include deletions, insertions, missense mutations, and splicing errors (Rizzo *et al.*, 2001). These mutations are scattered throughout the entire gene, except for the hydrophobic carboxy-terminal tail. Most amino acid substitutions are associated with highly conserved residues in the FALDH protein causing complete loss of enzyme activity whereas several non-critical substitutions result in a small amount of catalytic activity ranging from 1% to 9% of normal activity. Polymorphisms in the human FALDH gene are due to five sequence variations that do not cause disease. All polymorphisms are located in introns, and none are so far known to be associated with disease (Rizzo *et al.*, 2001).

## **FUNGAL ALDEHYDE DEHYDROGENASE GENES**

The fungal family of organisms is composed largely of the filamentous fungi, commonly known as molds and the single celled fungi called yeasts. Fungal ALDHs appear to separate into two subfamilies; mold ALDHs and yeast ALDHs. In the 2002 update of the ALDH superfamily a total of 41 ALDH/ALDH-like gene sequences were identified in fungi (Sophos and Vasiliou, 2003). With the genomes of several species of filamentous fungi completely sequenced, the majority of ALDHs in this group were mold ALDHs, whereas a few yeast ALDHs

were known from only two yeast genomes (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* which were completely sequenced by the time the 2002 compilation of ALDH sequences by Sophos and Vasiliou was made). Table 1.1.6 shows the different fungal species and the number of ALDH genes/ALDH-like gene sequences listed in the 2002 update of the ALDH superfamily (Sophos and Vasiliou, 2003). Sequences from recently completed genomes were not incorporated into this table.

**Table 1.1.6: A summary of fungal ALDH genes, as of June 2002** (Sophos and Vasiliou, 2003).

ALDH genes	Species	Number of ALDH genes	Total number of genes
Molds	<i>Agaricus bisporus</i>	2	11
	<i>Alternaria alternata</i>	1	
	<i>Aspergillus niger</i>	1	
	<i>Cladosporium spp.</i>	2	
	<i>Emericella nidulans</i>	3	
	<i>Pichia angusta</i>	1	
	<i>Ustilago maydis</i>	1	
Yeasts	<i>Saccharomyces cerevisiae</i>	13	21
	<i>Schizosaccharomyces pombe</i>	8	
Fungal ALDHs			32

The study of 145 aligned ALDH sequences by Perozich and co-workers (1999) showed a few characteristics unique but not exclusive to fungal ALDHs. The fungal ALDH family comprised both cytoplasmic and mitochondrial ALDHs of variable substrate specificity that were closely related to class 1 and 2 ALDHs. Phylogenetic analysis pointed to the fact that the ancestral “class1/2” gene may have diverged along two separate lines, one leading to fungal ALDHs and the other to the higher plant and animal forms, represented by the class 1 and 2 isozymes. In addition, type I  $\gamma$ -glutamyl semialdehyde dehydrogenase (GGSALDH) gene sequences from yeast, mushroom and human are found to

have a Phe residue index 373 lying at the beginning of the helix  $\alpha$ -D, immediately following critical coenzyme binding turn. The Phe residue is otherwise almost always Gly or Ala in other ALDHs. Substitution of Phe<sup>373</sup> for Gly in the lower bacterial forms (e.g. *Bacillus subtilis*) may suggest that substitution of this residue to Phe, which is a bulkier amino acid, is a characteristic of GGSALDH of higher organisms (Perozich, *et al.*, 1999).

The family 2 cytosolic and mitochondrial ALDHs has been characterized extensively in humans and yeast (Hsu *et al.*, 1995; Yoshida *et al.*, 1996; Navarro-Avino *et al.*, 1999; Vasiliou *et al.*, 2000). These tetrameric enzymes are involved in acetaldehyde and other short chain ALDH metabolism. There is no systematic study available yet about other families of ALDHs in fungi.

Though it is a well known fact that fungi, especially the majority of non-conventional yeasts grow very well on long chain alcohols and aldehydes very little information is available about genes coding for enzymes responsible for metabolism of these substrates by fungi. As observed from table 1.1.6 the majority of yeast ALDH genes studied were mainly from *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* whose genomes had been completely sequenced (before 2000). A few other yeast ALDHs were from unfinished genomes and had not yet been fully characterized. The number of fungal ALDHs will increase significantly in the 2004 update of the aldehyde dehydrogenase superfamily, which is expected in 2005 (also see chapter 4).

## 1.2 ALCOHOL DEHYDROGENASES

### Introduction

Alcohol dehydrogenase enzymes (ADH; EC 1.1.1.1) occur widely in a variety of living organisms, where they are responsible for detoxification and metabolism of ethanol and other alcohols. In addition, ADH reacts with a wide variety of aldehydes and ketones (Branden *et al.*, 1975) with the concomitant reduction of NAD(P)<sup>+</sup>.

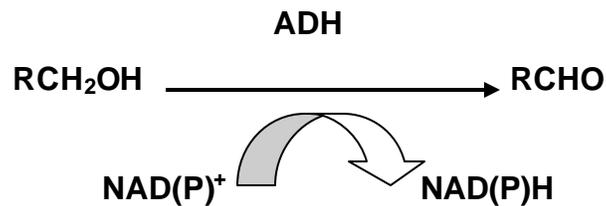


Figure 1.2.1: The overall reaction catalyzed by alcohol dehydrogenases.

Studies of many alcohol dehydrogenases and related enzymes aimed at defining functional properties, structural patterns, and evolutionary relationships have resulted in the following four major conclusions about these enzymes (Jörnvall *et al.*, 1997);

- (i) There are clearly multiple enzymes representing different protein families. The class origins of different ADH enzymes have been traced to repeated gene duplications at minimally four different evolutionary levels. This is clearly seen within the medium chain dehydrogenases/reductases (MDR) family where these repeated duplications have produced the different enzymes within each family, as well as different classes of the enzymes within each family and different isozymes within the classes, excluding the allelic variants.
- (ii) The medium chain Class III ADH, with its glutathione-dependent formaldehyde dehydrogenase activity appears to be of ancient origin, which has evolved to give rise to much of the ADH family, encompassed by one basic function, which is cellular detoxification (Jörnvall, 1999).

- (iii) Separate, molecular architectures are noticeable within the different ADH classes (Danielsson *et al.*, 1994). Two types of ADH in separate protein families are recognizable, namely the “medium chain” zinc enzymes and the “short chain” enzymes distinguishable by ethanol dehydrogenase activity. Class III ADH, which is a member of the MDR family appears to be constant in terms of evolutionary patterns and variability, thus it resembles the classical protein/enzyme pattern. The short chain enzymes on the other hand are highly variable in overall structure and functional properties (Krozowski, 1994). This protein-atypical manner has resulted in this group exhibiting different enzyme activities (including alcohol dehydrogenase) and they operate in a completely different catalytic mechanism. The medium-chain ADH family was initially called the long chain family (Jörnvall *et al.*, 1987), before other, still longer-chain families were recognized. These are the true long chain ADHs as well as the iron-activated ADHs.
- (iv) It appears that functional convergence towards ethanol activity has occurred in many lines. Consequently, ethanol-active enzymes are present in yeast, prokaryotes, plants, and animals even though they all appear to have separate lines (Jörnvall, 1994).

### **The different alcohol dehydrogenase classes**

The ADH family consists of enzymes catalyzing metabolism of a wide variety of substrates such as ethanol, retinol, other aliphatic alcohols, hydroxysteroids and lipid peroxidation products (Duester *et al.*, 1999). Such a wide variety of substrates suggest that the specific enzymes involved differ in terms of structure, genetics and enzymology. It is for this particular reason that Chase (1999) suggested that the name “alcohol dehydrogenase” is a misnomer. A special reference was made in this regard to the insect-type or short chain dehydrogenases, which with the exception of the *Drosophila* enzyme, do not oxidize ethanol or any aliphatic alcohols, but rather secondary alcohols and

steroids. Despite such a difference in substrate specificity all enzymes referred to as ADHs have one feature in common, i.e. in presence of NAD(P)<sup>+</sup> all catalyze the reversible transfer of hydrogen to the carbonyl group of aldehydes or related compounds (Ciriacy, 1997), and moreover ADH is stereospecific for which proton it removes, i.e. it removes the pro-R hydrogen. (<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehyznrt.html>).

**Table 1.2.1: The four classes of alcohol dehydrogenases (Chase, 1999)**

Class	Characteristics		Examples, EC numbers	References
Short chain			~250-300 residues: <i>Drosophila</i> Adh (EC 1.1.1.1), steroid reductases (EC 1.1.1.51)	Jörnvall <i>et al.</i> , 1995
Medium chain	Zinc-containing	dimeric forms	~375 residues contain a second, structural (non-catalytic) zinc, e.g. horse liver ADH (EC 1.1.1.1), Cinnamyl ADH (EC 1.1.1.195)	Jörnvall <i>et al.</i> , 1987 Persson <i>et al.</i> , 1994
		tetrameric forms	<i>S. cerevisiae</i> ADH (EC 1.1.1.1), sheep sorbitol dehydrogenase (EC1.1.1.14)	Jörnvall <i>et al.</i> , 1984 Magonet <i>et al.</i> , 1992
	Non-zinc-containing		quinone reductase/ $\beta$ -crystallins (EC 1.6.5.5), enoyl reductase (EC 1.3.1.10)	Borrás <i>et al.</i> , 1989
Long chain			600-750 residues, bacterial ethanol dehydrogenases, e.g. <i>Acetobacter aceti</i> ; pyrroloquinoline quinine as cofactor (EC 1.1.99.8)	Mackintosh and Fewson, 1987 Inoue <i>et al.</i> , 1989
Iron-activated			1,2-propanediol dehydrogenase of <i>E. coli</i> (EC 1.1.1.77), Adh2 of <i>Zymomonas mobilis</i> (EC 1.1.1.1); AdhIV of <i>S. cerevisiae</i>	Sridhara <i>et al.</i> , 1969 Scopes, 1983 Conway <i>et al.</i> , 1987 Conway and Ingram, 1989 Williamson and Paquin, 1987

Consequently, there are three well established groups of ADHs in prokaryotes and eukaryotes, namely, the medium chain zinc-dependent ADH, the short chain non-zinc and the polyol/sugar dehydrogenases and recently a fourth group of

iron-activated ADHs have come to light. Other ADHs, which do not fall into any of the previously mentioned groups, have also been reported. Even though the catalytic mechanisms of these different classes of ADH enzymes may be different, the overall reactions remain the same.

**(i) Zinc-dependent medium chain alcohol dehydrogenases**

The zinc-containing ADH line consists of dimeric and tetrameric proteins of about 350 residues, with a subunit made up of two domains; one catalytic and the other for cofactor binding. The group includes the classical ADHs, such as the liver and yeast enzymes as well as plant enzymes and several bacterial enzymes (<http://www.users.csbsju.edu/~hjakubow/classes/rasmolchime/99ch331proj/alcoholdehydro/i...>). The vertebrate ADH, which belongs to this group, consists of several enzymes of seven distinct classes, based on sequence alignment, phylogenetic tree analysis, catalytic properties, and gene expression studies (Duester *et al.*, 1999). Some organisms have multiple ADH isoenzymes within a single class, however no species so far is known that encodes all seven ADH classes. For example, in human are found at least nine ADH genes, seven of which encode ADHs of five different classes (class I, II, III, IV and V) and in mouse only four different ADH classes (class I, II, III, and IV) have been reported (Persson *et al.*, 1993; Duester *et al.*, 1999). Class IV ADH has been observed only in rat and deer mouse, whereas class VII has only been found in chicken (Duester *et al.*, 1999).

In common medium-chain ADH, subunits are characterized by presence of two zinc atoms, with one being catalytically active at the active centre, while the other one is not (<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehyznrt.html>).

Many enzymes of this group have a sequence motif known as the ADH signature: GHEX<sub>2</sub>GX<sub>5</sub>(G,A)X<sub>2</sub>(I,V,A,C,S) (Persson *et al.*, 1993). These enzymes prefer primary alcohols as substrates and exhibit hydride transfer with 4-*pro*-R stereospecificity in their reactions. Presence of catalytic zinc atom is however not

an absolute characteristic of the family as some members (e.g.  $\zeta$ -crystallin) lack the zinc atom even though they exhibit a typical overall protein homology of the group (Borras *et al.*, 1989; Jörnvall *et al.*, 1993). Even in presence of the catalytic zinc atom ligands to the zinc are quite variable amongst members of the group. Consequently the MDR family itself exhibits a complex of variations in metalloenzyme properties and active site structure.

Medium chain ADHs have been found to be related to polyol dehydrogenases, and both enzymes have approximately 350 amino acids per subunit. Examples of medium chain ADHs include the mammalian liver and sorbitol dehydrogenases, as well as ADHs from birds, plants, yeast and bacteria (Ribas de Pouplana, 1991). The classical ADH is part of a widely spread system of zinc-containing enzymes and at least six different classes of this enzyme are known to occur in mammalian tissues (Koivula and Uotila, 1990). Of the six classes only classes I, III and IV are well characterized whereas classes II, V and VI are not so well studied and have only been known to exhibit a few distinct properties. A considerable difference representing different stages between enzymes and ordinary isozymes is clearly visible amongst the different classes. For example, class I is the well-known liver enzyme with ethanol dehydrogenase activity, class III is the one proven to be identical with glutathione-dependent formaldehyde dehydrogenase and class IV is the form preferentially expressed in the mammalian stomach. The class origins of medium chain ADH have been traced to gene duplications early during evolution resulting in emerging activities towards ethanol (Jörnvall *et al.*, 1993). Class I enzyme properties show variable evolutionary patterns whereas class III forms have rather remained the most evolutionarily conserved and correspond to an ancestral form.

## **(ii) Short-chain alcohol dehydrogenases**

The short-chain dehydrogenases/reductases (SDR) family similarly has evolved into a large family comprising many different NAD(P)<sup>+</sup>-dependent oxidoreductase enzyme activities, though it exhibits a catalytic mechanism completely different

from the zinc-containing ADHs.  
(<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehyznrt.html>).

Short chain ADHs are characterized by their preference for secondary alcohols as substrates, do not require any metal cofactors and show 4-*pro*-S stereospecificity (Benner *et al.*, 1985). Short chain ADHs are found in mammalian tissues, insects and bacteria. Examples include *Drosophila* ADH, bacterial ribitol, glutamate and 20 $\beta$ -hydroxysteroid dehydrogenases and mammalian 17-hydroxysteroid and 15-hydroxyprostaglandin dehydrogenases.

Most members of the short chain ADH family are proteins of about 250 to 300 amino acids. Several crystal structures of members of this group have been solved (Benach *et al.*, 1998). The insect enzyme is a homodimer whereas most other members are tetramers. Despite low amino acid sequence identity (< 30%) in the group all members show a striking similarity in their overall folding and reaction mechanism (Qin *et al.*, 2000). A comparison of the short chain superfamily has revealed that this group consists of peroxisomal proteins, which are highly homologous to each other (Qin *et al.*, 2000). These enzymes are involved in fatty acid and cholesterol side-chain- $\beta$ -oxidation. Another common feature of these enzymes is that they all utilize CoA derivatives as substrates, a property which could well give an insight to the nature of the substrate-binding site of this group of enzymes. Studies of the short chain *Drosophila melanogaster* enzyme resulted in secondary structure predictions of ADH, suggesting that the N-terminal half of the enzyme folds into alternating  $\alpha$ -helix/ $\beta$ -strand structure characteristic of nucleotide binding domains, while the C-terminal half does not appear to have this alternating  $\alpha/\beta$  structure (Ribas de Pouplana *et al.*, 1991).

### **(iii) Iron-activated alcohol dehydrogenases**

This is the least characterized group of ADHs with only four members known to date, namely three from bacteria and one from yeast (Scopes, 1983; Conway *et al.*, 1987; Williamson and Paquin, 1987). All of these members are NAD(P)-dependent. The ADH isolated from *Zymomonas mobilis* had a high specificity for

ethanol and required ferrous ions for activity (Neale *et al.*, 1986). Inactivation of this enzyme by complexing agents could be reversed fully by addition of ferrous ions and partially with cobaltous ions. Other divalent metal ions, including zinc, however could not restore enzyme activity. Iron-activated ADHs show no obvious sequence similarity to other dehydrogenases except amongst themselves (Glasfeld and Benner, 1989). These enzymes are not so obviously homologous to the zinc-dependent ADHs either, yet they contain a metal ion that appears to play the same mechanistic role as the zinc in zinc-dependent dehydrogenases. Since it is generally accepted that function determines stereospecificity in metal-dependent dehydrogenases, it is therefore logical to assume that iron-dependent dehydrogenases exhibit the same stereospecificity as zinc-dependent enzymes. Indeed the ferrous-dependent ADH from *Zymomonas mobilis* was found to catalyze the transfer of the pro-R hydrogen of NADH just like the zinc-dependent enzymes (Glasfeld and Benner, 1989). No structural information is yet available about these proteins.

#### **(iv) Long chain alcohol dehydrogenases**

A fourth group of ADHs is the NAD(P)<sup>+</sup>-independent ADHs. Members of this group are those enzymes that use neither NAD<sup>+</sup> nor NADP<sup>+</sup> as cofactors and instead are linked to pyrrolo-quinoline quinones, flavins or haem centres (Mackintosh and Fewson, 1987). These proteins are about 600 to 750 residues long. Examples include several bacterial ADHs such as the one from *Paracoccus denitrificans*, *Methylobacterium organophilum* and *Acetobacter aceti* (Wales and Fewson, 1990). Several other enzymes from plants, bacteria, yeast and mammals, though structurally different are known to be evolutionarily related to these ADHs (Lindahl *et al.*, 1992). Structural information about these proteins is still lacking.

#### **The structure of alcohol dehydrogenase**

The first three-dimensional (3-D) structure of an ADH was elucidated by extensive crystallographic and sequence studies of the horse liver enzyme

(Eklund *et al*, 1976). The horse liver ADH (LADH) is a dimeric member of the zinc-containing ADHs with a molecular mass of 40kDa. Since elucidation of its 3-D structure it has thus been reasonably assumed that the other members of the homologous medium and long chain ADH family have the same topologies (Niederhut, *et al.*, 2001). A ribbon diagram of alcohol dehydrogenase indicating the positions of the side chains and the monomer as a whole is shown in figure 1.2.2.

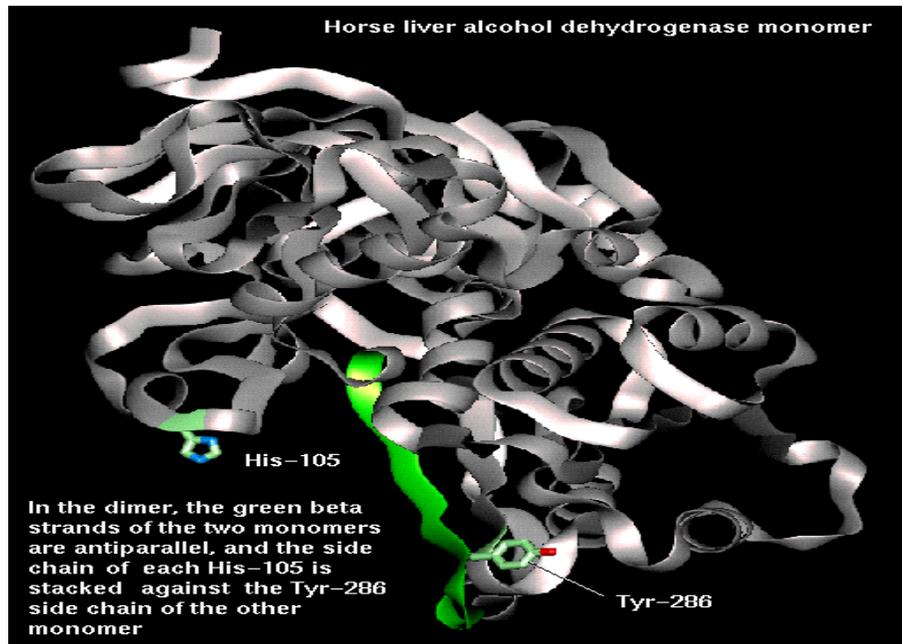
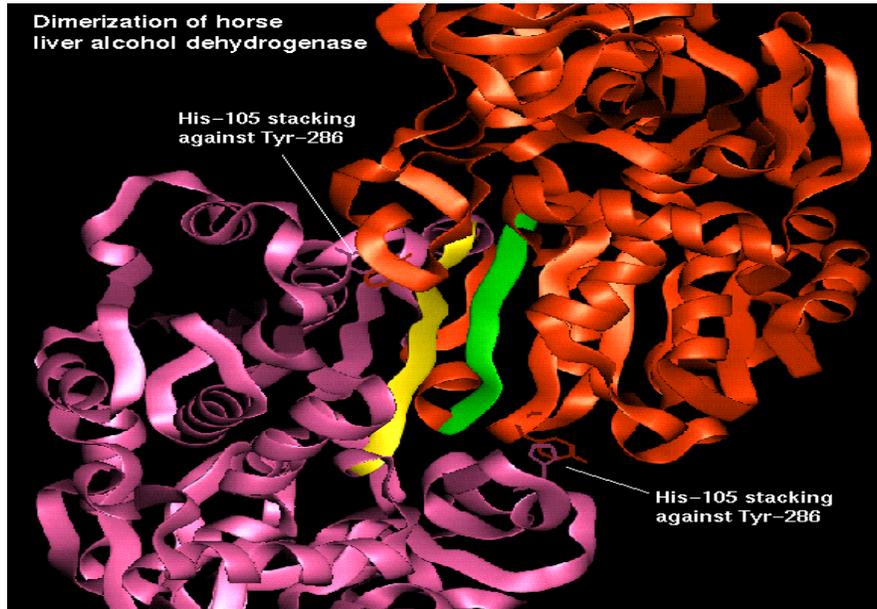


Figure 1.2.2: A ribbon diagram of horse liver alcohol dehydrogenase indicating the positions of the side chains in relation to the interfacial beta strands and the monomer as a whole.

The zinc-containing ADHs are known to bind two zinc atoms per subunit and one of the zinc atoms is essential for catalysis while the other one is not (<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehynrt.html>). The zinc atoms are coordinated by either cysteine or histidine residues, with the catalytic zinc atom coordinating with sulphurs of two cysteine and one histidine residues (Cys<sup>46</sup>, Cys<sup>174</sup> and His<sup>67</sup>) (<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehynrt.html>). An ionizable water molecule hydrogen bonds with the hydroxyl group of Thr<sup>48</sup> and occupies the fourth position on the catalytic zinc atom. The fifth and final position

on the zinc atom is occupied by an oxygen from the substrate, an alcohol. Thus an integral aspect of ADH catalysis is achieved through electrostatic stabilization of the alcohol's oxygen by the catalytic zinc atom, which consequently makes the proton on the alcohol more acidic.



**Figure 1.2.3: A ribbon diagram of horse liver ADH dimer illustrating the nature of interaction surfaces of the monomers prior to dimerization.**

Each subunit of the horse liver ADH folds into two domains, with the active site lying at the interface between the two domains (Eklund, 1989). It is at the interdomain region where  $\text{NAD}^+$  binds and makes contact with the two domains (figure 1.2.3). It is generally assumed that the size and shape of the substrate-binding pocket determine substrate specificity of an enzyme. This proposal was tested and proven by Green *et al* (1993) on the medium and long chain ADHs of horse liver and the short chain ADHs of *S. cerevisiae*. The yeast ADHs were found to have more restricted specificity than the mammalian liver enzymes. The lower activity of the short chain ADH is attributed to the smaller size of the substrate-binding pocket, which may restrict access of bulky substrates to the site. Mutagenesis studies have shown that enlarging the catalytic pocket (e.g. by removal of a methyl group on a branched chain alcohol) significantly increases

the catalytic efficiency on long chain primary alcohols (Green *et al.*, 1993). Green *et al* (1993) thus describe the short chain ADH enzyme as resembling a bottle, with a narrow opening that restricts access of bulky substrates to the catalytic site. The activity of the horse liver enzyme, on the other hand, increases with increasing substrate chain length until it reaches a maximum. This behaviour is generally attributed to the large and open barrel of the substrate-binding site of the medium and long chain ADHs, which allows unrestricted accessibility of the catalytic site to the solvent and substrates (Branden *et al*, 1975) This therefore allows increasing hydrophobic interactions between the hydrocarbon chain and the substrate-binding pocket.

In LADH the NAD<sup>+</sup> is bound in position by multiple residues of the Rossmann fold, a series of  $\alpha$ - $\beta$ - $\alpha$  folds (<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehyznrt.html>).

Some of the NAD-binding residues include Gly<sup>210</sup>, Asn<sup>225</sup>, Pro<sup>243</sup>, Asn<sup>242</sup>, Val<sup>268</sup>, Asp<sup>223</sup>, Try<sup>178</sup>, Arg<sup>47</sup>, Gly<sup>292</sup>, Val<sup>203</sup>. The ADH active site is made up of three amino acid residues, Phe<sup>93</sup>, Leu<sup>57</sup> and Leu<sup>116</sup>, and the three work in concert to provide a cyclohexonal, the three-point binding of the alcohol substrate. It is this mode of binding which accounts for the pro-R hydrogen stereospecificity of ADH.

From above explanation it is thus noticeable that both the size and shape of substrate binding pocket are important for catalytic activity. Moreover, it shows that it is possible to alter the specificity of an ADH enzyme by appropriate substitution of amino acid residues at the substrate-binding site without sacrificing catalytic power over the enzyme. In agreement with this an improved yeast enzyme was constructed by protein engineering where small residues were substituted for large ones (Ser<sup>48</sup> and Phe<sup>93</sup>) in the catalytic pocket (Jörnvall *et al*, 1987). The enzyme had better utilization of 1-octanol than it originally had.

Even though the *Drosophila* ADH (DADH) was the first member of the large family of short chain ADHs (SDRs) discovered, the 3-D structure of the protein

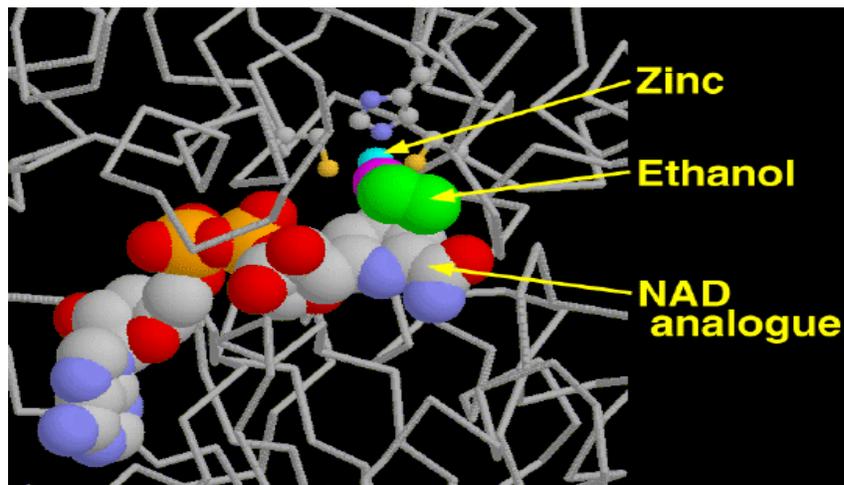
remained elusive for a long time. Predicted structures of members of the group were thus based on amino acid and spectroscopic data only. However, Benach *et al* (1998) provide a list of several SDR members whose crystal structures are now known and among them is the crystal structure of *Drosophila lebanonensis* ADH. The protein is a homodimer with a molecular mass of 27.4kDa. The DADH subunit has an  $\alpha/\beta$  single domain structure with a characteristic Rossmann fold. The active site consists of a deep cavity, covered by a 33-residue flexible loop and an 11-residue C-terminal from the neighbouring subunit. The catalytic triad (Ser<sup>138</sup>, Tyr<sup>151</sup>, Lys<sup>155</sup>) participates in enzymatic catalysis. Hydrogen bond distance to side chains of the triad is three well-ordered water molecules, which may be necessary for proton release during catalysis.

Due to lack of structural information, the iron-containing ADH group remains the least characterized class of ADHs. At this stage we can only assume that the members of this group have a similar topology as the homologous zinc-containing medium chain ADHs since both are metalloenzymes. However, it is obvious that there are features distinct for every class of ADH.

### **Mechanism of action of alcohol dehydrogenase**

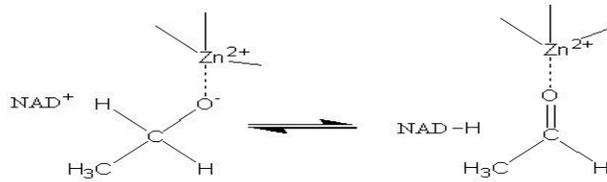
Kinetic studies show that ADH reaction mechanism is ordered, with the enzyme binding first, and for most secondary alcohols the rate-limiting step is the dissociation of NADH (Winberg *et al.*, 1982). In the inactive enzyme the zinc ion is held in place by three ligands to the protein and itself holds tightly to a water molecule as shown in figure 1.2.4. Substrate binding brings about a conformational change that results in expulsion of water molecules in the cleft of the catalytic domain (Eklund, 1989). Thus in presence of substrate or inhibitor the active site is free of water molecules. The bound alcohol becomes ionized and binds NAD<sup>+</sup>, thus helping the reaction to proceed. NAD binding transforms the enzyme from an inactive to an active species. Then there is proton transfer as the negatively charged His<sup>51-</sup> draws a proton from NAD, which in turn draws a proton from Thr<sup>48</sup>, making the threonine ready to accept a proton from the

alcohol. Thus ADH catalyzes oxidation of the alcohol by reducing NAD with a hydride, and this is facilitated by a zinc ion which electrostatically stabilizes the oxygen of the alcohol, thereby increasing the acidity of the proton on the alcohol as shown in figure 1.2.5 (<http://www.users.csbsju.edu/~hjakubow/classes/rasmolchime/99ch331proj/alcoholdehydro/i...>).



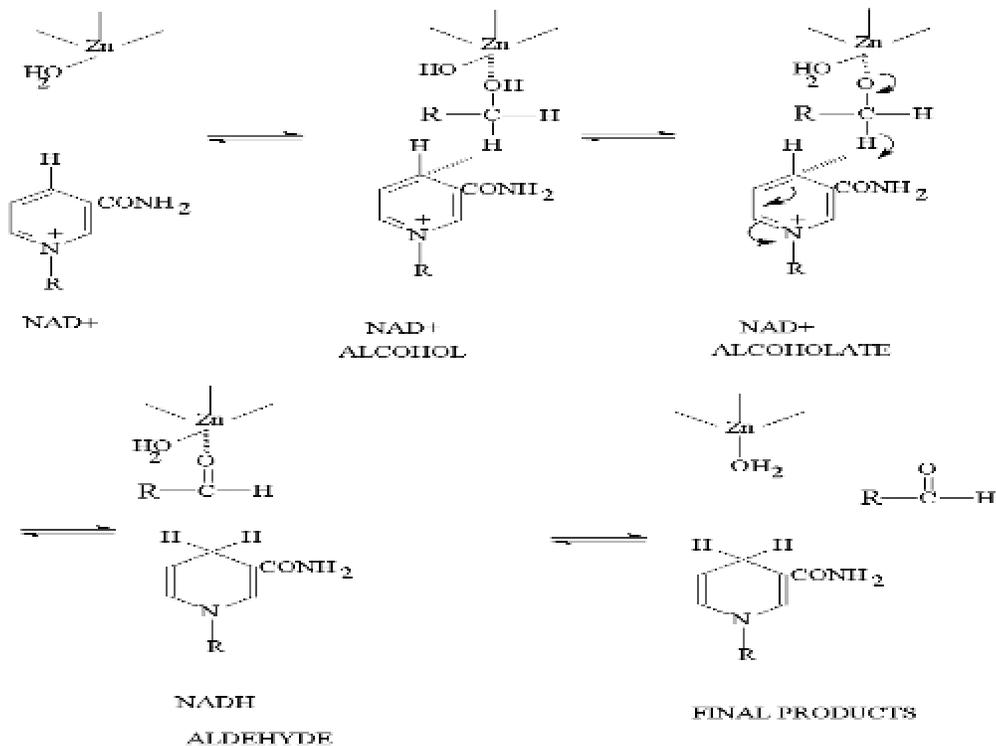
**Figure 1.2.4:** The 3D structure of horse liver ADH showing coenzymes and cofactors important to functioning of the enzyme ([http://www.rcsb.org/pdb/molecules/pdb13\\_3.html](http://www.rcsb.org/pdb/molecules/pdb13_3.html)).

ADH reaction mechanism is a true base catalysis where the substrate is directly involved, and the oxidation is concerted with a traditional hydride transfer to the NAD. The ADH transfer of the hydride to NAD and the oxidation of an alcohol to an aldehyde essentially rely on two key points, namely orientation of the amino acid proton donors and acceptors, as well as the electrostatic stabilization of the negative charge on the alcohol by a zinc ion.



**Figure 1.2.5: Schematic representation of mechanism of action of alcohol dehydrogenase.** Adopted from Enzymes and Enzyme Technology: Available at <http://www.lsbu.ac.uk/biology/enzyme/practical3.html>

The schematic representation of alcohol and NAD binding to ADH upon catalysis is shown in figure 1.2.6.



**Figure 1.2.6: Schematic representation of mechanism of action of alcohol dehydrogenase** available at <http://www.chemistry.ucsc.edu/~fink/231/lecture16.htm>).

### Nomenclature of alcohol dehydrogenases

Since the identification of multiple ADH isozymes in horse, human, mouse and rat a species-specific ADH nomenclature had been adopted. Such nomenclature sufficed at the time but with the extensive and continuous discovery of the ADH

system in almost every living organism so far studied this old type of nomenclature is no longer useful. The use of this old ADH nomenclature has resulted in many orthologous ADHs having the same protein and/or gene names, whereas many orthologous ADHs have different names. Consequently in 1998 the Human Genome Nomenclature Committee (HGNC) was assigned the task of drawing a proposal for revision of the ADH nomenclature (Wain *et al.*, 2002, updated at <http://www.gene.ucl.ac.uk/nomenclature/genefamily/ADH.shtml>). Names of members of the ADH nomenclature committee as well as other relevant data are also available at the ADH website, <http://www.uchsc.edu/sop/alcdibase/adhcov.html>.

The ADH nomenclature committee have reviewed and proposed for all vertebrate ADH a nomenclature consistent between human and mouse as well as for all mammalian ADH forms. This new nomenclature is based upon the human ADH class system of five ADH classes encoded by seven genes (HUGO, 2002; <http://www.gene.ucl.ac.uk/nomenclature/guidelines.html>.) The human ADH class I is encoded by three closely related genes and the other four classes, II- IV, are each encoded by a single gene. The mouse has three ADH classes each encoded by a single gene and all three are orthologs of the three human genes. Therefore, the new ADH gene and protein nomenclature for humans and mouse has resulted in changes in the old nomenclature, but no new name changes as shown in table 1.2.2.

The new ADH nomenclature system is essentially an extension of an ADH nomenclature system originally proposed by Jörnvall and Höög (1995), in which the ADH class system relies upon amino acid sequence homology and catalytic properties or expression patterns to identify orthologs. The ADH nomenclature proposal by Duester *et al.*, (1999) has also been incorporated into the new system. Duester *et al.*, (1999) have actually extended the original idea by Jörnvall and Höög (1995) by adding a set of protein and gene names such that there is only one name for the same protein/gene in different species. They have

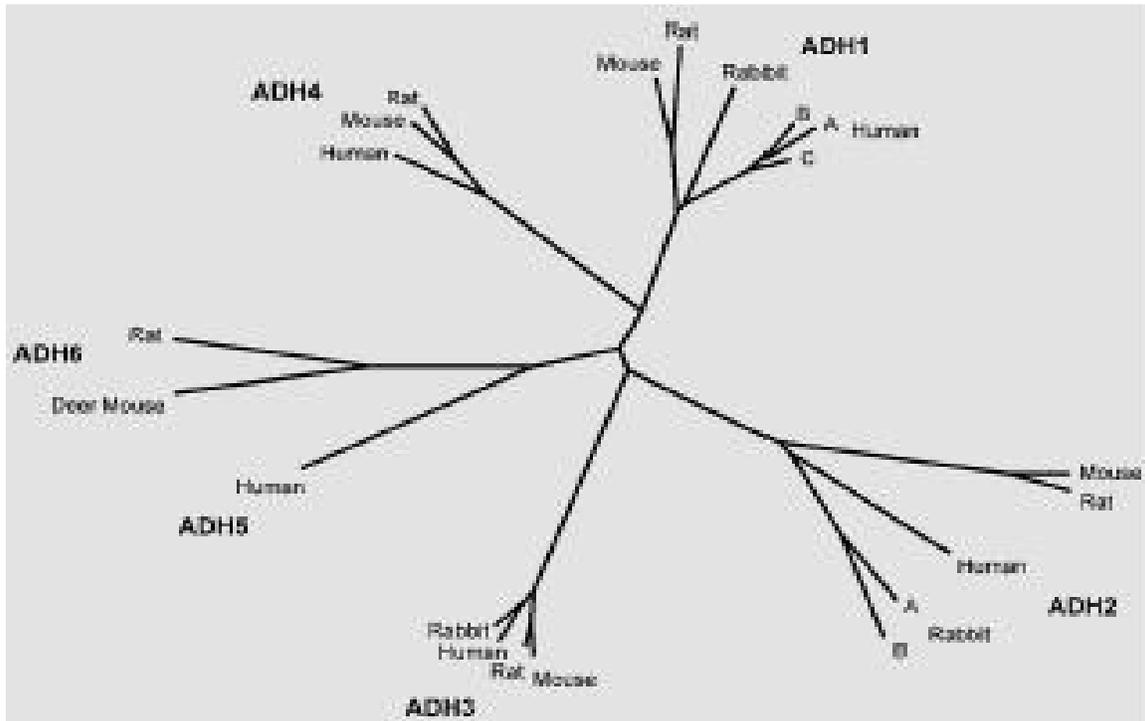
also eliminated the use of Roman numerals and Greek symbols as was recommended by the International Nomenclature Workshop (Blake *et al.*, 1997).

**Table 1.2.2: The new ADH gene and protein nomenclature for humans and mouse, available at <http://www.gene.ucl.ac.uk/nomenclature/genefamily/ADH.shtml>.**

	Old	New		Old	New
Class	human	<b>human</b>	Human Name	mouse	<b>mouse</b>
	symbol	<b>symbol</b>		symbol	<b>symbol</b>
I	ADH1	<b>ADH1A</b>	alcohol dehydrogenase 1A (class I), $\alpha$ -polypeptide	Adh1	<b>Adh1</b>
I	ADH2	<b>ADH1B</b>	alcohol dehydrogenase 1B (class I), $\beta$ -polypeptide		
I	ADH3	<b>ADH1C</b>	alcohol dehydrogenase 1C (class I), ?-polypeptide		
II	ADH4	<b>ADH4</b>	alcohol dehydrogenase 4 (class II), $\rho$ -polypeptide		
III	ADH5	<b>ADH5</b>	alcohol dehydrogenase 5 (class III), ?-polypeptide	Adh5	<b>Adh5</b>
III	ADH5P1	<b>ADH5P1</b>	alcohol dehydrogenase 5 (class III), ?- polypeptide, pseudogene1	Adh5-ps	<b>Adh5-ps</b>
IV	ADH7	<b>ADH7</b>	alcohol dehydrogenase 7 (class IV), $\mu$ - or $\sigma$ -polypeptide	Adh3	<b>Adh7</b>
V	ADH6	<b>ADH6</b>	alcohol dehydrogenase 6 (class V)		

Therefore the new names retain the root symbol “ADH” for “**alcohol dehydrogenase**” followed by an Arabic number instead of a Roman numeral, denoting the ADH class. For multiple isozymes within a class a capital letter instead of a Greek symbol is added after the Arabic number. Thus, the three human class I ADH isozymes previously known as ADH $\alpha$ , ADH $\beta$  and ADH $\gamma$  are now respectively called ADH1A, ADH1B and ADH1C. For the genes the italicized root symbol “*ADH*” is used for human and “*Adh*” for mouse. Thus the three human class I ADH isogenes are *ADH $\alpha$* , *ADH $\beta$*  and *ADH $\gamma$* , and the single mouse, rat and deer mouse class I gene is *Adh1*. The new ADH gene and protein

nomenclature for human and several rodents is shown in table 1.2.2. According to the new system the human class IV and class V genes were originally named out of order, so now the old *ADH6* and *ADH7* become *ADH5* and *ADH4* respectively.



**Figure 1.2.7: A tree showing phylogenetic relationship of the different ADH classes of human and mouse genes.**

Pseudogenes in human and mouse are named to reflect correct ADH class similarly following guidelines already established for human and mouse pseudogene nomenclature (White *et al.*, 1997; Maltais *et al.*, 1997). According to this naming system the human and mouse pseudogenes previously called *ADH5P1* and *Adh5-ps1* are now named *ADH5-P1* and *Adh5-p1* respectively.

Gene families encoding ADHs in plants have also been identified, necessitating the need for uniform ADH nomenclature in plants. Plant ADHs have thus been named following guidelines recommended by the Commission on Plant Gene

Nomenclature (Price *et al.*, 1996, updated at <http://mbclserver.rutgers.edu/CPGN/Guide.html>). Plant genes are placed in plant-wide gene families identified by product family number (PF No) and each gene family consists of genes encoding products with similar sequences (Chase, 1999). Thus names have been assigned only when biochemical function is known, even though in some instances enzymatic activity has not been demonstrated for all members of a sequence group. Genes with identifying function are named with three or four letters and a number, and a single PF No may include more than one function-defined gene family. The three fields, plant species, gene family and a member number, defining membership in multigene families within species, are used to identify genes in individual plant species. For example, ZEAm; Adh1;2 identifies the second member of the Adh1 gene family in maize (*Zea mays*), whereas ORYsa; Fdh1 identifies the glutathione-dependent formaldehyde dehydrogenase of *Oryza sativa*.

Generally, five principal classes of ADH are recognized in plants, and plant ADH genes are named to reflect these classes (Chase, 1999). The five classes are the classic ethanol active ADHs (EC 1.1.1.1) of the gene family *Adh1* and subfamily *Adh0*; the formaldehyde-active class III ADHs (S-hydroxymethylglutathione dehydrogenases, EC 1.2.1.1), family *Fdh1*; the cinnamyl ADHs active in lignin biosynthesis (EC 1.1.1.195), family *Cad2*; the aromatic (benzyl and terpenoid) ADHs (EC 1.1.1.91), families *Bad1* and *Cad1* and lastly, several short chain dehydrogenases (e.g. tropinone reductases and its relatives, EC 1.1.1.206, EC 1.1.1.236 and  $\beta$ -ketoacyl-ACP reductase, EC 1.1.1.212), families *Trr1*, *Trr2*, *Trr0* and *Ker1*.

ADH nomenclature with regard to microbial genes and enzymes is not clearly stipulated in literature. The rules for many organisms are modelled after those of model organisms whose genes/proteins in turn are named according to the recommendations published by an international committee in 1957 (ICGSN 1957). If no rules exist for an organism of interest, the researchers look to these

models for guidance. Thus most microbial ADH genes and enzymes have been named following guidelines for model organisms such as *Escherichia coli* for bacteria (Berlyn, 1998; Demerec, 1966) and, *S. cerevisiae* for yeast and fungi ([http://www.yeast\\_genome.org/gene\\_guidelines.shtml](http://www.yeast_genome.org/gene_guidelines.shtml)). A list of model organisms used in research together with the link to the organism's database which includes gene nomenclature guidelines for the particular organism are available at (<http://www.councilscienceeditors.org/publications/resources.cfm?printPage=1&>).

There are also a number of divergent proteins and/or genes with little overall sequence identity with ADHs, except for a few strongly conserved elements. It has also been observed that in many organisms there are additional proteins which have been identified as ADHs through enzyme activity assays but which lack sequence data to place them definitively within a class. Duester *et al.*, (1999) have recommended that such enzymes exhibiting catalytic properties and tissue distribution profiles of an already known ADH class be referred to as ADH1-like, ADH2-like, etc., until sequence data is available to certainly identify them. Establishment of additional ADH classes is also suggested but only in the presence of enough data to suggest that complete amino acid and other properties indicate a substantial difference from the existing classes.

### **Fungal long chain alcohol dehydrogenases**

As previously stated, the majority of non-conventional yeasts grow very well on long chain alcohols and aldehydes, however very little information is available about genes responsible for metabolism of these substrates by fungi. Only recently Vanhanen *et al.*, (2000) cloned three genes involved in long chain alcohol utilization from *Candida tropicalis* and *Candida cloacae*. These genes encoded fatty acid alcohol oxidases and not dehydrogenases.

Most literature suggests that yeast ADHs, in contrast to the horse ADH are specific for short chain alcohols. However studies by Dickinson and Dack (2001) on yeast ADH I and ADH II show that these enzymes are active towards alcohols

in the range C<sub>2</sub> –C<sub>14</sub>, as well as α- and β-diols. An SFA-encoded enzyme from *Saccharomyces cerevisiae* that oxidizes long chain alcohols, and formaldehyde in the presence of glutathione, has been characterized by Wehner *et al.*, (1993). The glutathione dependent formaldehyde dehydrogenase has also been purified from a mushroom, *Agaricus bisporus* (Norin *et al.*, 2004). ADHs activity specific for long chain alcohols have been described from the majority of alkane degrading yeasts (Yamada *et al.*, 1980; Ueda and Tanaka, 1990; Barth and Kunkel, 1979, Mauersberger *et al.*, 1996).

In actual fact all the ADHs described above are the zinc-dependent medium chain ADHs, even though they are commonly called long chain ADHs, probably from their ability to oxidize long chain alcohols. We can therefore reasonably assume that these enzymes possess the characteristics already described for the zinc-dependent medium chain ADHs. At present no real long chain ADH has been described from fungi.

### **Relationships among the different classes of ADH**

Progress has been increasing rapidly and steadily in the knowledge of ADH families, resulting in increase in size, species variation, characterized tertiary structures, additional forms, known evolutionary properties, novel isozymes and family relationships, including general principles of protein build. As a result a lot of information has been gathered showing that the ADH system consists of a considerable divergence in primary structure, isozyme development, substrate specificity and other functional characteristics. This brief review will therefore not suffice to cover such a huge body of information. However, a few points about similarities, differences and any other relationships among the different classes of ADH will be discussed.

The crucial class distinction in ADHs appears to be enzymatic rather than structure (Jörnvall *et al.*, 1995). Interpretation of single residue replacements at the active site from computer graphics modelling was done to explain the

difference in substrate specificity of the enzyme classes. The same conclusions were later directly confirmed by experimental proof from site-directed mutagenesis, which showed that the separate substrate specificities of ADH classes depend on just a few residue replacements at the active site. Similarly catalytic property differences arise from single active site exchanges between species variant enzymes of the same class. The overall class assignments of structures however change much more slowly, not as a result of such large changes from just limited replacements as seen in substrate specificity (Jörnvall *et al.*, 1995). Thus structural relationships reflect origins, but not necessarily functions, and hence give hint to assignments of classes, which have about 60-70% residue identity, and species variance, which usually exhibit higher percentage conservation (Jörnvall *et al.*, 1995).

Structural similarities between the ADH families involve only part of the molecules and presumably reflect rearranged molecular building units constituting founders for each separate line (Jörnvall *et al.*, 1993). Segments of maximum similarity involve the mononucleotide-binding units, which correspond to positional identities whose overall residue identity is well below 20% between the separate superfamilies, since each segment is only part of the entire subunits.

Consequently, sequence homologies amongst the different classes of NAD(P)-dependent ADHs are found within only two regions namely, the NAD-binding domain and the N-terminal region (Wales and Fewson, 1991). The “fingerprint” motif, GxGxxG (often GxGxxA in NADP-dependent enzymes, where x denotes any other amino acid) for the NAD-binding domain is found towards the C-terminal end (about position 200) of the medium chain zinc-dependent enzymes and close to the N-terminal end of the short chain non-zinc enzymes. This NAD-binding domain motif is also present in other dehydrogenases. The N-terminal regions of the long chain iron-activated and the medium chain zinc-dependent ADHs also appear to share some homology (about 20%), but no other significant

similarities are seen between these two groups of enzymes (Wales and Fewson, 1991). Further comparisons of the ADH system suggest that the zinc-containing ADH though clearly distantly related to the other classes of ADH represents the less evolved ancestral form of ADH and thus may well serve as a model for correlations with other classes of ADH (Jörnvall *et al.*, 1993). As a result this initiates the need for a further look into the zinc-containing group of ADH as a class.

It should however be noted that comparisons of and conclusions about, many known medium chain alcohol dehydrogenases have been made, but in spite of all these the overall comparisons regarding the presently known zinc-containing alcohol dehydrogenases are still lacking (Jörnvall *et al.*, 1997). Only broad and far-reaching relationships showing minor changes between allelic variants have been characterized and correlated with different enzyme properties (Hjelmqvist *et al.*, 2003). In some species isozymes have been characterized and show successive development of the isozymes and enzymes arising from multiple gene duplications at three different levels (Jörnvall *et al.*, 1993).

Another distinguishing feature of ADHs is the strict conservation of structurally important Gly residues within each ADH family, found at reverse turns and those involved in conformation at large (Jörnvall *et al.*, 1995). Such residues constitute the limits that actually define a family, reflecting the true importance of conservation of such turns and conformation in general. The difference in conformational change upon catalysis has also been observed in the different ADH classes, and this has been attributed to the difference in domain movements of three segments, namely a part of the entrance to the active site, a part of the major subunit interacting surface and the loop around the second zinc atom (Jörnvall *et al.*, 1995). Alignments of several alcohol/sorbitol dehydrogenases by Jörnvall *et al.*, (1987) show that few residues are conserved in these enzymes and most of the conserved residues are likely to form typical characteristics of the zinc-containing alcohol/sorbitol dehydrogenases. Several of

these residues may define minimal requirements for correct folding and functional interactions. For example, the notably many glycine residues, are a typical over-representation of distantly related proteins within a family of largely conserved conformations (Jörnvall *et al.*, 1984). The two critical zinc ligands (Cys-46 and His-67) give the alcohol/sorbitol dehydrogenases their general properties. Moreover, some group of acidic residues appears to have defined and critical binding interactions, and the presence of proline and a single branched-chain residue are also possibly important for elements of secondary structure.

Rough estimates of evolutionary relationships amongst 17 alcohol dehydrogenases aligned by Jörnvall *et al.*, (1987) gave values of the same order of magnitude, suggesting that the functional relationships on the corresponding enzymes have remained comparable in the various evolutionary lines. This becomes even more interesting when considering the fact that neither the exact metabolic role nor the substrate specificity of alcohol dehydrogenases is established. As a result whatever may be the role(s) or principal substrate(s), the overall functional properties of ADH have not differed greatly over long periods of time and in widely diverse organisms. This is further confirmed by studies of phylogenetic relationships and rates of evolution of the zinc-containing ADHs from various organisms (Yokoyama *et al.*, 1987, 1990, 1993).

Estimates of the rate of evolutionary change also show that the degree of change is quite large when ADHs are compared with other oligomeric enzymes and other proteins, whereas little is noticeable in comparison to other dehydrogenase enzymes (Jörnvall *et al.*, 1997). It thus appears that in relation to function, the restrictions on evolutionary change have been fewer or more variable in ADH than those on other dehydrogenases. To conclude, change in various lines of ADHs as estimated by amino acid alignments shows values and patterns comparable with those of other protein families, and these can be interpreted in relation to the functional roles of ADHs.

### **1.3 FATTY ALCOHOL AND FATTY ALDEHYDE DEHYDROGENASES OF *YARROWIA LIPOLYTICA***

*Yarrowia lipolytica* is a non-conventional yeast that can use *n*-alkanes and 1-alkenes as sole sources of carbon (Barth and Gaillardin, 1996). *Y. lipolytica* is also a dimorphic, non-pathogenic, ascomycetous yeast (Casaregola *et al.*, 1997). Interest in *Y. lipolytica* research stemmed from the yeast's rather uncommon physiological characteristics. Strains of this species are often isolated from lipid- or protein-rich substrates such as cheese, olive oil and sewage. This feature has been attributed to an efficient synthesis and excretion of proteolytic and hydrophobic substrate degrading enzymes (Barth and Gaillardin, 1997). Sexuality in this yeast was discovered several decades ago (Wickerham *et al.*, 1970), however many unusual features such as low mating frequencies, low fertility of hybrids, irregular meiotic segregation and mitotic hybridization (Casaregola *et al.*, 1997) hampered rapid progress of genetic studies in this organism. In fact, phylogenetic relationship studies based on sequence comparison of genes encoding well-conserved functions, such as glycolytic genes and ribosomal RNA genes, isolate *Y. lipolytica* from the rest of other ascomycetous yeasts and clearly separate it from the *Saccharomyces* yeasts (Barth and Gaillardin, 1996). It has thus become common knowledge that genomic organization of this yeast is radically different from what is seen in other yeasts. Due to this characteristic it has always been a joke among *Y. lipolytica* researchers that the yeast never does anything as expected; rather that one would be looking for one thing and instead find another (Barth and Gaillardin, 1996). Consequently, these unfavourable characteristics seemed to exclude the possibility of performing genetic studies, thus slackening progress of molecular research studies using this organism.

Despite all these problems research into the biology of *Y. lipolytica* did not stop. As research progressed data on genetic and physiological aspects of this yeast were accumulated (Barth and Gaillardin, 1996, 1997) and in 2004 sequencing of the six chromosomes of *Y. lipolytica* was completed (Dujon *et al.*, 2004). The

natural populations of this species widely consist of divergent haploid lines, with absolute requirement for centromeric function for extra chromosomal plasmids (Barth and Gaillardin, 1996). The haploid type strain of the yeast was sequenced. Among many other things it has been found that even though the yeast has a larger genome size (~20kb versus ~12kb in other yeasts) the overall gene density is significantly lower than in most other yeasts (Dujon *et al.*, 2004). The yeast has a highly redundant genome with a strong tendency for map distribution. For example, seven rDNA repeat loci were found in its subtelomeric regions, which is three- to seven-fold the situation in other yeasts (Dujon *et al.*, 2004). Also, 105 copies of rRNA genes (including 11 pseudogenes) were found dispersed throughout the genome of *Y. lipolytica* whereas only one or two copies occur in tandem repeats in other yeasts. Moreover, only 25% of tDNAs has been observed in other yeasts whereas in *Y. lipolytica* about 60% (26 of 44) contain introns (Dujon *et al.*, 2004). This situation has never been observed in other eukaryotes. Horizontal gene transfer is very rare in hemiascomycetes. Only a few examples of genes occurring in only one yeast species that have close homologues in bacteria have been observed, one in *Debaryomyces hansenii*, five in *Kluyveromyces lactis* and as usual *Y. lipolytica* outnumbers them with eight genes (Dujon *et al.*, 2004). The origins and functions of these genes are unknown, and have only been observed to be species specific and form paralogous families in the only species in which they occur.

Following entry into the yeast cell the alkane is oxidized by the cytochrome-P450 monooxygenases to the corresponding *n*-alkan-1-ol, which in turn is oxidized to the corresponding fatty aldehyde. Several NAD(P)-dependent alcohol dehydrogenases which oxidize the alkanol to the corresponding aldehyde have been observed in *Y. lipolytica* (Barth and Kunkel, 1979). However a growing body of evidence suggests that fatty alcohol oxidation in alkane-assimilating yeasts is carried out by fatty alcohol oxidase (FAOD), a non-NAD-requiring enzyme which instead uses molecular oxygen and in the process generates H<sub>2</sub>O<sub>2</sub> (Blasig *et al.*, 1988; Kemp *et al.*, 1991; Mauersberger *et al.*, 1996). The FAOD enzyme has

been purified from *Y. lipolytica*, and in addition presence of several FAOD activities in this yeast has been suggested (Ilchenko *et al.*, 1994). Thus co-existence of both fatty alcohol dehydrogenase (FADH) and fatty alcohol oxidase (FAOD) is strongly suggested in *Y. lipolytica*.

Oxidation of fatty aldehyde into fatty acids has been undoubtedly demonstrated to be effected by a membrane-bound, NAD(P)-dependent fatty aldehyde dehydrogenase (FALDH), following its detection and purification in several yeasts including *Y. lipolytica*, *C. maltosa* and *C. tropicalis* (Barth and Gaillardin, 1996). During growth of *Y. lipolytica* on alkanes many enzymes including FAOD, FADH and FALDH, are induced (Heslot, 1990). However, with the exception of the cytochrome P450 monooxygenases, all the other enzymes are poorly studied in non-conventional yeasts in general. It is thus obvious that as an alkane assimilating yeast *Y. lipolytica* can be used to study enzymes and genes involved in biodegradation of alkanes. Moreover, with the entire genome sequence of this yeast now available (Dujon *et al.*, 2004) the next task is identification and characterization of unknown genes which will reveal more about the uncommon biological and physiological features of this yeast.

## MOTIVATION

The development of research on non-conventional yeasts such as *Candida maltosa*, *Candida tropicalis* and *Yarrowia lipolytica* was initially stimulated in the early 1960's by the ability of these yeasts to grow on a variety of substrates, especially hydrocarbons (Barth and Gaillardin, 1996). The *n*-paraffins being cheap and abundant at the time led to the petrochemical industry being interested in the application of these microorganisms for production of single-cell protein (SCP) as animal feed as well as biochemical synthesis of amino acids, fatty acids, vitamins, sterols and many other biochemicals of commercial interest using *n*-paraffins as raw material (Barth and Gaillardin, 1996). Thus this potential application acted as a stimulant to research into the physiology, biochemistry and genetics of many non-conventional yeasts and to their application in the development of other significant applications. Since then a large amount of data on genetics and physiology of many non-conventional yeasts has been accumulated.

Production of higher-value speciality products by *Y. lipolytica* is of increasing interest for industrial, medical and analytical purposes. During recent years molecular and genetic tools have been developed for this yeast, and it is currently being used as a model organism for fundamental studies on protein secretion, peroxisome biogenesis, dimorphism, alkane and fatty acid biodegradation, (Barth and Gaillardin, 1996; Madzack *et al.*, 2004; Juretzek *et al.*, 2004). In these studies *Y. lipolytica* emerged as one of the most attractive alternative host organism for use in secretion and expression cloning (Muller *et al.*, 1998).

Our research project is part of a bigger project, which aims to develop and evaluate yeast-based processes (microbial processes) for the bioconversion of alkanes produced by the unique South African Fischer-Tropsch process, to a range of value-added products. Specific products being considered are the long-chain mono alcohols, long-chain dicarboxylic acids, biosurfactants, and

(heterologously expressed proteins) enzymes for biocatalysis. One of the aims is thus to produce *Y. lipolytica* strains that will accumulate dicarboxylic acids and mono-alcohols during growth on alkanes. Through our collaboration with Dr. Jean-Marc Nicaud of the INRA-CNRS at Grignon, France, we had access to the necessary cloning systems and to the sequence information as it became available from the *Y. lipolytica* sequencing project.

It is generally assumed that the fatty alcohols and  $\omega$ -hydroxy acids produced, as alkane hydroxylation products by cytochrome P450 monooxygenases in yeasts such as *Y. lipolytica* are further oxidized to the corresponding aldehydes and acids by fatty alcohol oxidases, fatty alcohol dehydrogenases and long-chain aldehyde dehydrogenases. At the beginning of 2001 we therefore started this project with the aim to identify the genes coding for these enzymes in *Y. lipolytica* using the information from the *Y. lipolytica* genome sequencing project as it became available. Using this information the following steps were to be followed;

- (i) Use of northern blot analysis to determine expression and relative importance of these genes during growth on alkanes.
- (ii) Disruption of gene(s) of importance singly or in combination to investigate substrate specificity and to test whether it might lead to accumulation of fatty alcohols and/or fatty aldehydes.

## References

- <http://www.chem.qmw.ac.uk/iupac/bibliog/white.html> Enzyme Nomenclature: Recommendations (1972) of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry on the nomenclature and classification of enzymes, together with their units and the symbols of enzyme kinetics (1973), Elsevier Publishing Co., Amsterdam.
- <http://www.chemistry.ucsc.edu/~fink/231/lecture16.htm> Alcohol dehydrogenase: Chem 231 Lecture No. 16 Sp 1999 (9/10/2005)
- [ICGSN] International Committee on Genetic Symbols and Nomenclature. (1957) Report of the International Committee on Genetic Symbols and Nomenclature. International Union of Biological Sciences, (30): 1-6.
- [No authors listed] (1989) Enzymology and molecular biology of carbonyl metabolism 2: aldehyde dehydrogenase, alcohol dehydrogenase, and aldo-keto reductase. Proceedings of the fourth international workshop. Gifu, Japan, July 4-8, 1988. Progress in clinical and biological research, 290:1-432.
- Barth G, Gaillardin C.** (1996) *Yarrowia lipolytica*. In: K. Wolf, ed. Non-conventional Yeasts in Biotechnology. Berlin, Germany: Springer Verlag, pp 313-388.
- Barth G, Kunkel W.** (1979) Alcohol dehydrogenase (ADH) in yeasts. II. NAD<sup>+</sup> and NADP<sup>+</sup>-dependent alcohol dehydrogenases in *Saccharomycopsis lipolytica*. Zeitschrift für allgemeine Mikrobiologie, 19(6): 381-90.
- Benach J, Atrian S, Gonzalez-Duarte R, Ladenstein R.** (1998) The Refined Crystal Structure of *Drosophila lebanonensis* alcohol dehydrogenase at 1.9Å resolution. Journal of Molecular Biology, 282: 383-99.
- Benner SA, Nambiar KP, Chambers GK.** (1985) A stereochemical imperative in dehydrogenases: new data and criteria for evaluating function-based theories in bioorganic chemistry. Journal of the American Chemical Society, 107(19); 5513-17.

- Berlyn MK.** (1998) Linkage map of *Escherichia coli* K-12. *Microbiology and Molecular Biology Reviews*, 62(3): 814-984.
- Blake JA, Davisson MT, Eppig JT, Maltais LJ, Povey S, White JA, Womack JE.** (1997) MEETING REPORT. *Genomics*, 45(2): 464-8.
- Blasig R, Mauersberger S, Riege P, Schunck W-H, Jockisch W, Franke P, Müller H-G.** (1988) Degradation of long-chain *n*-alkanes by the yeast *Candida maltosa*. II. Oxidation of *n*-alkanes and intermediates using microsomal membrane fractions. *Applied Microbiology and Biotechnology*, 28: 589-597.
- Borras T, Persson B, Jörnvall H.** (1989) Eye lens *zeta*-crystallin relationships to the family of "long-chain" alcohol/polyol dehydrogenases. Protein trimming and conservation of stable parts. *Biochemistry*, 28(15): 6133-9.
- Branden CI, Jörnvall H, Eklund H, Furugren B.** (1975) In: *The Enzymes* (3<sup>rd</sup> edition), Vol. XI, pp103-190, Academic Press, London.
- Casaregola S, Feynerol C, Diez M, Fournier P, Gaillardin C.** (1997) Genomic organization of the yeast *Yarrowia lipolytica*. *Chromosoma*, 106(6): 380-90.
- Chang C, Mann J, Yoshida A.** (1995) Transgenesis of the aldehyde dehydrogenase-2 (ALDH2) locus in a mouse model and in cultured human cells. *Advances in Experimental Medicine and Biology*, 372: 131-136.
- Chang C, Yoshida A.** (1997) Human fatty aldehyde dehydrogenase gene (ALDH10): Organization and tissue-dependent expression. *Genomics*, 40: 80-85.
- Chase T.** (1999) Alcohol Dehydrogenases: Identification and Names for Gene Families. *Plant Molecular Biology Reporter*, 17: 333-50.
- Chervitz SA, Hester ET, Ball CA, Dolinski K, Dwight SS, Harris MA, Juvik G, Malekian A, Roberts S, Roe T, Scafe C, Schroeder M, Sherlock G, Weng S, Zhu Y, Cherry JM, Botstein D.** (1999) Using the *Saccharomyces* Genome Database (SGD) for analysis of protein similarities and structure. *Nucleic Acids Research*, 27(1): 74-78.

- Ciriacy M.** (1997) Alcohol dehydrogenases: In Yeast sugar metabolism: Biochemistry, Genetics, Biotechnology and Applications, pages 213-223, Technomic Publishing Co., Inc., USA.
- Cobessi D, Tete-Favier F, Marchal S, Azza S, Branlant G, Aubry A.** (1999) Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. Journal of Molecular Biology, 290(1): 161-73.
- Conway T, Ingram LO.** (1989) Similarity of *Escherichia coli* propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. Journal of Bacteriology, 171(7): 3754-9.
- Conway T, Sewell GW, Osman YA, Ingram LO.** (1987) Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. Journal of Bacteriology, 169(6): 2591-7.
- Cooper DL, Isola NR, Stevenson K, Baptist EW.** (1993) Members of the ALDH gene family are lens and corneal crystallins. Advances in Experimental Medicine and Biology, 328:169-79.
- Corbier C, Clermont S, Billard P, Skarzynski T, Branlant C, Wonacott A, Branlant G.** (1990) Probing the coenzyme specificity of glyceraldehyde-3-phosphate dehydrogenases by site-directed mutagenesis. Biochemistry, 9(30): 7101-6.
- Danielsson O, Atrian S, Luque T, Hjelmqvist L, Gonzalez-Duarte R, Jörnvall H.** (1994) Fundamental molecular differences between alcohol dehydrogenase classes. Proceedings of the National Academy of Sciences of the United States of America, 91(11): 4980-4.
- Davis BK.** (2002) Molecular evolution before the origin of species. Progress in biophysics and molecular biology, 79(1-3): 77-133.
- Dayhoff MO.** (1976) The origin and evolution of protein superfamilies. Federation Proceedings, 35: 2132-38.
- Demerec M, Adelberg EA, Clark AJ, Hartman PE.** (1966) A proposal for a uniform nomenclature in bacterial genetics. Genetics, 54(1): 61-76.

- Dickinson FM, Dack S.** (2001) The activity of yeast ADH I and ADH II with long-chain alcohols and diols. *Chemico-Biological Interactions*, 130-132(1-3):417-23.
- Dickinson FM.** (1989) Aldehyde dehydrogenase: Kinetics and mechanism. *Biochemical Society Transactions*, 17(2): 299-300.
- Dolinski K, Balakrishnan R, Christie K, Costanzo M, Dwight S, Engel S, Fisk D, Hong E, Issel-Tarver L, Sethuraman A, Theesfeld C, Binkely G, Lane C, Schroeder M, Dong S, Kaloper M, Weng S, Andrada R, Botstein D, Cherry JM.** (SGD - *Saccharomyces Genome Database*: <http://genome-www.stanford.edu/Saccharomyces/>) (12/8/2004).
- Duester G, Farres J, Felder MR, Holmes RS, Höög JO, Pares X, Plapp BV, Yin SJ, Jörnvall H.** (1999) Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochemical Pharmacology*, 58(3): 389-95.
- Edenberg JH, Brown CJ, Hur MW, Kotagiri S, Li M, Zhang L and Zhi X** (1997) Regulation of the seven human alcohol dehydrogenase genes. *Advances in Experimental Medicine and Biology*, 414: 339-45.
- Eklund H** (1989) Nicotinamide nucleotide dependent enzymes. *Biochemical Society Transactions*, 17(2): 293-96.
- Eklund H, Nordstrom B, Zeppezauer E, Soderlund G, Ohlsson I, Boiwe T, Soderberg BO, Tapia O, Branden CI, Akeson A.** (1976) Three-dimensional structure of horse liver alcohol dehydrogenase at 2-4 Å resolution. *Journal of Molecular Biology*, 102(1): 27-59.
- Fan F, Plapp BV.** (1999) Probing the affinity and specificity of yeast alcohol dehydrogenase I for coenzymes. *Archives of Biochemistry and Biophysics*, 367(2): 240-49.
- Feimer J, Xie Y, Takimoto K, Asman D, Pitot H, Lindahl R.** (1995) Class 3 aldehyde dehydrogenase: A northern perspective in the land down under. *Advances in Experimental Medicine and Biology*, 372: 137-41.

- Fox MG, Dickinson FM, Ratledge C.** (1992) Long-chain alcohol and aldehyde dehydrogenase activities in *Acinetobacter calcoaceticus* strain HO1-N. *Journal of General Microbiology*, 138 (Pt 9): 1963-72.
- Gao Z, Loescher WH.** (2000) NADPH supply and mannitol biosynthesis. Characterization, cloning, and regulation of the non-reversible glyceraldehyde-3-phosphate dehydrogenase in celery leaves. *Plant Physiology*, 124(1): 321-30.
- Glasfeld A, Benner SA.** (1989) The stereospecificity of the ferrous-ion-dependent alcohol dehydrogenase from *Zymomonas mobilis*. *European Journal of Biochemistry*, 180(2): 373-5.
- Green DW, Sun HW, Plapp BV.** (1993) Inversion of the substrate specificity of yeast alcohol dehydrogenase. *The Journal of Biological Chemistry*, 268(11): 7792-98.
- Gupta RS.** (1997) Protein phylogenies and signature sequences: evolutionary relationships within prokaryotes and between prokaryotes and eukaryotes. *Antonie Van Leeuwenhoek*, 72(1): 49-61.
- Hempel J, Lindahl R, Perozich J, Wang B-C, Kuo I, Nicholas H.** (1999) Aldehyde dehydrogenase catalytic mechanism: A proposal. *Advances in Experimental Medicine and Biology*, 463: 295-300.
- Hempel J, Nicholas H, Lindahl R** (1993) Aldehyde dehydrogenases: widespread structural and functional diversity within a shared framework. *Protein Science*, 2(11): 1890-900.
- Hempel J, Perozich J, Chapman T, Rose J, Boesch, JS, Liu Z-J, Lindahl R, Wang B-C.** (2001) Beyond the catalytic core of ALDH: a web of important residues begins to emerge. *Chemico-Biological Interactions*, 130-132:39-46.
- Hempel J, Perozich J, Wymore T, Nicholas H** (2003) Algorithm for identification and ranking of family specific residues, applied to the ALDH3 family. *Chemico-Biological Interactions*, 143-144:23-28.

**Heslot H.** (1990) Genetics and Genetic Engineering of the Industrial yeast *Yarrowia lipolytica*. *Advances in Biochemical Engineering/Biotechnology*, 43: 43-73.

**Hjelmqvist L, Norin A, El-Ahmad M, Griffiths W, Jörnvall H.** (2003) Distinct but parallel evolutionary patterns between alcohol and aldehyde dehydrogenases: addition of fish/human betaine aldehyde dehydrogenase divergence. *Cellular and Molecular Life Sciences*, 60: 2009-2016

**Höög JO, Hedberg JJ, Strömberg P, Svensson S.** (2001) Mammalian Alcohol Dehydrogenase – Functional and Structural Implications, *Journal of Biomedical Science*, 8: 71–76.

**Hsu LC, Chang WC, Lin SW, Yoshida A.** (1995) Cloning and characterization of genes encoding four additional human aldehyde dehydrogenase isozymes. *Advances in Experimental Medicine and Biology*, 372: 159-168.

<http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehzyznr.html> ;  
Alcohol Dehydrogenases (6/4/2002).

<http://www.lsbu.ac.uk/biology/enzyme/practical3.html> ; Enzymes and Enzyme  
Technology (4/15/2005).

<http://www.users.csbsju.edu/~hjakubow/classes/rasmolchime/99ch331proj/alcoholdehydro/i...> ; Alcohol Dehydrogenases (4/4/2005).

**Hurley TD, Steinmetz CG, Weiner H.** (1999) Three-dimensional structure of mitochondrial aldehyde dehydrogenase: Mechanistic implications. *Advances in Experimental Medicine and Biology*, 463: 15-25.

**Hurley TD, Weiner H.** (1999) Evaluation of the roles of the conserved residues of aldehyde dehydrogenase. *Advances in Experimental Medicine and Biology*, 463: 45-52.

**Il'chenko AP, Morgunov IG, Honeck H, Mauersberger S, Vasil'kova NN, Müller H-G** (1994) Purification and properties of alcohol oxidase from the yeast *Yarrowia lipolytica* H-222. *Biochemistry (Moscow)*, 59(9): 969-974.

- Inoue T, Sunagawa M, Mori A, Imai C, Fukuda M, Takagi M, Yano K.** (1989) Cloning and sequencing of the gene encoding the 72-kilodalton dehydrogenase subunit of alcohol dehydrogenase from *Acetobacter aceti*. *Journal of Bacteriology*, 171(6): 3115-22.
- IUB-International Union of Biochemistry, Webb EC.** (1973) *Enzyme Nomenclature: Recommendations (1972) of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry on the nomenclature and classification of enzymes, together with their units and the symbols of enzyme kinetics* (1973), Elsevier Publishing Co., Amsterdam
- Jez JM, Flynn TG, Penning TM.** (1997) A new nomenclature for the aldo-keto reductase superfamily. *Biochemical pharmacology*, 54(6): 639-47.
- Jörnvall H, Danielsson O, Eklund H, Hjelmqvist L, Höög JO, Pares X, Shafqat J.** (1993) Enzyme and isozyme developments within the medium-chain alcohol dehydrogenase family. *Advances in Experimental Medicine and Biology*, 328:533-44.
- Jörnvall H, Danielsson O, Hjelmqvist L, Persson B, Shafqat J.** (1995) The alcohol dehydrogenase system. *Advances in Experimental Medicine and Biology*, 372: 281-94.
- Jörnvall H, Höög JO, von Bahr-Lindstrom H, Vallee BL.** (1987) Mammalian alcohol dehydrogenases of separate classes: intermediates between different enzymes and intraclass isozymes. *Proceedings of the National Academy of Sciences of the United States of America*, 84(9): 2580-4.
- Jörnvall H, Höög JO.** (1995) Nomenclature of alcohol dehydrogenases. *Alcohol and Alcoholism*, 30(2): 153-61.
- Jörnvall H, Persson B, Jeffery J.** (1987) Characteristics of alcohol/polyol dehydrogenases: The zinc-containing long-chain alcohol dehydrogenases. *European Journal of Biochemistry*, 167: 195-201.
- Jörnvall H, Persson M, Jeffery J.** (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. *Proceedings of the*

- National Academy of Sciences of the United States of America, 78(7): 4226-30.
- Jornvall H, Shafqat J, el-Ahmad M, Hjelmqvist L, Persson B, Danielsson O.** (1997) Alcohol dehydrogenase variability. Evolutionary and functional conclusions from characterization of further variants. *Advances in Experimental Medicine and Biology*, 414: 281-9.
- Jörnvall H, von Bahr-Lindstrom H, Jeffery J.** (1984) Extensive variations and basic features in the alcohol dehydrogenase-sorbitol dehydrogenase family. *European Journal of Biochemistry*, 140(1): 17-23.
- Jörnvall H.** (1999) Multiplicity and complexity of SDR and MDR enzymes. *Advances in Experimental Medicine and Biology*, 463: 359-64.
- Juretzek T, Le Dall M, Mauersberger S, Gaillardin C, Barth G, Nicaud J.** (2004) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast*, 18(2):97-113.
- Kallberg Y, Oppermann U, Jörnvall H, Persson B.** (2002) Short-chain dehydrogenases/reductases (SDRs): Coenzyme-based functional assignments in completed genomes. *European Journal of Biochemistry*, 269: 4409-4417.
- Karlsson A, El-Ahmad M, Johansson K, Shafqat J, Jörnvall H, Eklund H, Ramaswamy S.** (2003) Tetrameric NAD-dependent alcohol dehydrogenase. *Chemico-Biological Interactions*, 143-144: 239-45.
- Kelson TL, McVoy JRS, Rizzo WB.** (1997) Human liver fatty aldehyde dehydrogenase: microsomal localization, purification, and biochemical characterization. *Biochimica et Biophysica Acta*, 1335: 99-110.
- Kemp GD, Dickinson FM, Ratledge C.** (1991) Activity and substrate specificity of the fatty alcohol oxidase of *Candida tropicalis* in organic solvents. *Applied Microbiology and Biotechnology*, 34: 441-445.
- Kirch H-H, Bartels D, Wei Y, Schnable PS, Wood AJ.** (2004) The ALDH gene superfamily of Arabidopsis. *Trends in plant science*, 9(8): 371-7.
- Kirch HH, Nair A, Bartels D.** (2001) Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant

- Craterostigma plantagineum* and *Arabidopsis thaliana*. *The Plant Journal*, 28(5): 555-567.
- Koivusalo M, Uotila L.** (1993) Glutathione-dependent formaldehyde dehydrogenase/class III alcohol dehydrogenase: further characterization of the rat liver enzyme. *Advances in Experimental Medicine and Biology*, 328:465-74.
- Krozowski Z.** (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *The Journal of Steroid Biochemistry and Molecular Biology*, 51(3-4): 125-30.
- Kutzenko AS, Lamzin VS, Popov VO.** (1998) Conserved supersecondary structural motif in NAD-dependent dehydrogenases. *FEBS Letters*, 423: 105-9.
- Larroy C, Pares X, Biosca JA.** (2002) Characterization of a *Saccharomyces cerevisiae* NADP(H)-dependent alcohol dehydrogenase (ADHVII), a member of the cinnamyl alcohol dehydrogenase family. *European Journal of Biochemistry*, 269(22): 5738-45.
- Lee P, Kuhl W, Gelbart T, Kamimura T, West C, Beutler E.** (1994) Homology between a human protein and a protein of the green garden pea. *Genomics*, 21(2): 371-8.
- Liébecq C.** (1992) *Biochemical Nomenclature and Related Documents*, 2nd edition, Portland Press [ISBN 1-85578-005-4].
- Lin Z, Carney C, Rizzo WB.** (2000) Genomic organization, expression, and alternate splicing of the mouse fatty aldehyde dehydrogenase gene. *Molecular Genetics and Metabolism*, 71: 496-505.
- Lindahl R, Hempel J.** (1991) Aldehyde dehydrogenases: what can be learned from a baker's dozen sequences? *Advances in Experimental Medicine and Biology*, 284:1-8.
- Lindahl R.** (1992) Aldehyde dehydrogenases and their role in carcinogenesis. *Critical Reviews in Biochemistry and Molecular Biology*, 27(4-5): 283-335.
- Liu ZJ, Hempel J, Sun J, Rose J, Hsiao D, Chang WR, Chung YJ, Kuo I, Lindahl R, Wang BC.** (1997a) Crystal structure of a class 3 aldehyde

- dehydrogenase at 2.6 Å resolution. *Advances in Experimental Medicine and Biology*; 414:1-7.
- Liu Z-J, Sun Y-J, Rose J, Chung Y-J, Hsiao C-D, Chang WR, Kuo I, Perozich J, Lindahl R, Hempel J, Wang B-C.** (1997b) The structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Structural Biology*, 4(4): 317-326.
- Luque T, Atrian S, Danielsson O, Jörnvall H, González-Duarte R.** (1994) Structure of the *Drosophila melanogaster* glutathione-dependent formaldehyde dehydrogenase/octanol dehydrogenase gene (class III alcohol dehydrogenase): Evolutionary pathway of the alcohol dehydrogenase genes. *European Journal of Biochemistry*, 225: 985-93.
- Lutstorf U, Megnet R.** (1968) Multiple forms of alcohol dehydrogenase in *Saccharomyces cerevisiae*: Physiological control of ADH-2 and properties of ADH-2 and ADH-4. *Archives of Biochemistry and Biophysics*, 126: 933-44.
- Mackintosh RW, Fewson CA.** (1987) Microbial aromatic alcohol and aldehyde dehydrogenases. *Progress in Clinical and Biological Research*, 232: 259-73.
- Madzak C, Gaillardin C, Beckerich JM.** (2004) Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *Journal of Biotechnology*, 109(1-2):63-81.
- Magonet E, Hayen P, Delforge D, Delaive E, Remacle J.** (1992) Importance of the structural zinc atom for the stability of yeast alcohol dehydrogenase. *Biochemical Journal*, 287: 361-65.
- Maltais LJ, Blake JA, Eppig JT, Davisson MT.** (1997) Rules and guidelines for mouse gene nomenclature: a condensed version. *International Committee on Standardized Genetic Nomenclature for Mice. Genomics*, 45(2): 471-6.
- Masaki R, Yamamoto A, Tashiro Y.** (1994) Microsomal aldehyde dehydrogenase is localized to the endoplasmic reticulum via its carboxy-terminal 35 amino acids. *Journal of Cell Biology*, 126: 1407-20.

- Mauersberger S, Ohkuma M, Schunck WH, Takagi M.** (1996) *Candida Maltosa*  
In: Non-conventional Yeasts in Biotechnology; A handbook. K. Wolf,  
Springer-Verlag, Berlin, Germany, pp435-458.
- Miyauchi K, Masaki R, Taketani S, Yamamoto A, Akayama M, Tashiro Y.**  
(1991) Molecular cloning, sequencing, and expression of cDNA for rat liver  
microsomal aldehyde dehydrogenase. *Journal of Biological Chemistry*,  
266(29): 19536-42.
- Muller S, Sandal T, Kamp-Hansen P, 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(14):1267-83.
- Navarro-Avino JP, Prasad R, Miralles VJ, Benito RM, Serrano R.** (1999) A  
proposal for nomenclature of aldehyde dehydrogenases in  
*Saccharomyces cerevisiae* and characterization of the stress-inducible  
ALD2 and ALD3 genes. *Yeast*, 15: 829-42.
- Neale AD, Scopes RK, Kelly JM, Wettenhall REH.** (1986) The two alcohol  
dehydrogenases of *Zymomonas mobilis*; Purification by differential dye  
ligand chromatography, molecular characterization and physiological  
roles. *European Journal of Biochemistry*, 154: 119-124.
- Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ,  
Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, et al.** (1989) The  
P450 superfamily: updated listing of all genes and recommended  
nomenclature for the chromosomal loci. *DNA*, 8(1): 1-13.
- Nebert DW, Nelson DR, Coon MJ, Estabrook RW, Feyereisen R, Fujii-  
Kuriyama Y, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF,  
et al.** (1991) The P450 superfamily: update on new sequences, gene  
mapping, and recommended nomenclature. *DNA and Cellular Biology*;  
10(1): 1-14.

- Nebert DW.** (2000) Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics*, 10(4): 279-90.
- Nebert DW.** Current status of the IUBMB-recommended nomenclature program for several gene superfamilies. (<http://www.uchsc.edu/sp/sp/alcdbase/aldhcov.html>) (10/5/2005).
- Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, et al.** (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and Cell Biology*, 12(1): 1-51.
- Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW.** (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, 6(1): 1-42.
- Niederhut MS, Gibbons BJ, Perez-Miller S, Hurley TD.** (2001) Three-dimensional structures of the three human class I alcohol dehydrogenases. *Protein Science*, 10(4): 697-706.
- Nordling E, Jörnvall H, Persson B.** (2002) Medium-chain dehydrogenases/reductases (MDR): Family characterizations including genome comparisons and active site modeling. *European Journal of Biochemistry*, 269: 4267-76.
- Norin A, Shafqat J, El-Ahmad M, Alvelius G, Cederlund E, Hjelmqvist L, Jörnvall H.** (2004) Class III alcohol dehydrogenase: consistent pattern complemented with the mushroom enzyme. *FEBS Letters*, 559: 27-32.
- Oppenheimer NJ, Henehan GTM, Huete-Pérez JA, Ito K.** (1997) *Pseudomonas putida* formaldehyde dehydrogenase: An alcohol dehydrogenase masquerading as an aldehyde dehydrogenase. *Advances in Experimental Medicine and Biology*, 414: 417-423.

- Perozich J, Kuo I, Lindahl R, Hempel J.** (2001) Coenzyme specificity in aldehyde dehydrogenase. *Chemico-Biological Interactions*, 130-132(1-3): 115-24.
- Perozich J, Kuo I, Wang BC, Boesch JS, Lindahl R, Hempel J.** (2000) Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. *European Journal of Biochemistry*, 267(20): 6197-203.
- Perozich J, Nicholas H, Lindahl R, Hempel J.** (1999) The big book of aldehyde dehydrogenase sequences: An overview of the extended family. *Advances in Experimental Medicine and Biology*, 463: 1-7.
- Perozich J, Nicholas H, Wang B-C, Lindahl R, Hempel J.** (1999) Relationships within the aldehyde dehydrogenase extended family. *Protein Science*, 8: 137-146.
- Persson B, Bergman T, Keung WM, Waldenstrom U, Holmquist B, Vallee BL, Jörnvall H.** (1993) Basic features of class-I alcohol dehydrogenase: variable and constant segments coordinated by inter-class and intra-class variability. Conclusions from characterization of the alligator enzyme. *European Journal of Biochemistry*, 216(1): 49-56.
- Persson B, Kallberg Y, Oppermann U, Jörnvall H.** (2003) Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chemico-Biological Interactions*, 143-144: 271-78.
- Persson B, Zigler JS, Jörnvall H.** (1994) A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *European Journal of Biochemistry*, 226(1): 15-22.
- Plapp BV, Ganzhorn AJ, Gould RM, Green DW, Jacobi T, Warth E, Kratzer DA.** (1991) Catalysis by yeast alcohol dehydrogenase. *Advances in Experimental Medicine and Biology*, 284: 241-251.
- Plapp BV, Green DW, Sun HW, Park DH, Kim K.** (1993) Substrate specificity of alcohol dehydrogenases. *Advances in Experimental Medicine and Biology*, 328:391-400.

- Price CA, Reardon EM, Lonsdale DM.** (1996) A guide to naming sequenced plant genes. *Plant molecular Biology*, 30: 25-27. Also available at: <http://mbclserver.rutgers.edu/CPGN/Guide.html>
- Qin YM, Poutanen MH, Novikov DK.** (2000) Substrate specificities of peroxisomal members of short-chain alcohol dehydrogenase superfamily: expression and characterization of dehydrogenase part of *Candia tropicalis* multifunctional enzyme. *Journal of Lipid Research*, 41: 93-8.
- Ramaswamy S, Eklund H, Plapp BV.** (1994) Structures of horse liver alcohol dehydrogenase complexed with NAD<sup>+</sup> and substituted benzyl alcohols. *Biochemistry*, 33(17): 5230-7.
- Reid MF, Fewson CA.** (1994) Molecular characterization of microbial alcohol dehydrogenases. *Critical Reviews in Microbiology*, 20(1): 13-56.
- Rizzo WB, Lin Z, Carney G.** (2001) Fatty aldehyde dehydrogenase: structure, expression and mutation analysis in Sjögren-Larsson syndrome. *Chemico-Biological Interactions*, 130-132:297-307.
- Rodriguez-Zavala J, Weiner H.** (2001) Role of the C-terminal tail on the quaternary structure of aldehyde dehydrogenases. *Chemico-Biological Interactions*, 130-132:151-60.
- Rogers GR, Markova NG, De Laurenzi V, Rizzo WB, Compton JG.** (1997) Genomic organization and expression of the human fatty aldehyde dehydrogenase gene (FALDH). *Genomics*, 39: 127-135.
- Rosenstein R, Futter-Bryniok D, Götz F.** (1999) The Choline-Converting Pathway in *Staphylococcus xylosus* C2A: Genetic and Physiological Characterization. *Journal of Bacteriology*, 181(7): 2273-78.
- Rzhetsky A, Ayala FJ, Hsu LC, Chang C, Yoshida A.** (1997) Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory. *Proceedings of the National Academy of Sciences*, 94: 6820-25.
- Saccharomyces Genome Database:** SGD Gene Naming Guidelines. ([http://www.yeastgenome.org/gene\\_guidelines.shtml](http://www.yeastgenome.org/gene_guidelines.shtml)) (4/7/2005).

- Sanfacon R, Rouillard R, Goupil M, Heick HMC.** (1971) Alcohol dehydrogenase activity in the yeast *Saccharomyces fragilis*. *Canadian Journal of Microbiology*, 18: 35-44.
- Scopes RK.** (1983) An iron-activated alcohol dehydrogenase. *FEBS Letters*, 156(2): 303-6.
- Shafqat J, El-Ahmad M, Danielsson O, Martinez MC, Persson B, Pares X, Jörnvall H.,** (1996) Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). *Proceedings of the National Academy of Sciences*, 93: 5595-99.
- Shafqat J, Höög J-O, Hjelmqvist L, Opperman U, Ibanez C, Jörnvall H.** (1999) Studies on variants of alcohol dehydrogenase and its domains. *Advances in Experimental Medicine and Biology*, 463: 285-329.
- Shah P, Pietruszko R.** (1999) Reaction-chemistry sequence alignment of aldehyde dehydrogenases. *Advances in Experimental Medicine and Biology*, 463: 9-14.
- Singer ME, Finnerty WR.** (1985) Fatty aldehyde dehydrogenases in *Acinetobacter sp.* strain HO1-N: role in hexadecane and hexadecanol metabolism. *Journal of Bacteriology*, 164(3): 1011-16.
- Sophos NA, Pappa A, Ziegler TL, Vasiliou V** (2001) Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chemico-Biological Interactions*, 130-132:323-37.
- Sophos NA, Vasiliou V.** (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chemico-Biological Interactions*, 143-144: 5-22
- Sridhara S, Wu TT, Chused TM, Lin EC.** (1969) Ferrous-activated nicotinamide adenine dinucleotide-linked dehydrogenase from a mutant of *Escherichia coli* capable of growth on 1, 2-propanediol. *Journal of Bacteriology*, 98(1): 87-95.
- Steinmetz CG, Xie P, Weiner H, Hurley TD.** (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure*, 5: 701-11.

- Tasayco J, Prestwich GD.** (1990) A specific affinity reagent to distinguish aldehyde dehydrogenases and oxidases. *Journal of Biological Chemistry*, 265: 3090-3101.
- Tatusov RL, Koonin EV, Lipman DJ.** (1997) A genomic perspective on protein families. *Science*, 278: 631-637.
- Ueda M, Tanaka A** (1990) Long-chain aldehyde dehydrogenase of *Candida* yeast. *Methods in Enzymology*, 188:176-8.
- Vanhanen S, West M, Kroon JT, Lindner N, Casey J, Cheng Q, Elborough KM, Slabas AR.** (2000) A consensus sequence for long-chain fatty-acid alcohol oxidases from *Candida* identifies a family of genes involved in lipid omega-oxidation in yeast with homologues in plants and bacteria. *Journal of Biological Chemistry*, 275(6): 4445-52.
- Van Loon APGM, Young ET.** (1986) Intracellular sorting of alcohol dehydrogenase isozymes in yeast: a cytosolic location reflects absence of an amino-terminal targeting sequence for the mitochondrion. *The EMBO Journal*, 5(1): 161-65.
- Van Ophem PW, van Beeumen J, Duine JA.** (1992) NAD-linked, factor-dependent formaldehyde dehydrogenase or trimeric, zinc-containing, long-chain alcohol dehydrogenase from *Amycolatopsis methanolica*. *European Journal of Biochemistry*, 206: 511-18.
- Vasiliou V,** Alcohol Dehydrogenase Gene Superfamily (<http://www.uchsc.edu/sop/alcdbase/adhcov.html>) (10/5/2004).
- Vasiliou V, Kozak CA, Lindahl R, Nebert DW.** (1996) Mouse microsomal Class 3 aldehyde dehydrogenase: AHD3 cDNA sequence, inducibility by dioxin and clofibrate, and genetic mapping. *DNA and Cell Biology*, 15(3): 235-45.
- Vasiliou V, Pappa A, Petersen DR** (2000) Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-Biological Interactions*, 129:1-19.

- Vasiliou V, Weigner H, Marselos M, Nebert DW.** (1995) Aldehyde dehydrogenase genes: classification based on evolution, structure and regulation. *European Journal of Biochemistry*, 20: 53-64.
- Vasiliou V.** (1997) Aldehyde dehydrogenase genes. *Advances in Experimental Medicine and Biology*, 414: 595-600.
- Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S.** (2002) Guidelines for Human Gene Nomenclature. *Genomics*, 79(4): 464-70 (<http://www.gene.ucl.ac.uk/nomenclature/guidelines.html>).
- Wales MR, Fewson CA.** (1991) Comparison of NAD(P)-dependent bacterial alcohol dehydrogenases. *Advances in Experimental Medicine and Biology*, 372: 295-300.
- Wang X, Bai Y, Ni L, Weiner H.** (1997) *Saccharomyces cerevisiae* aldehyde dehydrogenases: Identification and expression. *Advances in Experimental Medicine and Biology*, 414: 277-80.
- Wehner EP, Rao E, Brendel M.** (1993) Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Molecular and General Genetics*, 237(3): 351-8.
- Wenger JI, Bernofsky C.** (1971) Mitochondrial alcohol dehydrogenase from *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta*, 227:479-90.
- White JA, McAlpine PJ, Antonarakis S, Cann H, Eppig JT, Frazer K, Frezal J, Lancet D, Nahmias J, Pearson P, et al.** Guidelines for human gene nomenclature (1997). *Genomics*, 45(2): 468-71. Also available from <http://www.gene.ucl.ac.uk/nomenclature/guidelines.html>.
- Wickerham LJ, Kurtzman CP, Herman AI.** (1970) Sexual reproduction in *Candida lipolytica*. *Science*, 167(921):1141.
- Williamson VM, Paquin CE.** (1987) Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. *Molecular and general genetics*, 209(2): 374-81.
- Wood AJ, Saneoka H, Rhodes D, Joly RJ, Goldsbrough PB.** (1996) Betaine aldehyde dehydrogenase in sorghum. *Plant Physiology*, 110(4): 1301-8.

- Wymore T, Deerfield DW, Nicholas HB, Hempel J, Field MJ.** (2002) What happens at the active site: Better understanding of an enzyme that metabolizes chemotherapy drugs may permit better cancer-fighting with smaller doses. Projects in Scientific Computing. [http://www.psc.edu/science/2002/wymore/what\\_happens\\_at\\_the\\_active\\_site.html](http://www.psc.edu/science/2002/wymore/what_happens_at_the_active_site.html) (4/15/2005).
- Wymore T, Field MJ, Deerfield DW, Nicholas HB, Hempel J.** (2004) Mechanistic details of aldehyde dehydrogenase chemistry from MM and QM/MM simulations. <http://www.psc.edu/biomed> (4/15/2005).
- Wymore T, Nicholas HB, Hempel J, Field MJ,** (2001) Molecular dynamics simulation of class 3 aldehyde dehydrogenase. *Chemico-Biological interactions*, 130: 201-07.
- Yamada K, Mori H, Yamaki S.** (1999) Identification and cDNA cloning of a protein abundantly expressed during apple fruit development. *Plant and Cell Physiology*, 40(2): 198-204.
- Yamada T, Nawa H, Kawamoto S, Tanaka A, Fukui S.** (1980) Subcellular localization of long-chain alcohol dehydrogenase and aldehyde dehydrogenase in n-alkane-grown *Candida tropicalis*. *Archives of Microbiology*, 128(2): 145-51.
- Yokoyama S, Harry DE.** (1993) Molecular phylogeny and evolutionary rates of alcohol dehydrogenases in vertebrates and plants. *Molecular Biology and Evolution*, 10(6): 1215-26.
- Yokoyama S, Yokoyama R, Kinlaw CS, Harry DE.** (1990) Molecular evolution of the zinc-containing long-chain alcohol dehydrogenase genes. *Molecular Biology and Evolution*, 7(2): 143-54.
- Yokoyama S, Yokoyama R.** (1987) Molecular evolution of mammalian class I alcohol dehydrogenase. *Molecular Biology and Evolution*, 4(5): 504-13.
- Yoshida A, Rzhetsky A, Hsu LC, Wang C.** (1998) Human aldehyde dehydrogenase gene family: Review. *European Journal of Biochemistry*, 259: 549-57.

**Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E.** (2001) Differences in nucleotide specificity and catalytic mechanism between *Vibrio harveyi* aldehyde dehydrogenase and other members of the aldehyde dehydrogenase gene superfamily. *Chemico-Biological Interactions*, 130-132:29-38.

**Zinovieva RD, Tomarev SI, Piatigorsky J.** (1993) Aldehyde dehydrogenase-derived omega-crystallins of squid and octopus. Specialization for lens expression. *Journal of Biological Chemistry*, 268(15): 11449-55.

## CHAPTER 2

### An exploratory study into the presence of long chain alcohol and aldehyde dehydrogenases in *Yarrowia lipolytica*

#### 2.1 Introduction

Hydroxylation of *n*-alkanes by alkane-assimilating microorganisms produces the long chain alcohols, which are oxidized to the corresponding fatty acids via aldehydes by the action of fatty alcohol oxidase (FAOD) and/or fatty alcohol dehydrogenase (FADH), and fatty aldehyde dehydrogenase (FALDH) (Blasig *et al.*, 1989, I'chenko *et al.*, 1994, Krauzova *et al.*, 1984). The results of investigations on subcellular distribution of FAOD, FADH and FALDH of alkane-degrading yeasts have been described in several papers (Yamada *et al.*, 1980; Mauersberger *et al.*, 1987). The activity of these enzymes has been detected in membranes of different subcellular organelles, namely mitochondria, peroxisomes or microsomes (Yamada *et al.*, 1980). Although these enzymes have been shown to be abundant in yeast cells (e.g. *Candida tropicalis*, *Yarrowia lipolytica*) growing on alkanes, only FAOD has been purified from these yeasts (Dickinson and Wadforth, 1992; I'chenko *et al.*, 1994; Hommel *et al.*, 1994), but no FADH or FALDH has yet been purified to homogeneity.

Since the late 1980's several papers have disproved involvement of FADH in alkane degradation by yeasts and instead endorse existence of the non-NAD(P)<sup>+</sup>-dependent FAOD (Kemp *et al.*, 1988, 1994; I'chenko *et al.*, 1994; Mauersberger *et al.*, 1992; Hommel *et al.*, 1994). These authors argued that reduction of NAD(P)<sup>+</sup> by membrane fractions observed during oxidation of fatty alcohols is only observable under aerobic but not under strictly anaerobic conditions, indicating that the NAD(P)<sup>+</sup> reduction is due to the activity of NAD(P)<sup>+</sup>-dependent FALDH oxidizing aldehyde formed by the FAOD reaction to the corresponding fatty acid (Krauzova *et al.*, 1984; Mauersberger *et al.*, 1987).

Thus according to these authors there is no membrane-bound FADH in alkane-degrading yeasts so far tested (*Candida*, *Torulopsis*, *Debaryomyces*). On the contrary, soluble FADH and FALDH activities have been demonstrated in *C. maltosa* (Mauersberger *et al.*, 1996). Moreover, a number of discrete alcohol dehydrogenases but no alcohol oxidases have been identified in several alkane-degrading bacteria (Fox *et al.*, 1993).

From this discussion then a question arises whether fatty alcohol oxidizing enzymes, FADH and FAOD, co-exist in alkane-utilizing cells of *Y. lipolytica*. Presence of FALDH activity is indisputable, but presence of multiple FALDH enzyme activities in this yeast has never been mentioned before. Consequently we undertook an exploratory study into the expression of genes encoding these enzymes in this yeast.

## **2.2 MATERIALS AND METHODS**

### **2.2.1 Growth of organisms**

#### **2.2.1 (a) Pre-cultures**

Cultures were grown in 10ml YP broth in 100ml flasks. The YP broth incubated on a shaker for 24h at 30<sup>0</sup>C, contained 2% v/v hexadecane, 2% w/v peptone and 1% w/v yeast extract. This preculture was used to inoculate main cultures to an OD<sub>620nm</sub> of between 0.03 and 0.05.

#### **2.2.1(b) Growth of cells for FALDH and FADH enzyme assays**

Duplicate 100ml cultures in 2-litre flasks were grown in YNB broth supplemented with 0.5% w/v yeast extract and 1% each of the following substrates as sole source of carbon; glucose, glycerol and hexadecane. Growth in these substrates was followed by taking OD<sub>620nm</sub> readings every 4 hours until stationary phase. At the same time 10ml cell sample were withdrawn from each culture and cells harvested by centrifugation (3000 x g, 10 min). The cell pellet obtained was washed twice in deionized water and used immediately or frozen away at -20<sup>0</sup>C for later use.

### **2.2.1(c) Glucose derepressed cells for total RNA isolation**

The preculture was used to inoculate 11 main cultures (50ml in one litre flasks) to an initial OD<sub>620nm</sub> reading of 0.03 to 0.05 in YP media containing 2% glucose. The cells were grown in glucose for 24 hours after which they were harvested and then starved by resuspension and shaking for 4 hours in yeast nitrogen base (YNB) supplemented with 0.5% w/v yeast extract, but without any carbon source. After 4 hours different carbon sources were added to the starved cells and shaking continued for 5 hours. The carbon sources were 4% glucose, 4% glycerol, 2% each of dodecane, dodecanol, 12-hydroxy dodecanoic acid plus 0.1% Tween80 and dodecanediol, 2% docosane plus 2% pristane, 2% docosanol plus 2% pristane, 4% glucose plus 1% dodecanol, 4% glycerol plus 1% dodecanol and 4% glycerol plus 1% dodecanol. Two control cultures were included with one containing 0.1%v/v Tween80 but no carbon source and another one containing starved cells in YNB without either carbon source or Tween80. Pristane is often used as co-solvent for insoluble alkanes such as C<sub>22</sub> alkane (McKenna and Kallio, 1971), while Tween80 is used as an emulsifier for organic acids (Wu *et al.*, 2001). Total RNA was isolated from all above cultures using a combination of liquid nitrogen and the Trisol<sup>®</sup> reagent (Invitrogen Life Technologies). The isolated RNA was stored in formamide at -80<sup>0</sup>C until later use.

### **2.2.2 Preparation of enzyme extracts**

#### **2.2.2 (a) Preparation of cell-free enzyme extracts**

The cell pellet obtained from cultures described in section 2.2.4(c) was mixed with 1 volume glass-beads (450-600µm) and ½ volume Tris-HCl buffer (50mM, pH 8.5) and then broken by vortexing in bursts of 1 min vortexing and 1 min cooling on ice for 3 min. The crude cell extract thus obtained was centrifuged (500 x g, 10 min) to remove cell debris and unbroken cells. The cell-free extract was used to perform the enzyme assays described in section 2.2.6 and for subcellular fractionation.

### **2.2.2(b) Subcellular fractionation of enzyme activities**

The cell-free extract was centrifuged at 5000 x g for 15 min to obtain the supernatant fraction (S1) and the particulate fraction pellet (P1). The S1 fraction was centrifuged at 22 000 x g for 30 min resulting in the second supernatant fraction (S2) and a pellet (P2). All fractions obtained (including particulate pellet fractions) were normalized to same volume as the original crude extract (about 10ml) with buffer prior to assaying. All fractions obtained were assayed for long chain alcohol and aldehyde dehydrogenase (FADH and FALDH respectively).

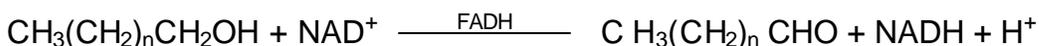
### **2.2.2 (c) Y-PER™ treatment of cells**

Crude enzyme extract from cells treated with Y-PER™ protein extraction reagent (Pierce Chemical Co., US) before cell disruption, were also used for subcellular fractionation. The pellet from harvested cells was resuspended uniformly in Y-PER™ reagent (1ml Y-PER™ reagent/1g wet cells) by gentle vortexing. The suspension was incubated for 40 min at room temperature with gentle agitation. The cell pellet was then separated from cell lysate (Y-PER extract) by centrifugation at 13 000 x g for 10 min. The cells were then broken with glass beads as described in section 2.2.5(a). When using the Y-PER™ protein extraction reagent prior to cell disruption the reagent is supposed to extract all soluble cytoplasmic fractions thereby separating them from the non-soluble, membrane-bound enzyme fractions. As a control the short chain alcohol dehydrogenase (ADH), which is a soluble cytoplasmic protein, was therefore assayed, by using ethanol as a substrate in the FADH/FALDH assay.

## 2.2.3 Enzyme assays

### 2.2.3(a) Long chain alcohol dehydrogenase (FADH) and fatty aldehyde dehydrogenase (FALDH) assays

#### Reactions:



#### Assay Method

*Principle:* The reduction of  $\text{NAD}^+$  (*i.e.* NADH formation, noticeable by increase in absorbance) is measured at  $30^\circ\text{C}$  with a recording spectrophotometer at 340nm (Ueda and Tanaka, 1990) for 5 min at 10 sec interval. The rate of activity was calculated using an absorbance coefficient ( $\epsilon$ ) of  $6.27\text{M}^{-1}\text{cm}^{-1}$  for NADH, with one mole of aldehyde/acid being produced per mole of substrate oxidized. Enzyme activity was expressed as nmol/min/mg protein.

*Procedures:* All reagents were added to a 96-well microtitre plate as shown below. The reaction was initiated by adding the enzyme extract. The blank was prepared with enzyme extract and all other reagents except for the substrate. Each sample was assayed in triplicate and the experiment was repeated twice.

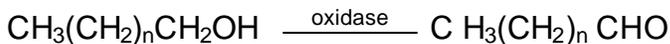
<i>Reagents</i>	<i>Assay mixture (volumes)</i>
50mM Tris-HCl buffer, pH8.5	37ul
20mM substrate (alcohols in DMSO, aldehydes in dioxane)	13ul
13mM $\text{NAD}^+$	50ul
13mM $\text{NaN}_3$ (sodium azide)	50ul
Enzyme extract (10X diluted)	50ul
Total volume	200ul

### 2.2.3(b) NAD and NADP dependence of fatty aldehyde dehydrogenase

Enzyme assays as described above for FADH and FALDH were carried out using both NAD<sup>+</sup> and NADP<sup>+</sup> (Roche Diagnostics) as cofactors. Cell-free crude extracts from glucose- and hexadecane-grown cells were assayed. The experiment was repeated twice with each sample assayed in triplicate.

### 2.2.3(c) Fatty alcohol oxidase assay

#### Reaction:



#### Assay Method:

Fatty alcohol oxidase activity is determined spectrophotometrically at 405nm in coupled assays with 2,2' azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) and peroxidase (Kemp *et al.*, 1988). The method involves two successive enzymatic reactions (Verduyn *et al.*, 1984);

- (i) Substrate + O<sub>2</sub>  $\xrightarrow{\text{alcohol oxidase}}$  product + H<sub>2</sub>O<sub>2</sub> followed by,
- (ii) H<sub>2</sub>O<sub>2</sub> + ABTS  $\xrightarrow{\text{peroxidase}}$  H<sub>2</sub>O + ABTS<sub>ox</sub>

<i>Reagents</i>	<i>Assay mixture (volumes)</i>
50mM glycine-NaOH buffer (PH 9.0 plus 6mM sodium azide)	88ul
3.5mM substrate (alcohol) in DMSO	20ul
Peroxidase (0.5ug/ul)	26ul
10mM ABTS	16ul
Enzyme extract (10X diluted)	50ul
Total volume	200ul

*Principle:* The spectrophotometric assay method measures the liberation of ABTS<sub>ox</sub> at 405nm, and it is based on the methanol oxidase assay of Haywood

and Large (1981). Readings were taken for 5 min at 10 sec interval. The molar extinction coefficient ( $\epsilon$ ) for the radical cation of ABTS is  $18.4\text{mM}^{-1}\text{cm}^{-1}$  with 1mM radical cation of ABTS being produced per mol of substrate oxidized. Enzyme activity is expressed as nmol of product formed/ min per mg of protein.

*Procedures:* All reagents were added to a 96-well microtitre plate as shown above. The blank was prepared with enzyme extract and all other reagents except for the substrate. The reaction was initiated by adding the enzyme extract. Each sample was assayed in triplicate and the experiment was repeated twice.

#### **2.2.4 BLAST searches for putative FAOD, FADH and FALDH genes**

Shortly after completion of phase I of the Genolevures sequencing project (Gènolevures consortium, personal communication, <http://cbi.labri.u-bordeaux.fr/Genolevures/index.php>, described by Artiguenave *et al.*, 2000; Sherman *et al.*, 2004; Dujon *et al.*, 2004) the *Y. lipolytica* database was searched (BLASTP and TBLASTN searches) for putative FAOD, FALDH and FADH encoding genes. The *Saccharomyces cerevisiae* SFA1 sequence (encoding FADH) (Wehner *et al.*, 1993; accession no. X68020) was used as query in searches for putative FADH-like gene sequences. The deduced amino acid sequence from the FAOD encoding gene of *C. tropicalis* (Vanhanen *et al.*, 2000; accession no. AJ242496) was used to search for FAOD encoding genes and the protein sequence of the human FALDH (Accession no. P51648 or U46689) was used to search for FALDH encoding genes.

#### **2.2.5 Identification of FAOD, FALDH and FADH genes in *Y. lipolytica***

##### **2.2.5(a) PCR amplification of FALDH and $\beta$ -actin genes**

Sequence information of four \*clones containing partial gene sequences of *ALDH* 1 and 2, and *FALDH* 3 and 4, and the appropriate oligonucleotide primers (table 2.1) were used for PCR amplification of the four partial *ALDH* gene

---

\* Courtesy of Dr J-M Nicaud, France.

sequences. The four partial *ALDH* DNA sequences were also amplified using genomic DNA of *Y. lipolytica*. The PCR mixture (50µl) contained approximately 500ng/µl template DNA, 0.2mM dNTPs, 100pM of each of the two primers, MgCl<sub>2</sub>-containing PCR buffer and one unit of Taq polymerase (Roche Diagnostics). PCR conditions for 25 cycles were as follows; denaturation at 94<sup>0</sup>C for 1 min, annealing at 50<sup>0</sup>C for 1 min, and extension at 72<sup>0</sup>C for 1 min. For PCR amplification of *ALDH2*, however the annealing temperature was 48<sup>0</sup>C. PCR amplification of the β-actin gene was also performed under similar conditions with annealing temperature of 55<sup>0</sup>C. The size of PCR products was determined on a 1% agarose gel electrophoresed at 80V. The PCR products were then gel purified using a DNA purifying kit (Amersham) and the isolated PCR products were stored at -20<sup>0</sup>C for later use in Southern and northern blot analyses.

**Table 2.1: Oligonucleotide sequence of primers used in this study.**

Gene	Primer	Sequence (5'® 3')
ALDH1	FA 1-1F	ATCAACAACGAGTGGTC
	FA 1-1R	CAGCGAAATGGAGATGG
ALDH2	FA 2-1F	GAGAACTACCTCCAGAC
	FA 2-2R	GACCATCTCGACCCATC
FALDH3	FA 3-1F	CGATTCTCTGATGCAGG
	FA 3-2R	TTGGCATCCACTCCGTC
FALDH4	FA 4-1F	AAAATGTCTACCTTTGATTGG
	FA 4-1R	CATCGTTACATGCTGGC
FADH	LADH1-1F	CGC CTA CAC TCT CAG TGG CTC TGA
	LADH 1-1R	TGA AAG TCG CAA ACC TCC GTC GGG
*FAOD ( <i>C. tropicalis</i> )	C-trop-FAOD-1F	ATGGCTAGTTTTTTACCAGACAAAG
	C-trop-FAOD-1R	CTACAATTTAGCCTTGGATTTCAAG

\* FAOD from *Y. lipolytica* failed to PCR amplify

### **2.2.5(b) PCR amplification of *FADH* gene**

A pair of oligonucleotides was designed based on available partial <sup>S</sup> *FADH* gene sequence (table 2.1). The partial *FADH* DNA sequence was PCR amplified from *Y. lipolytica* H222 genomic DNA under the following conditions; the PCR mixture (50µl) contained approximately 500ng/µl genomic DNA, 0.2mM dNTPs, 100pM of each of the two primers, MgCl<sub>2</sub>-containing PCR buffer and one unit of Taq polymerase (Roche Diagnostics). PCR conditions for 25 cycles were as follows; denaturation at 94<sup>0</sup>C for 1 min, annealing at 50<sup>0</sup>C for 1 min, and extension at 72<sup>0</sup>C for 1 min. The size of the resulting PCR product was determined on a 1% agarose gel electrophoresed at 80V. The PCR product was then isolated, purified using a DNA purifying kit (Amersham) and the purified PCR product was stored at -20<sup>0</sup>C for later use in Southern and northern blot analyses.

### **2.2.5(c) PCR amplification of *FAOD* gene in *C. tropicalis* OC3**

From the nucleotide sequence of *C. tropicalis* described by Vanhanen *et al.* (2000) (Accession no. AJ242496) a pair of synthetic oligonucleotides (25-mers) were constructed from both ends of the sequence (table 2.1). The PCR reaction mixture (50µl) contained approximately 500ng/µl genomic DNA, 0.2mM dNTPs, 100pM of each of the two primers, MgCl<sub>2</sub>-containing PCR buffer and 1µl Taq polymerase (Roche Diagnostics). Amplification was performed under the following conditions for 25 cycles; denaturation at 94<sup>0</sup>C for 1 min, annealing at 55<sup>0</sup>C for 1 min, and extension at 72<sup>0</sup>C for 2 min. The amplified DNA fragment was Dig-labelled using the DIG DNA labelling and Detection kit (Roche Diagnostics) following the manufacturer's protocol, and then used as a probe in Southern hybridization experiments.

---

\* Courtesy of Dr J-M Nicaud, France.

### **2.2.5(d) Southern Hybridization**

*Yarrowia lipolytica* H222 genomic DNA (700ng) digested individually with *Acc65I*, *BamHI*, *EcoRI*, *HindIII*, *PstI*, *Sall* and *XbaI*, was electrophoresed on a 1% agarose gel and then blotted onto a nylon transfer membrane (0.22 $\mu$ , magnacharge, Osmonics Inc.) using capillary method of transfer in 10X SSC buffer. All *FAOD* and *ALDH* DNA sequences obtained by PCR (section 2.2.1(a), (b) and (c)) were Dig-labelled (Roche Diagnostics) and then used individually for overnight hybridization of the blotted membranes. In addition, Dig-labelled  $\lambda$  DNA was added to the hybridization mixture to allow visualization of the molecular weight marker. Low and high stringency hybridizations were performed at 52 $^{\circ}$ C and 68 $^{\circ}$ C respectively. Thereafter the blots were similarly washed with both low and high stringency buffers. The Dig labelling efficiency of PCR probes was also determined as a control measure. Southern hybridization using  $\beta$ -actin as a probe was also included as a control.

### **2.2.6 Expression analysis of *ALDH* and *FADH* genes**

#### **2.2.6(a) Northern Hybridization**

Total RNA isolated from cells grown in different carbon sources as described in section 2.2.1(c) was electrophoresed on a 1% agarose gel containing 2% formaldehyde. The RNA was transferred onto a nylon membrane (0.22 $\mu$ , magnacharge, Osmonics Inc.) by capillary blotting, using 10X SSC. The membranes were probed by overnight hybridization at 65 $^{\circ}$ C with the radiolabelled ( $^{32}$ P-dATP) PCR amplified gene sequences of *ALDH* 1 and 2, *FALDH* 3 and 4, *FADH*, *FAOD* and  *$\beta$ -actin*. The membranes were then washed twice each with a low stringency buffer (2X SSC, 0.1% SDS at room temperature) followed by high stringency buffer (0.1X SSC, 0.1% SDS at 65 $^{\circ}$ C).

#### **2.2.6(b) RT-PCR of *ALDHs* 1 and 2, and *FALDHs* 3 and 4**

To determine presence of mRNA for *ALDH* 1, 2 and *FALDH* 3, 4 genes, RT-PCR reactions were performed using total RNA isolated from cells grown in glycerol

and C<sub>12</sub> alkane (dodecane). As determined by northern blot analysis glycerol represented those substrates where the genes were least induced whereas dodecane represented those substrates where the genes were strongly induced. Isolation of mRNA was carried out using the mRNA capture kit (Roche Diagnostics), following the manufacturer's protocol. A two-step RT-PCR method was used using the Access RT-PCR system (Promega). In this method both the RT and the PCR steps are carried out in the same tube one after the other as described in the manufacturer's protocol.

#### **2.2.6(c) Preparation of radiolabelled probes for northern blot analyses**

PCR amplified DNA sequences of the four *ALDHs*, *FADH* and *β-actin* genes were radiolabelled with <sup>32</sup>P-dATP using High Prime DNA Labelling kit (Roche). These radiolabelled gene sequences were used to probe the RNA blots.

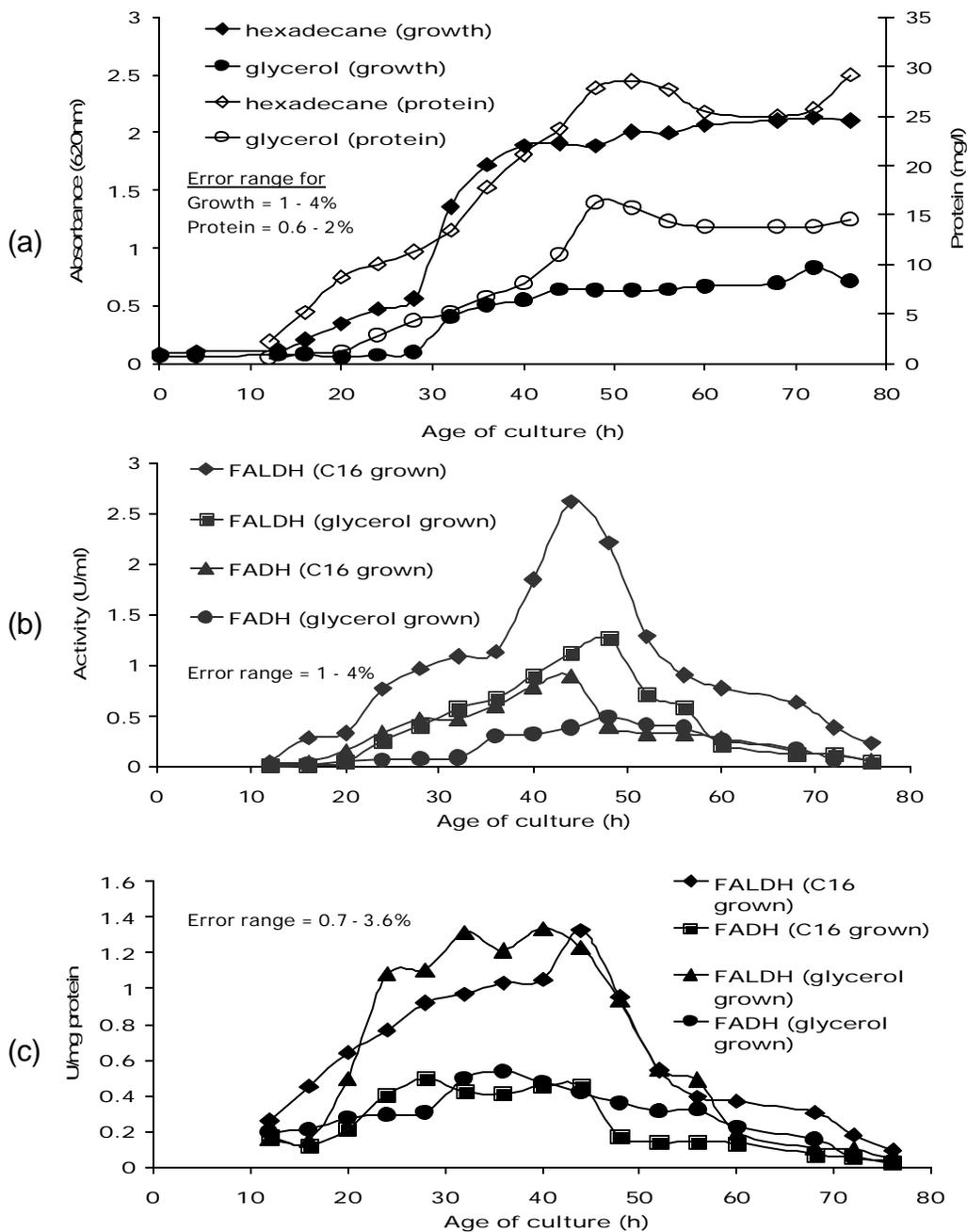
## 2.3 RESULTS AND DISCUSSION

### 2.3.1 FAOD, FADH and FALDH enzyme activities in *Y. lipolytica*

#### 2.3.1.1 FAOD, FADH and FALDH activities in hexadecane and glycerol grown cells

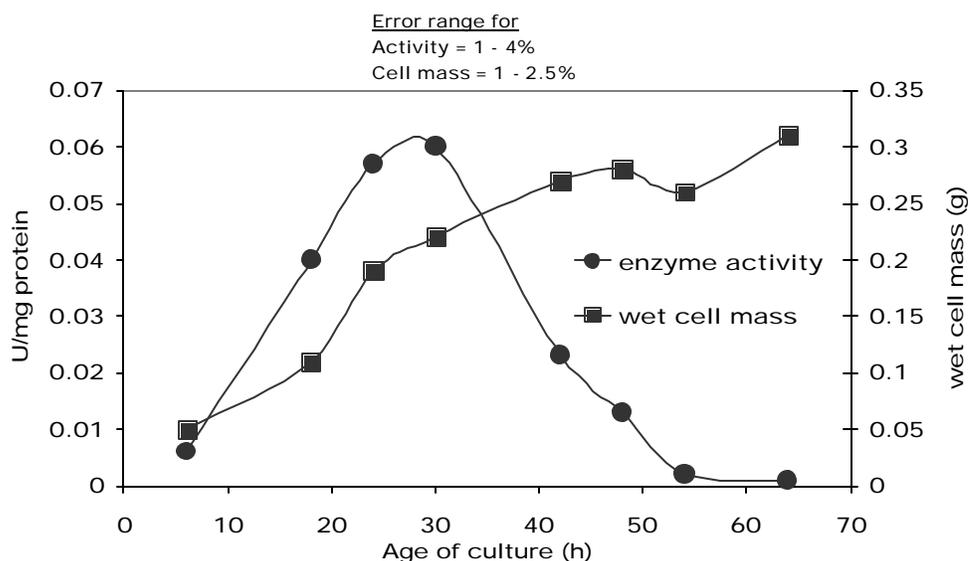
Long chain alcohol and aldehyde dehydrogenase (FADH and FALDH respectively) activities were monitored in cells growing on hexadecane and glycerol using cell-free enzyme extracts. The C<sub>10</sub> aldehyde (1-decanal) and C<sub>12</sub> alcohol (1-dodecanol) were used as substrates for determining the FALDH and FADH enzyme activities, respectively. The dehydrogenase activities were determined by monitoring NADH formation spectrophotometrically at 340nm. The crude cell extracts obtained from *Y. lipolytica* cells grown in alkane were also assayed for fatty alcohol oxidase (FAOD) activity using 1-dodecanol as substrate in the ABTS assay.

FALDH activity was higher than FADH activity in extracts obtained from both glycerol and hexadecane grown cells (figure 2.1). Although growth and protein production were higher in hexadecane grown cells, specific FALDH and FADH activities were very similar for hexadecane and glycerol grown cells. No FAOD activity was detected in glycerol grown cells. Very low FAOD activity was detected in hexadecane grown cells. The known thermolability and photosensitivity of this enzyme (I'ichenko *et al.*, 1994; Hommel *et al.*, 1994), might have contributed to FAOD activity being underestimated.



**Figure 2.1: Induction of FALDH and FADH activities in *Y. lipolytica* H222 cells during batch culture. (a) Production of long chain alcohol (FADH) and aldehyde (FALDH) dehydrogenase in hexadecane (C<sub>16</sub>) and glycerol was followed over time by taking OD<sub>620nm</sub> readings and determining protein concentration of the cells. FADH and FALDH activities were assayed using 1-dodecanol and 1-decanal as substrates respectively. (b) FADH and FALDH activity is expressed in units per volume of enzyme extract. (c) FADH and FALDH activity is expressed as specific activity.**

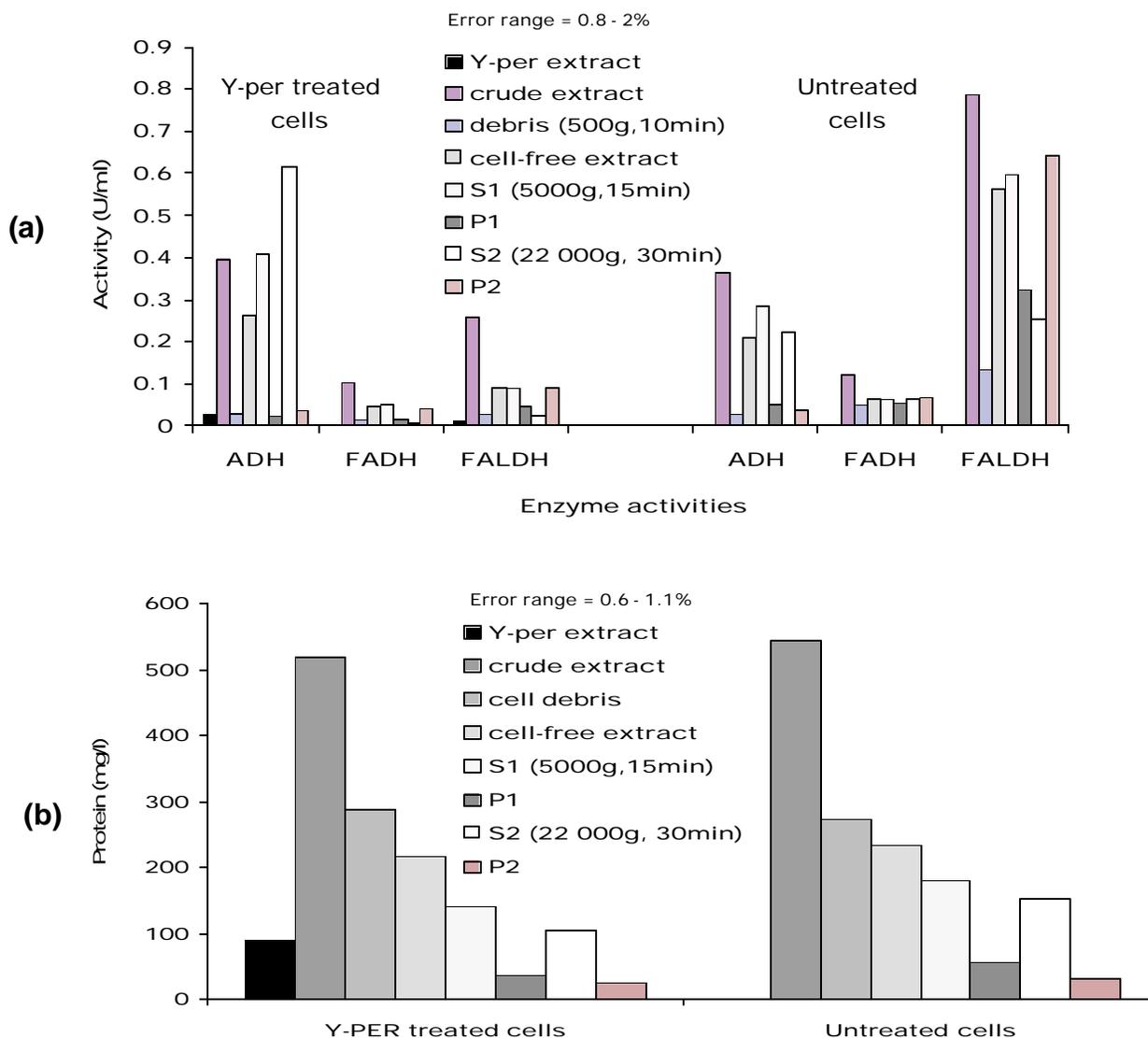
However, compared to FALDH activity both FADH and FAOD activities were very low in the extracts of alkane grown cells (figures 2.1(b) and 2.2), and this left a question as to how *Y. lipolytica* is able to oxidize the long chain alcohols to the corresponding aldehydes.



**Figure 2.2:** A comparison of FAOD induction in relation to wet cell mass in *Y. lipolytica* H222 cells growing in hexadecane. The increases in wet cell mass and specific activity of FAOD were monitored over time for cells grown in hexadecane (C<sub>16</sub>). FAOD activity was assayed using 1-dodecanol as substrate.

### 2.3.1.2 Subcellular fractionation of FADH and FALDH activity

The enzymes FADH, FALDH and FAOD are known to be membrane-bound and associated with peroxisomes, mitochondria and endoplasmic reticulum (Yamada *et al.*, 1980; Mauersberger *et al.*, 1987; Ueda and Tanaka, 1990). Subjecting crude enzyme extracts obtained after cell disruption to several centrifugation steps at different speeds, afforded subcellular fractionation. According to literature it is expected that at a centrifugation speed of 5000 x g for 10 min heavy particles (e.g. nucleus) will be collected and at 22 000 x g for 30 min peroxisomes and mitochondria will be collected while microsomal and cytosolic fractions will remain in solution (Yamada *et al.*, 1980; Ueda and Tanaka, 1990).



**Figure 2.3: Subcellular fractionation of FADH, FALDH and FAOD of *Y. lipolytica* H222 cells growing in hexadecane. (a) Long chain alcohol and aldehyde dehydrogenase and short chain alcohol dehydrogenase activities (FADH, FALDH and ADH respectively) of subcellular fractions of *Y. lipolytica* H222 cells grown on hexadecane were determined for cells treated with Y-PER™ protein extraction reagent before cell disruption as well as extracts from untreated cells. All fractions were normalized to original volume of the crude extract with buffer prior to enzyme activity determinations. FADH, FALDH and ADH activities were assayed using 1-dodecanol, 1-decanal and ethanol as substrates respectively. (b) The protein concentration in all fractions from Y-per treated cells and untreated.**

When using the Y-PER™ protein extraction reagent (Pierce Chemical Co., US) prior to cell disruption the reagent is supposed to extract all soluble cytoplasmic fractions thereby separating them from the non-soluble, membrane-bound enzyme fractions, which will include the microsomal fractions. In an attempt to investigate the subcellular localization of the FALDH and FADH activity in *Y. lipolytica*, enzyme extracts from Y-PER treated and untreated cells were subjected to the different centrifugation steps (figure 2.3).

Since yeast ADH (the short chain alcohol dehydrogenase) is a soluble cytoplasmic enzyme most ADH activity was expected in the Y-PER extract, but this was not the case. Very little protein was also present in the Y-PER extracts, indicating that Y-PER treatment did not have the expected effect in *Y. lipolytica*. This is probably due to the reagent having been designed for *Saccharomyces* and related yeasts but not non-conventional yeasts such as *Y. lipolytica*, which may have a different cell wall structure. FALDH activity was lower (or lost) in Y-PER treated cells, possibly due to interference of the Y-PER reagent with membranes. However, the Y-PER™ protein extraction reagent helped to soften the yeast cells and shorten the time normally required to break the cells. In the case of untreated extracts most FALDH activity was eventually present in the P2 pellet, which should contain peroxisomes and mitochondria. As the amount of protein in solution was reduced at each centrifugation step FALDH activity also increased, which suggests that some proteins initially present in solution and interfering with the assay were removed at each centrifugation step. The relative increase in FALDH activity as enzyme fractions were separated, might thus indicate interference of other proteins with the NAD(P) based assay.

### **2.3.1.3 NAD and NADP dependence of FADH and FALDH activity**

In previous experiments the presence of FALDH enzyme activity in *Y. lipolytica* was confirmed by monitoring reduction of NAD<sup>+</sup>. According to literature the enzyme uses both NAD<sup>+</sup> and NADP<sup>+</sup> (Cobessi *et al.*, 1999, Zhang *et al.*, 2001). The following experiment was aimed at determining exactly which cofactor, NAD<sup>+</sup> or NADP<sup>+</sup>, is

preferred and if one can be substituted for the other. In this case activities were determined for both glucose and hexadecane grown cells. Hexadecane-grown cells had more FALDH activity than glucose-grown cells (table 2.2), indicating that the alkane induces this enzyme. We also found that FALDH uses both NAD<sup>+</sup> and NADP<sup>+</sup> as cofactor. As was expected enzyme activity was higher when using NADP<sup>+</sup> than NAD<sup>+</sup>. The same effect was reported for the FALDH activity of *Vibrio harveyi* (Zhang *et al.*, 2001; Ahvazi *et al.*, 2000). Apparently this is common in all class 3 ALDHs (Perozich *et al.*, 2001), of which FALDH is a member. This behaviour is a result of the larger space in the NAD(P)<sup>+</sup> binding pocket as well as structural shift that accompanies NADP<sup>+</sup> binding which are only seen in class 3 ALDHs (Perozich *et al.*, 2001; Ahvazi *et al.*, 2000). FADH results were not conclusive and are therefore not shown.

**Table 2.2: NAD<sup>+</sup> and NADP<sup>+</sup> dependence of FALDH activities of *Y. lipolytica* H222 cells grown in hexadecane and glucose. FALDH activities were assayed using 1-dodecanal as a substrate.**

Growth media	FALDH activity (U/mg protein)	
	1-dodecanal+ NAD <sup>+</sup>	1-dodecanal + NADP <sup>+</sup>
Hexadecane	1.90 ± 0.71	4.58 ± 0.53
Glucose	1.35 ± 0.36	1.65 ± 0.18

### 2.3.2. Identification of FAOD, FALDH and FADH genes in *Y. lipolytica*.

#### 2.3.2.1. BLAST searches for putative FAOD, FADH and FALDH encoding genes

Shortly after completion of phase I of the Genolevures sequencing project (Genolevures consortium, personal communication, <http://cbi.labri.u-bordeaux.fr/Genolevures/index.php>, described by Artiguenave *et al.*, 2000; Sherman *et al.*, 2004; Dujon *et al.*, 2004) the *Y. lipolytica* database was searched for putative FAOD, FALDH and FADH encoding genes. The deduced amino acid sequence from the *SFA1* gene from *S. cerevisiae* which encodes a long chain alcohol dehydrogenase (Wehner *et al.*, 1993; accession no. X68020) was used to search for FADH encoding genes, the deduced amino acid sequence from the FAOD encoding gene from *C.*

*tropicalis* (Vanhanen *et al.*, 2000; accession no. AJ242496) was used to search for FAOD encoding genes and the amino acid sequence of the human FALDH (Accession no. P51648 or U46689) was used to search for FALDH encoding genes.

No genes coding for proteins with significant homology to the *C. tropicalis* FAOD could be detected. The search was repeated after completion of the *Y. lipolytica* genome sequencing project (Gènolevures consortium, personal communication, <http://cbi.labri.u-bordeaux.fr/Genolevures/index.php>; Dujon *et al.*, 2004), but still without success. Detection of FAOD enzyme activity in *Y. lipolytica* cells in the absence of any candidate genes may thus suggest one of the two possibilities; presence of a novel enzyme or the ability of a related enzyme to catalyze the same reaction, both of which would require further and extensive work to verify. The ability of the P450 monooxygenase system to solely carry out complete oxidation of *n*-alkanes to the corresponding fatty acids has previously been suggested by Scheller *et al* (1998).

### **One putative FADH encoding gene**

BLAST search using the *S. cerevisiae* *SFA1* protein sequence as query against the *Y. lipolytica* gene sequence database yielded one gene fragment coding for a putative FADH. The partial *FADH* DNA sequence was translated into an amino acid sequence. The protein sequence translation gave two frames with interrupted open reading frames. The open reading frames were combined into one continuous amino acid sequence, which was then aligned against the amino acid sequence of *S. cerevisiae*, *SFA1* as shown in figure 2.4. Close similarity (69% amino acid identity) of this protein with *S. cerevisiae* SFA is observable which confirmed this to be indeed a putative long chain ADH protein sequence.

SFA	1	MSAATVGKPIKICIAAVAYDAKKPLSVEEITVDAPKAHEVRIKIEYTAVCHTDAYTLGSDP	61
FADH	1	EVRVKVYTYTGVCHTAAYTLGSDP	24
SFA	62	EGLFPCVLCHEGAG--IVESVGDDVITVKPGDHVIALYTAECGKCKFCTSGKTNLCGAVRA	120
FADH	25	EGNLPWFVHEGAGYRIVESIGEGVTSVKPGDSVLLLYTAACKECKFKSGKSTLCQKVRA	85
SFA	121	TQKGVMMPDGTTRFHNAGKEDLYHFMGCSFSEYTVVADVSVAIDPKAPLDAACLIGCGV	181
FADH	86	TQKGVLPAQTTR---SKGKALS SYMGCSFSQYTVVADVSLVAVDPSAPQDRTCLLGCGV	143
SFA	182	TTGFCAALKTANVQKGDIVAVFGCGTVGLSVTQAKLRGASKIIAIDINNKKKQYCSQFGA	242
FADH	144	TTGYAAATVTANVQKGDNVAVFGAGCVGLAVVMGAKARGAAKIIVIDIHWQ-----QGGL	198
SFA	243	TDFVNPKEDLAKDQTIIVEKLIEMTDGGLDFTFDCTGNTKIMRDALEACHKQWGQSIIIGVA	303
FADH	199	WGFF-----S-----FGCHLTLFNPHQSYPEGHPLICLTSLVLEI	233
SFA	304	AAGEEISTRPFQLVTGRVWKGSAFSGGIKGRSEMGLIKDYQKGALKVEEFI THRRPFKEIN	364
FADH	234	YFTEVCDHFPEVL	246

**Figure 2.4: Comparison of the SFA protein sequence with the partial FADH amino acid sequence obtained through BLAST search of the *Y. lipolytica* genome. The partial FADH amino acid sequence obtained through BLAST search of the *Y. lipolytica* genome was aligned against the SFA1 protein sequence of *S. cerevisiae*, using the DNAssist program. Identical regions are coloured in pink whereas similar regions are green-coloured.**

Figure 2.5 shows the PCR product obtained from amplification of the partial *FADH* gene sequence from *Y. lipolytica* H222 genomic DNA. This PCR product was radiolabelled with  $^{32}\text{P}$ -dATP using High Prime DNA Labelling kit (Roche), and used for northern blot analysis of FADH expression in *Y. lipolytica* (see section 2.3.3.2). As the *Y. lipolytica* sequencing project progressed the full-length FADH gene was obtained and the protein sequence is now available at GenBank with accession number gi|49651085|emb|CAG78022.1|. This FADH protein sequence was used in BLASTP searches of the NCBI database to find other similar proteins in yeasts. The BLASTP search showed that the *Y. lipolytica* FADH was a member of the zinc-binding dehydrogenases belonging to the FAD/NAD(P)-binding Rossmann fold superfamily. Alignment of the *Y. lipolytica* FADH sequence against some members of this group is shown in figure 2.6.

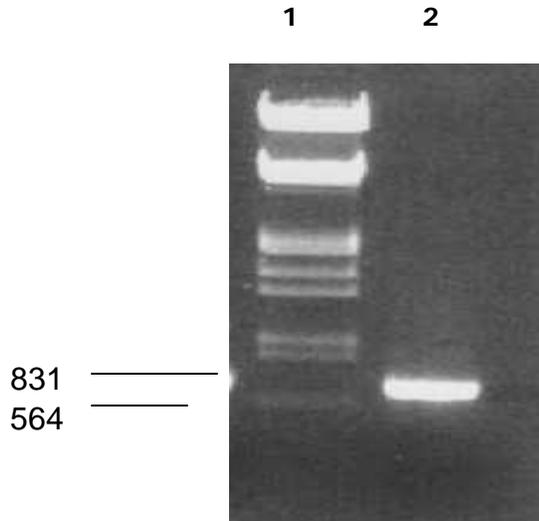


Figure 2.5: The PCR product (Lane 2, 746bp) obtained from amplification of the partial FADH gene sequence from *Y. lipolytica* H222 genomic DNA. The  $\lambda$  (*EcoRI/HindIII*) marker (Lane 1) is shown along side for size comparison. The sizes of the  $M_R$  marker fragments appear in appendix B.

```

P-methanolica  --MSTAGKTITCKAAVAWEAGKDL SIETIEVAPPKAHEVRVKIAYTGVCHTDAFTLSGKD  58
C-boidinii     --MSTVGTKITCKAAVAWKPGEDLSIETIEVAPPKAHEVRIKIEYTGVCHTDAYTLSGAD  58
C-maltosa      MSESTVGTKPITCKAAVAWEAAKPLSIEDVTVAPPKRHEVRIKLYDTGVCHTDAYTLSGVD  60
Y-lipolytica   --MSTEGKTITCKAAVAWEAGKPLTLEDVEVAPPQAHEVRVKVYTYTGVCHTDAYTLSGSD  58
S-cerevisiae   MSAATVGTKPIKICAAVAYDAKKPLSVEEITVDAPKAHEVRIKIEYTAVCHTDAYTLSGSD  60
      :* **.*.* *****.. : *::: * .*: *****:* * .*****:**** *

```

```

P-methanolica  AEGEFVPIVGHGAGIVESIGEGVTSVKVGDHVCLYTPECKECKFCCKSGKTNLCGKIRA  118
C-boidinii     AEGEFVPIVGHGAGIVESVGEVGVTSVKVGDVSVICLYTPECKECKFCCKSGKTNLCGKIRA  118
C-maltosa      PEGAFPVILGHGAGIVESIGEGVTNVKVGDHVIALYTPECKECKFCCKSGKTNLCGKIRA  120
Y-lipolytica   PEGIFPSVILGHGAGIVESIGEGVTSVKPGDSVILLYTAECKECKFCCKSGKTNLCQKVRA  118
S-cerevisiae   PEGLFPCVILGHGAGIVESVGDVDTVVKPGDHVIALYTAECGKCKFCTSGKTNLCGAVRA  120
      .** ** : *****:*** ** *: **.*.***** :***

```

Zinc-ADH signature

```

P-methanolica  TQKGKLPDGTSRFT-CKGKLLHYMGCSTFSQYTVLADISVVAVDPKAPMDRTCLLGCG  177
C-boidinii     TQKGKLPDGTSRFT-CKGQPLLHYMGCSTFSQYTVVADISVVAVNPEAPKDRCLLGCG  177
C-maltosa      TQKGVMPTDGTSRFT-CKGKEILHFMGCSTFSQYTVVADISVVAINPKAEFDKACLLGCG  179
Y-lipolytica   TQKGVLPTDGTTRFK-CKGKDLYSYMGCSFSQYTVVADVSLVAVDPSAPQDRCLLGCG  177
S-cerevisiae   TQKGVMPTDGTTRFHNAKGEDIVHFMGCSTFSEYTVVADVSVVAIDPKAPLDAACLLGCG  180
      *****:*****:** .**:: : *****:**:**:**:**:**:**:**:**:**:**

```

```

P-methanolica  ITTGYGAATNTAKISPGDNIAVFGAGCIGLSVIQGAVERKAGKIIVVDINDDKKEWADKF  237
C-boidinii     ITTGYRAATNTVMNEGDNIAVFGVGCIGLSVIQGAVERKAGKIIAIDINDGKTEWAEKF  237
C-maltosa      ITTGYGAATITANVQKGDVAVFGGGIVGLSVIQCAERGAQIILVDISDKKEEWGQKL  239
Y-lipolytica   VTTGYGAATVTANVQKGDVAVFGAGCVGLAVVMGAKERGAAKIIVIDINGNKEAWAYKF  237
S-cerevisiae   VTTGFGAALKTANVQKGDVAVFGCGTVGLSVIQGAKLRGASKIIAIDINNKKQYCSQF  240
      :***: ** *::: **:*::**:*::**:*::**:*::**:*::**:*::**:*::**:*::**

```

NAD-binding site

```

P-methanolica  GATDFVNPTK-LPAGVTIVDKLVEMTDGGCDFTFDCTGNVGMVRNALEACHKGWGESIVI  296
C-boidinii     GATDFVNPTK-LAEGETIVSKLIDMTDGGCDFTFDCTGTQVGMVRNALEACHKGWGESVII  296
C-maltosa      GATAFVNPTK-LPEGTTIVDKLIEMTDGGCDFTFDCTGNVGMVRNALEACHKGWGTSVII  298
Y-lipolytica   GATDFVNPTK-LPEGTTIVDKLVEMTDGGCDFTFDCTGNVTVMRQALEACHKGWGESIII  296
S-cerevisiae   GATDFVNPKEDLAKDQTIVEKLIEMTDGGLDFTFDCTGNTKIMRDALEACHKGWQGSIII  300
      *** **.*. : * . . **.*.*::***** *****.. :**:****** **::*

```

```

P-methanolica  GVAPAGAEISTRPFQLVTGRVWKGCAFGGVKGGRSQMPGLVQDYLQGGKIKVDEFITHRHPL 356
C-boidinii     GVAAAGKEISTRPFQLVTGRVWRGCAFGGVKGRTQMPGIVEDYLSGTLKVDEFITHRHKL 356
C-maltosa      GVAAAGKEISTRPFQLVTGRTWKGAAFGGVKGGRSQLPGIVNNYLDGKLKVEEFITHREPL 358
Y-lipolytica   GVAAAGQEIASTRPFQLVTGRVWKGSAFGGVKGRTQLPEIVKRYKDGSEIDNFITHSKPL 356
S-cerevisiae   GVAAAGEEISTRPFQLVTGRVWKGSAFGGIKGRSEMGLIKDYQKGALKVEEFITHRRPF 360
***.* **.******.*.*.***:***::: ::: * .* :::::**** . :

P-methanolica  EDINDAFHDMHAGDCIRAVITMDE-- 380
C-boidinii     DDINKAFHDMHAGDCIRAVITMEE-- 380
C-maltosa      AAINKAFEEMHAGDCIRAVVDLS--- 381
Y-lipolytica   KDINSAFTDLHKGDCIRTVVDMWQV- 381
S-cerevisiae   KEINQAFEDLHNGDCLRTVLKSDEIK 386
**.* **.* :.* ***:***:

```

**Figure 2.6: Alignment of the *Y. lipolytica* long chain alcohol dehydrogenase (FADH) protein against known class III Zn-dependent alcohol dehydrogenases (also called glutathione-dependent formaldehyde dehydrogenases) of other yeasts. The zinc-ADH signature sequence and the NAD-binding site in these sequences are shown in boxes. Accession numbers for sequences used are as follows: *Yarrowia lipolytica*, CAG78022.1; *Pichia methanolica*, BAD15032.1; *Candida boidinii*, BAC16635.1; *Saccharomyces cerevisiae*, CAA98742.1; *Candida maltosa*, JN0447.**

Similarity of the *Y. lipolytica* sequence to the other members of this group is observable. In addition, presence of the zinc-ADH signature sequence [GHEX<sub>2</sub>GX<sub>5</sub>(G,A)X<sub>2</sub>(I,V,A,C,S)] and the NAD-binding site (GxGxxG or GxTxxG) (shown in boxes) in these sequences shows that all were indeed members of this family. In most species there are more than one zinc-ADH isozymes (e.g. human have at least six while *S. cerevisiae* has 12 according to Pfam and COG databases) and similarly at least four FADH isogenes have been identified in *Y. lipolytica* (Barth and Kunkel, 1979). The long chain alcohol dehydrogenase activity has been confirmed only for SFA<sub>P</sub> of *S. cerevisiae* (Wehner *et al.*, 1993) and human ADH3<sub>P</sub> (Jornvall and Höög, 1995, Duester *et al.*, 1999).

#### **Four putative FALDH encoding genes**

BLASTP and TBLASTN searches of the first releases of the *Y. lipolytica* genome using the human FALDH as query initially delivered four putative FALDH encoding genes. A DNA similarity of 45% between two of these putative FALDHs and 54% between the other two sequences with a similarity of not more than 20% between the two pairs (table 2.3) showed that we had two groups of ALDHs. CLUSTALW alignments of the deduced

amino acid sequences with each other and with other ALDH and FALDH sequences revealed that two of these sequences (labelled FALDH3 and FALDH4) most likely encoded FALDHs while the other two (labelled ALDH1 and ALDH2) probably encoded other ALDHs. These results were later confirmed with northern blot analysis and RT-PCR (see section 2.3.3).

**Table 2.3: A comparison of the four putative ALDHs from phase I of the *Y. lipolytica* genome sequencing project with each other and with other ALDHs. Values were obtained from ClustalW alignments using default settings.**

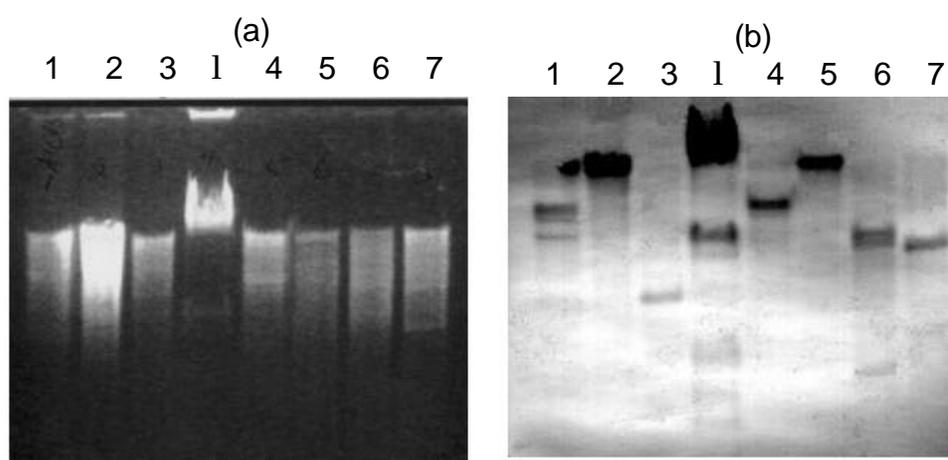
% Similarity								
	DNA sequence				Amino acid sequence			
Sequence	ALDH1	ALDH2	FALDH3	FALDH4	ALDH1	ALDH2	FALDH3	FALDH4
ALDH1								
ALDH2	57				45			
FALDH3	30	3			20	18		
FALDH4	30	3	58		21	19	54	
Human FALDH	n.d.	n.d.	n.d.		24	21	37	36
<i>S. cerevisiae</i> ALDH1	n.d.	n.d.	n.d.		43	50	18	21

\*n.d. - not determined

BLAST searches done after completion of the *Y. lipolytica* genome sequencing project yielded a further two putative FALDH encoding genes, which were labelled *FALDH1* and *FALDH2*. Further BLAST searches of the NCBI database yielded at least 28 putative FALDH sequences from yeasts and molds. A comparative study of all the putative fungal FALDHs is described in chapter 4. The deletion of the four putative FALDHs in *Y. lipolytica* is described in Chapter 3.

### 2.3.2.2. Southern hybridization to detect FAOD genes in *Y. lipolytica*

According to literature FAOD enzyme activity has been detected from *Y. lipolytica* cells growing on alkanes (Kemp *et al.*, 1990; Hommel *et al.*, 1994), and the protein has been purified from this organism (Il'chenko *et al.*, 1994). However all efforts to amplify a FAOD gene from the genome of *Y. lipolytica* H222 using degenerate oligomers based on the FAOD gene sequence of *C. tropicalis* failed. On the contrary, Southern hybridization analysis using the *C. tropicalis* FAOD gene PCR product as a probe suggested the presence of several FAOD encoding genes in *Y. lipolytica* (figure 2.7).



**Figure 2.7: Southern hybridization to probe for FAOD in *Y. lipolytica*. (a) A restriction digest of *Y. lipolytica* H222 genomic DNA. Lanes 1, 2, 3, 4, 5, 6 and 7 contained digests using restriction enzymes *Acc65I*, *BamHI*, *EcoRI*, *Hind III*, *Pst I*, *Sall* and *XbaI*, respectively and the 1 (*HindIII/EcoRI*) marker is in the middle. (b) Southern hybridization of *Y. lipolytica* using FAOD PCR product of *Candida tropicalis* as a probe.**

The enzyme work and our Southern blot analysis results pointed to the presence of an enzyme whose gene(s) could not be located with PCR or BLAST searches of the genome sequence. A lot of work is thus necessary to explain this behaviour and to establish whether the FAOD gene is perhaps only present in some strains of *Y. lipolytica*. Certain conditions (e.g. temperature, growth media, aeration, pH, etc) are known to cause preferential morphological changes among strains of this yeast, however nothing is known about the genes involved in regulation of such changes. Unlike laboratory strains, wild-type isolates are for some unknown reason very prone to

such instability (Barth and Gaillardin, 1996). The strain used for the genome sequencing was CLIB99 (E150) whereas the one used for FAOD activity (literature and this work) was H222, the wild-type strain. This difference may therefore explain why our result appear different from the sequencing results.

### **2.3.3. Expression of FADH and FALDH genes**

#### **2.3.3.1. Northern Hybridization Analysis of FALDH and FADH genes**

PCR products for *ALDH1*, *ALDH2*, *FALDH3*, *FALDH4*, *FADH* and  $\beta$ -actin were purified, radiolabelled with  $^{32}\text{P}$ -dATP and used as probes for northern blot analysis. Two experiments were carried out in which total RNA was isolated from cells that had been grown on glucose, glycerol,  $\text{C}_{12}$  alkane,  $\text{C}_{16}$  alkane,  $\text{C}_{16}$  alkane with pristane,  $\text{C}_{22}$  alkane with pristane and  $\text{C}_{16}$  acid with Tween80. Pristane is often used as co-solvent for insoluble alkanes such as  $\text{C}_{22}$  alkane (McKenna and Kallio, 1971) while Tween80 is used as an emulsifier for organic acids (Wu *et al.*, 2001). RNA extracted from cells incubated with only pristane and Tween80 served as controls. Figure 2.8 shows representative results from one of these experiments. Both experiments gave unsatisfactory results with very weak hybridization and varying levels of  $\beta$ -actin although an effort was made to use the same amounts of RNA. The fact that the  $\beta$ -actin levels were not the same for all substrates made it difficult to use  $\beta$ -actin induction for normalization of the results. However, results obtained in the two experiments were similar, indicating induction of *ALDH1*, *FALDH4*, and *FADH* but no observable induction of *ALDH2* and *FALDH3*.

In figure 2.8 there is an indication that *FALDH4* was transcribed during growth on  $\text{C}_{12}$ ,  $\text{C}_{16}$ ,  $\text{C}_{22}$  alkanes and the  $\text{C}_{16}$  acid, which suggests that this gene is induced by growth on alkanes as suggested by the sequence analysis which showed that it probably codes for a FALDH. *ALDH1* induction levels are low and constant in all substrates. *FADH* appeared to be induced in glucose, glycerol and the  $\text{C}_{12}$  and  $\text{C}_{16}$  alkanes. Thus one may assume that the gene is induced in actively growing yeast cells regardless of the type of substrate.

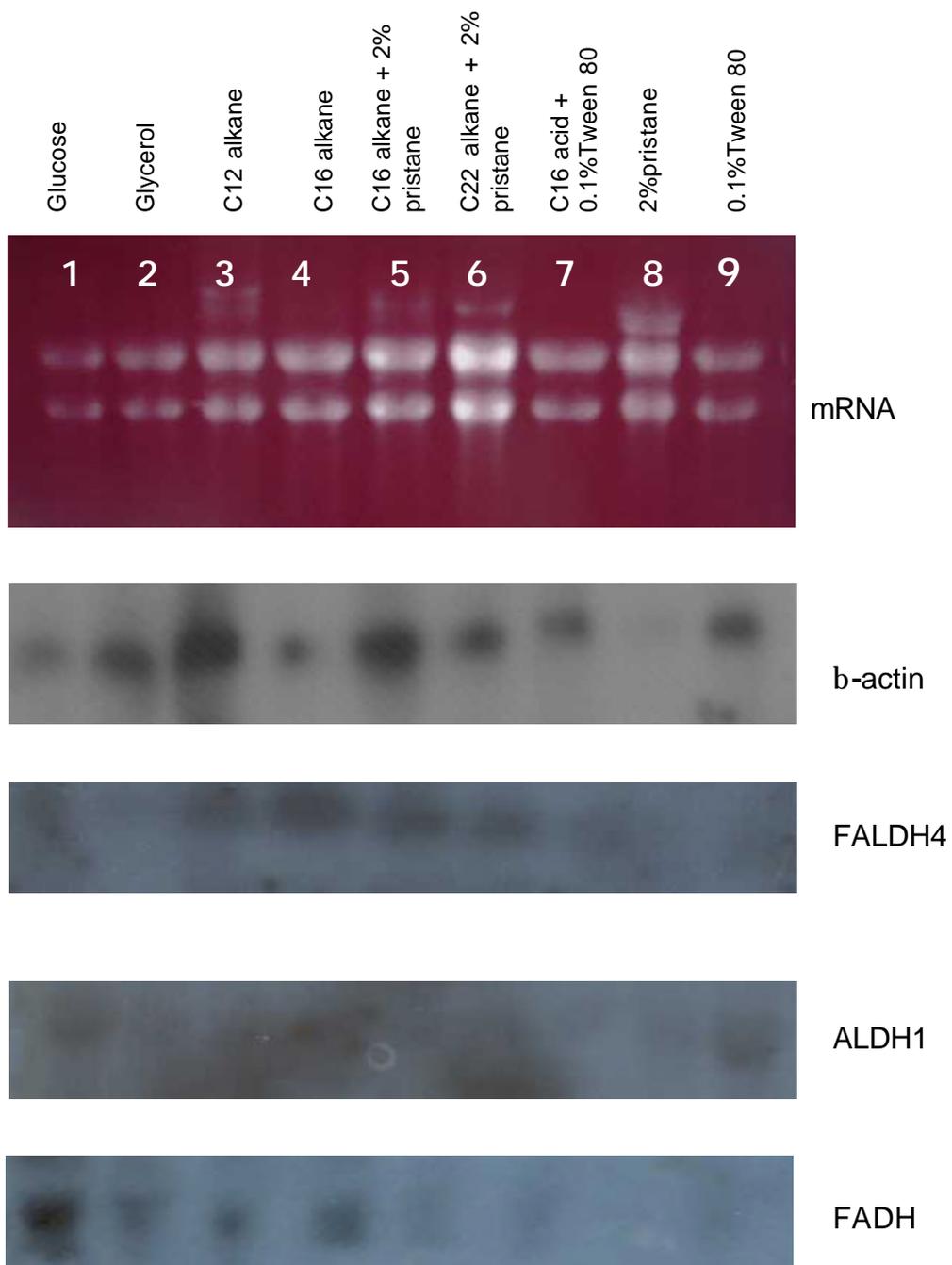
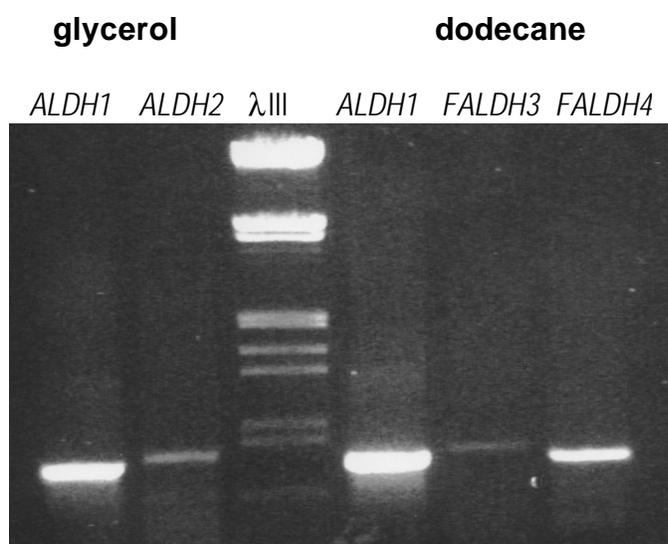


Figure 2.8: Northern blot analysis for induction of *ALDH*, *FALDH* and *FADH* genes. mRNA from *Y. lipolytica* H222 was probed with b-actin, *ALDH1*, *FALDH4* and *FADH* PCR products. Lanes 1, 2, 3, 4, 5, 6 and 8 respectively represent RNA isolated from cells grown in 1% each of the following carbon sources; glucose, glycerol, C<sub>12</sub> alkane, C<sub>16</sub> alkane, C<sub>16</sub> alkane + 2% pristane, C<sub>22</sub> alkane + 2% pristane and C<sub>16</sub> acid + 0.1% Tween80. Lanes 7 and 9 represent glucose-derepressed cells incubated in 2% pristane and 0.1% Tween80 respectively.

### 2.3.3.2. Analysis of ALDH/FALDH expression by RT-PCR

RT-PCR reactions with primers for *ALDH1*, *ALDH2*, *FALDH3* and *FALDH4* were performed using RNA isolated from cells grown in glycerol and C<sub>12</sub> alkane (dodecane) to confirm transcription of these genes as observed with Northern blots. Glycerol as a carbon source is not known to induce any enzymes of the alkane degradation pathway, and was thus included in this study as a control. RNA isolated from cells grown in glycerol yielded PCR products with primers for *ALDH1* and *ALDH2*, while RNA isolated from cells grown in dodecane gave products with primers for *ALDH1*, *FALDH3* and *FALDH4* (figure 2.9).



**Figure 2.9:** Analysis of ALDH/FALDH expression of *Y. lipolytica* cells grown in glycerol and dodecane by RT-PCR. RT-PCR products obtained from mRNA of cells grown in glycerol and dodecane are shown. The  $\lambda$  (*HindIII/EcoRI*) marker is in the middle.

Low levels of PCR products were detected for *ALDH2* and *FALDH3*, as might have been expected from the fact that the northern blot analysis did not show transcription of these two genes. The PCR products for *ALDH1* were more intense than would have been expected from the northern blot analysis while the PCR product for *FALDH4* corresponded with the northern blot analyses. All four RT-PCR products shown in figure 2.9 were sequenced, and all the sequences showed without any doubt that they are indeed products of the corresponding DNA sequences.

## 2.4 CONCLUSIONS

As initially stated, the work presented here was an exploratory study carried out in the early stages of the project to establish the expression of FAOD, FADH and FALDH genes in *Y. lipolytica*. Induction of the enzymes during growth on *n*-alkanes was investigated. Both FADH and FAOD activity were very low, but there is a possibility to have underestimated FAOD levels, since it has been observed that this enzyme is unstable (Kemp *et al.*, 1990). Contrarily, FALDH activity was significantly induced during growth on alkanes. Experiments to establish cellular distribution of the FADH and FALDH enzymes indicated that both enzymes in this yeast exhibit an identical subcellular distribution, i.e. they are enriched in the mitochondrial/peroxisomal fraction as has been described for other alkane-assimilating yeasts (Yamada *et al.*, 1980; Mauersberger *et al.*, 1987).

At the early stages of the *Y. lipolytica* sequencing project BLAST searches of the genome database delivered fragments of four putative *FALDH* gene sequences and one partial *FADH* but no *FAOD* gene sequence. However presence of a *FAOD* gene homologue was indicated through Southern blot analysis. Sequence analysis revealed that there were only two putative *FALDH* sequences (designated *FALDH3* and *FALDH4*) while the other two were short chain ALDH genes (named *ALDH1* and *ALDH2*). Northern blot analysis showed that transcription of *ALDH1*, *FADH* and *FALDH4* was probably induced during growth of *Y. lipolytica* on alkanes. RT-PCR results indicated transcription of *ALDH1* during growth on glycerol and alkane (dodecane), while *FALDH4* was only transcribed during growth on alkanes. It thus appeared that *ALDH1* was constitutively expressed.

At this stage of the project new sequence data from the sequencing project became available in which four full-length FALDH and one full-length FADH gene sequences were identified. The sequences included the two initial FALDHs, *FALDH3* and *FALDH4*, plus two more *FALDH* genes which were named *FALDH1* and *FALDH2*.

## References

1. **Artiguenave F, Wincker P, Brottier P, Duprat S, Jovelin F, Scarpelli C, Verdier J, Vico V, Weissenbach J, Saurin W.** (2000) Genomic exploration of the hemiascomycetous yeasts: 2. Data generation and processing. *FEBS Letters*, 487(1): 13-6.
2. **Ahvazi B, Coulombe R, Delarge M, Vedadi M, Zhang L, Meighen E, Vrielink A.** (2000) Crystal structure of the NADP<sup>+</sup>-dependent aldehyde dehydrogenase from *Vibrio harveyi*: structural implications for cofactor specificity and affinity. *The Biochemical Journal*, 349 (Pt 3):853-61.
3. **Blasig R, Huth J, Franke P, Borneleit P, Schunck W-H, Müller H-G.** (1989) Degradation of long-chain *n*-alkanes by the yeast *Candida maltosa*. III. Effect of solid *n*-alkanes on cellular fatty acid composition. *Applied Microbiology and Biotechnology*, 31: 571-576.
4. **Blasig R, Mauersberger S, Riege P, Schunck W-H, Jockisch W, Franke P, Müller H-G.** (1988) Degradation of long-chain *n*-alkanes by the yeast *Candida maltosa*. II. Oxidation of *n*-alkanes and intermediates using microsomal membrane fractions. *Applied Microbiology and Biotechnology*, 28: 589-597.
5. **Cummings L, Riley L, Black L, Souvorov A, Resenchuk S, Dondoshansky I, Tatusova T.** (2002) Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiology Letters*, 216:133-138.
6. **Delaisse JM, Martin P, Verheyen-Bouvy MF, Nyns EJ.** (1981) Subcellular distribution of enzymes in the yeast *saccharomyopsis lipolytica*, grown on *n*-hexadecane, with special reference to the omega-hydroxylase. *Biochimica et Biophysica Acta*, 676(1): 77-90.
7. **Dickinson FM, Wadforth C.** (1992) Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. *Biochemical Journal*, 282 ( Pt 2):325-31.
8. **Duester G, Farres J, Felder MR, Holmes RS, Höög JO, Pares X, Plapp BV, Yin SJ, Jörnvall H.** (1999) Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochemical Pharmacology*, 58(3): 389-95.

9. **Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, etc.** (2004) Genome evolution in yeasts. *Nature*, 430(6995): 35-44.
10. **Fox MG, Broadway N, Dickinson FM, Ratledge C.** (1993) Carbonyl metabolizing enzymes in alkane-grown microorganisms. *Advances in Experimental Medicine and Biology*, 328:621-30.
11. **Gallo M, Roche B, Aubert L, Azoulay E.** (1973) Distribution of enzymes and cytochromes in *Candida tropicalis* cultured on alkanes. *Biochimie*. 1973;55(2):195-202.
12. **Gallo M, Roche-Penverne B, Azoulay E.**(1974) Localization of mitochondrial alcohol dehydrogenase in *Candida tropicalis*. *FEBS Letters*, 46(1): 78-82.
13. **Haywood GW, Large PJ.** (1981) Microbial oxidation of amines. Distribution, purification and properties of two primary-amine oxidases from the yeast *Candida boidinii* grown on amines as sole nitrogen source. *The Biochemical Journal*, 199(1): 187-201.
14. **Hommel RK, Lassner D, Weiss J, Kleber H-P.** (1994) The inducible microsomal fatty alcohol oxidase from alkane-grown *Candida tropicalis* and *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, 40: 729-734.
15. **Jörnvall H, Höög JO.** (1995) Nomenclature of alcohol dehydrogenases. *Alcohol and Alcoholism*, 30(2): 153-61.
16. **Il'chenko AP, Morgunov IG, Honeck H, Mauersberger S, Vasil'kova NN, Müller H-G** (1994) Purification and properties of alcohol oxidase from the yeast *Yarrowia lipolytica* H-222. *Biochemistry (Moscow)*, 59(9): 969-974.
17. **Kemp GD, Dickinson FM, Ratledge C.** (1988) Light sensitivity of the *n*-alkane-induced fatty alcohol oxidase from *Candida tropicalis*. *Applied Microbiology and Biotechnology*, 29: 370-374.
18. **Kemp GD, Dickinson FM, Ratledge C.** (1990) Inducible long chain alcohol oxidase from alkane-grown *Candida tropicalis* and *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, 32: 461-464.
19. **Kemp GD, Dickinson FM, Ratledge C.** (1991) Activity and substrate specificity of the fatty alcohol oxidase of *Candida tropicalis* in organic solvents. *Applied Microbiology and Biotechnology*, 34: 441-445.

- 20. Kemp GD, Dickinson FM, Ratledge C.** (1994) Occurrence of fatty alcohol oxidase in fatty-acid-utilizing yeasts and molds. *Applied Microbiology and Biotechnology*, 40: 873-875.
- 21. Krauzova VI, Komarova GN, Il'chenko AP, Gulevskaia SA.** (1984) Alcohol oxidation by *Candida guilliermondii* yeasts grown on hexadecanol] *Mikrobiologiya*, 53(4): 621-7.
- 22. Lebeault JM, Meyer F, Roche B, Azoulay E.** (1970 a) Oxidation of higher alcohols in *Candida tropicalis* cultivated on hydrocarbons. *Biochimica et Biophysica Acta*, 220(3): 386-395.
- 23. Lebeault JM, Roche B, Duvnjak Z, Azoulay E.** (1970 b) Alcohol and aldehyde dehydrogenases of *Candida tropicalis* cultivated on hydrocarbons. *Biochimica et Biophysica Acta*, 220(3): 373-85.
- 24. Iida T, Sumita T, Ohta A, Takagi M.** (2000) The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning and characterization of genes encoding for new CYP52 family members. *Yeast*, 16: 1077-1087.
- 25. Mauersberger S, Drechsler H, Oehme G, Müller H-G.** (1992) Substrate specificity and stereoselectivity of fatty alcohol oxidase from the yeast *Candida maltosa*. *Applied Microbiology and Biotechnology*, 37: 66-72.
- 26. Mauersberger S, Kärigel E, Matyashova RN, Müller H-G.** (1987) Subcellular organization of alkane oxidation in the yeast *Candida maltosa*. *Journal of Basic Microbiology*, 27: 565-582.
- 27. Mauersberger S, Ohkuma M, Schunck W-H, Takagi M.** (1996) *Candida maltosa* In: *Non-conventional Yeasts in Biotechnology; A Handbook*. Edited by Wolf K, Springer-Verlag, Germany, pp435-458.
- 28. McKenna EJ, Kallio RE.** (1971) Microbial metabolism of the isoprenoid alkane pristane. *Proceedings of the National Academy of Sciences of the United States of America*, 68(7):1552-4.
- 29. Ohkuma M, Zimmer T, Iida T, Schunck WH, Ohta A, Takagi M.** (1998) Isozyme function of n-alkane-inducible cytochromes P450 in *Candida maltosa* revealed by sequential gene disruption. *The Journal of Biological Chemistry*, 273(7): 3948-53.

30. **Patterton HG, Graves S.** (2000) DNAssist: the integrated editing and analysis of molecular biology sequences in windows. *Bioinformatics*, 16(7): 652-2.
31. **Picataggio S, Deanda K, Mielenz J.** (1991) Determination of *Candida tropicalis* acyl coenzyme A oxidase isozyme function by sequential gene disruption. *Molecular and Cellular Biology*, 11(9): 4333-9.
32. **Sasnauskas K, Jomantiene R, Januska A, Lebediene E, Lebedys J, Janulaitis A.** (1992) Cloning and analysis of a *Candida maltosa* gene which confers resistance to formaldehyde in *Saccharomyces cerevisiae*. *Gene*, 122: 207-11.
33. **Scheller U, Zimmer T, Becher D, Schauer F, Schunck WH.** (1998) Oxygenation cascade in conversion of *n*-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A2. *Journal of Biological Chemistry*, 273(49):32528-34.
34. **Sherman D, Durrens P, Beyne E, Nikolski M, Souciet JL; Genolevures Consortium.** (2004) Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. 1: *Nucleic Acids Research*, 32 Database issue: D315-8.
35. **Sophos NA, Vasiliou V.** (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chemico-Biological Interactions*, 143-144: 5-22.
36. **Ueda M, Tanaka A.** (1990 a) Long-chain alcohol dehydrogenase of *Candida* yeast. *Methods in Enzymology*, 188:171-5.
37. **Ueda M, Tanaka A.** (1990 b) Long-chain aldehyde dehydrogenase of *Candida* yeast. *Methods in Enzymology*, 188:176-8.
38. **Vanhanen S, West M, Kroon JT, Lindner N, Casey J, Cheng Q, Elborough KM, Slabas AR.** (2000) A consensus sequence for long-chain fatty-acid alcohol oxidases from *Candida* identifies a family of genes involved in lipid omega-oxidation in yeast with homologues in plants and bacteria. *Journal of Biological Chemistry*, 275(6): 4445-52.
39. **Verduyn C, van Dijken JP, Scheffers WA** (1984) Colorimetric alcohol assays with alcohol oxidase. *Journal of Microbiological Methods*, 2: 15-25.
40. **Wang HJ, Le Dall MT, Wach Y, Laroche C, Belin JM, Gaillardin C, Nicaud JM.** (1999) Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the *n*-

alkane-assimilating yeast *Yarrowia lipolytica*. *Journal of Bacteriology*, 181(17): 5140-8.

41. **Wehner EP, Rao E, Brendel M.** (1993) Molecular structure and genetic regulation of *SFA*, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Molecular and General Genetics*, 237(3):351-8.
42. **Wu H, Ramachandran C, Weiner ND, Roessler BJ.** (2001) Topical transport of hydrophilic compounds using water-in-oil nanoemulsions. *International Journal of Pharmaceutics*, 220(1-2):63-75.
43. **Yamada T, Nawa H, Kawamoto S, Tanaka A, Fukui S.** (1980) Subcellular localization of long-chain alcohol dehydrogenase and aldehyde dehydrogenase in n-alkane-grown *Candida tropicalis*. *Archives of Microbiology*, 128(2): 145-51.
44. **Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E.** (2001) Differences in nucleotide specificity and catalytic mechanism between *Vibrio harveyi* aldehyde dehydrogenase and other members of the aldehyde dehydrogenase superfamily. *Chemico-Biological Interactions*, 130-132(1-3): 29-38.

## CHAPTER 3

### Gene disruption and Expression analysis of fatty aldehyde dehydrogenase isozymes in *Yarrowia lipolytica*

#### 3.1 Introduction

Existence of multiple gene families in *Y. lipolytica* and other alkane-degrading yeasts is a common phenomenon. For example, several authors have identified multiple gene families such as the POX genes in *Candida tropicalis* (Picataggio *et al.*, 1991) and *Y. lipolytica* (Wang *et al.*, 1999), and the P450 ALK genes in *Candida maltosa* (Ohkhuma *et al.*, 1998) and *Y. lipolytica* (Iida *et al.*, 2000). Studies of physiological function of the different isozymes encoded by these genes have shown that in most cases the isoforms have overlapping functions in terms of substrate specificity and chain length. In some cases however, some of them are not significantly involved in the assimilation of alkanes whereas some are very important. Deletion of some isoforms in some cases is lethal, rendering the organism incapable of growing on alkanes.

With aldehyde dehydrogenases emerging as crucial enzymes in metabolic and detoxifying systems in a wide range of organisms many ALDH genes have been characterized to date (Kotchoni and Bartels, 2003; Davydov *et al.*, 2004, Demosay *et al.*, 2004). However, very little is known about the fatty alcohol and fatty aldehyde oxidizing enzymes of the alkane-degrading organisms. It is therefore of interest to us to determine the molecular and physiological importance of the four putative FALDH isogenes in *Y. lipolytica*, which we believe will shed light on the role of these isozymes as regards the utilization of *n*-alkanes in this organism and perhaps fungi in general.

Depending on the induction level and substrate specificity, cellular metabolic processes are often affected by a specific ensemble of isoforms, making it difficult to distinguish the individual biological role of each, hence the need for sequential disruption of the isoform genes in order to be able to study the independent metabolic function of the individual isozymes. BLAST searches of the *Y. lipolytica* genome database delivered

gene sequences of four putative FALDHs (see chapter 4). This study involved disruption of these four fatty aldehyde dehydrogenase (*FALDH*) isogenes in the yeast *Y. lipolytica*, and examination of the phenotypes displayed by mutants containing all possible combinations of isozyme gene disruptions. Disruption of the four *FALDH* isogenes in *Y. lipolytica* was achieved using a method described by Fickers *et al.* (2003). The method uses the commonly used *URA3* selectable marker gene together with the *Cre-loxP* recyclable tools.

## **3.2 MATERIALS AND METHODS**

### **3.2.1 Plasmids, strains and media**

Plasmids, *E. coli* and *Y. lipolytica* strains used in this study are listed in table 3.1 and the oligonucleotides used for PCR amplification appear in table 3.2. *E. coli* Top10 (Invitrogen Life Technologies) used in transformation and amplification of recombinant plasmid DNA was grown at 37<sup>0</sup>C in Luria-Bertani broth. Selection of Ampicillin resistance in *E. coli* was performed on media containing 100µg ml<sup>-1</sup> ampicillin (Roche Diagnostics). The media and techniques used for *Y. lipolytica* are described by Barth and Gaillardin (1996), and those for *E. coli* are described by Sambrook *et al.* (1989). Yeast cells were grown on YPD broth supplemented with auxotrophic requirements. Ura3<sup>+</sup> and Leu2<sup>+</sup> transformants were selected by growing the yeast cells on yeast nitrogen base (YNB) broth containing all amino acids but no uracil (YNB ura-) for Ura3<sup>+</sup> selection and YNB broth containing 0.01% uracil and all amino acids except leucine (YNB leu-) for Leu2<sup>+</sup> selection. The amino acid mixture and uracil were first dissolved in water, filter-sterilized and then added to autoclaved YNB broth cooled to 60<sup>0</sup>C.

### **3.2.2 General Molecular Biology techniques**

Unless otherwise stated all standard molecular techniques used in this study are as described in Sambrook *et al.* (1989). Restriction enzymes, T4 DNA ligase and the meganuclease I-Sce I were purchased from Roche diagnostics (Germany).

**Table 3.1: Organisms, strains and plasmids used in this study**

Strain (host strain)	Plasmid, genotype	Reference
<b><i>E. coli</i> strains</b>		
	<i>F mcrA</i> ?( <i>mrr-hsdRMS-mcrBC</i> ) <i>F80lacZ</i> ?M15	
Top 10	? <i>lacX74 deoR recA1 araD139</i> ?( <i>ara-leu</i> )7697 <i>galK rpsL</i> (Str <sup>R</sup> ) <i>endA1 nupG</i>	Invitrogen
DH5 $\alpha$	?80d <i>lacZ</i> $\Delta$ m15, <i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi-1</i> , <i>hsdR17</i> ( <i>r<sub>k</sub></i> -, <i>m<sub>k</sub></i> +), <i>supE44</i> , <i>relA1</i> , <i>deoR</i> , $\Delta$ ( <i>lacZYA-argF</i> )U169	Promega
JME459 (DH5 $\alpha$ )	Bluescript KS+ ( <i>ColE1 LacZ bla</i> )	Stratagene
JME130 (DH5 $\alpha$ )	<i>URA3</i> in Bluescript KS+	Nicaud <i>et al.</i> unpublished
JME512 (DH5 $\alpha$ )	JMP112 (LPR synthetic oligonucleotide in Bluescript KS+)	Fickers <i>et al</i> 2003
JME507 (DH5 $\alpha$ )	JMP113 (1.2 kb <i>URA3</i> fragment in JMP112, MU cassette)	Fickers <i>et al</i> 2003
JME461 (DH5 $\alpha$ )	pRRQ2 ( <i>Cre ARS68 LEU2</i> in Bluescript KS+)	Richard <i>et al</i> 2001
<b><i>Y. lipolytica</i> strains</b>		
Po1d	<i>MATa ura3-302 leu2-270 xpr2-322</i>	CLIB139
H222	<i>MATa</i> (wild type)	Barth and Gaillardin, 1996
E150	<i>MATb ura3-302 leu2-270 xpr2-322 lip2::hph</i>	Barth and Gaillardin, 1996

PCR amplification was performed on a Perkin Elmer Gene Amp 2400 PCR system (Perkin Elmer Biosystem). The reaction mixture contained in a final volume of 50 $\mu$ l, 5 $\mu$ l 10X PCR buffer (Roche diagnostics), 10mM dNTP mix (containing 10mM each of dATP, dTTP, dGTP and dCTP), 20-50ng of template DNA, 100 $\mu$ M of appropriate primer pairs and one unit of SuperthermoTaq polymerase (Takara). The PCR amplification consisting of 30 cycles was carried out under the following conditions; denaturation at 94 $^{\circ}$ C for 30s, annealing at 60 $^{\circ}$ C for 30s and extension at 72 $^{\circ}$ C for 2 min. PCR fragments were purified using Amersham DNA purification kit. DNA sequencing was performed on an automated DNA sequencer (ABI Prism model 377, Perkin Elmer) using synthetic primers and the dye terminator procedure. The computer programs

*ClustalW* (Thompson *et al.*, 1994; available at the European Bioinformatics Centre: [www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw)) and DNAssist (Patterton and Graves, 2000) were used for sequence analysis.

### 3.2.3 Construction of disruption cassettes

**Construction of the promoter-terminator (PT) cassette:** The PT cassette was obtained in a two-step PCR reaction in which first the promoter (P) and terminator (T) regions of the relevant *FALDH* isogene were separately amplified. Then the combined P and T PCR products were used as template in a second PCR reaction to obtain the full *FALDH* PT PCR product, which was then purified and cloned into pGEM<sup>®</sup>-T Easy vector (Promega). Verification of the right construct was carried out by restriction digest(s) as well as sequence analysis.

**Construction of the final promoter-URA3-terminator (PUT) cassette:** The deletion cassette loxR-URA3-loxP was rescued from plasmid JMP113 by I-Sce I digestion and cloned into pGEM<sup>®</sup>-T PT at corresponding I-Sce I site. Thus the correct plasmid *FALDH*- pGEM<sup>®</sup>-T -PUT carried the full *FALDH* deletion cassette containing the *URA3* marker gene. Linear disruption PUT cassettes for all four *FALDH* isogenes were generated by PCR amplification of the relevant *FALDH*- pGEM<sup>®</sup>-T-PUT using oligonucleotide primer pair *FALDH*-F1/*FALDH*-R2.

**Deletion of *FALDH* genes:** Yeast cells were transformed by the lithium acetate method (Barth and Gaillardin, 1996) using approximately 1µg of purified PUT PCR product of the relevant *FALDH* gene. Ura<sup>+</sup> transformants, which appeared after approximately 36h, were selected on YNB ura<sup>-</sup> plates. Verification of disruption of the relevant *FALDH* gene was done by PCR on genomic DNA of transformants using primer pair *FALDH*-F1M/*Ura3*-R1M. The primer *FALDH*-F1M was designed for each individual *FALDH* isogene, from a region further upstream (~300bp) from the promoter part of the relevant gene while the common *Ura3*-R1M primer was from within the *URA3* marker gene (~250bp). Thus, a PCR product of the correct size using this primer pair could only be obtained for each individual *FALDH* isogene if the PUT cassette was integrated at the

correct locus and in the right orientation, i.e. if the relevant *FALDH* gene had successfully been deleted.

**Table 3.2: Oligonucleotide sequence of primers used in this study**

Primer	Sequence (5 — 3')
FALDH1-FIM	CGG GTT CGA AGC CGT CAA AAA CC
FALDH1-F1	CGT GGA ATC GGC TGG GAA TGA AAG
FALDH1-R1	C <u>ATT ACC CTG TTA TCC CTA</u> ATA TTT AGA GTT CGG GAT AAA G
FALDH1-F2	<u>CTA GGG ATA ACA GGG TAA TGA</u> ATC GTC TTT CAA ACA A
FALDH1-R2	CGC GTG TGT GGT AATTGA AAT TGG GC
FALDH2-FIM	GCT GCG CCT ACG GAT ACT TCT TC
FALDH2-F1	CTG GGA AGT GCG ACC ATA GAT TGC
FALDH2-R1	C <u>ATT ACC CTG TTA TCC CTA</u> TGT TGT CAG AGG AAT TAC
FALDH2-F2	<u>CTAG GGA TAA CAG GGT AAT</u> GCT TTT TGC CTA AAG TAT A
FALDH2-R2	GCA CTT GGT GAC AAG CAC ACA GT
FALDH3-FIM	CTT TGA CGT CAC AGC ACA TTG TCT GTA TT
FALDH3-F1	TTA CCC TGC CTT GAA CGT CGG TGA AAC
FALDH3-R1	C <u>ATT ACC CTG TTA TCC CTA</u> GGA CAA GTG TCA ATG TTG
FALDH3-F2	<u>CTAG GGA TAA CAG GGT AAT</u> TGG AGT TTA TGG GTG AGT AAT
FALDH3-R2	CCC TTC AAC TAT GAT CAA CCG CTG AAG AGA
FALDH4-FIM	CTA CGT ACA CGT ACA TGC TGT AGA TCT GA
FALDH4-F1	GAT GAC ATG ATG AAG TTG GCC CCT TCA AAT
FALDH4-R1	<u>CAT TAC CCT GTT ATC CCT A</u> TT TTA TTG GTG GTG TGT TT
FALDH4-F2	<u>CTAG GGA TAA CAG GGT AAT</u> ACT AAC CCT ACT TCC TCA TA
FALDH4-R2	ATC TTA GTG GTT GTG GCA ACA GTG TAC GTA
URA3-R1M	GTA GGT GAA GTC GTC AAT GAT GTC GAT AT

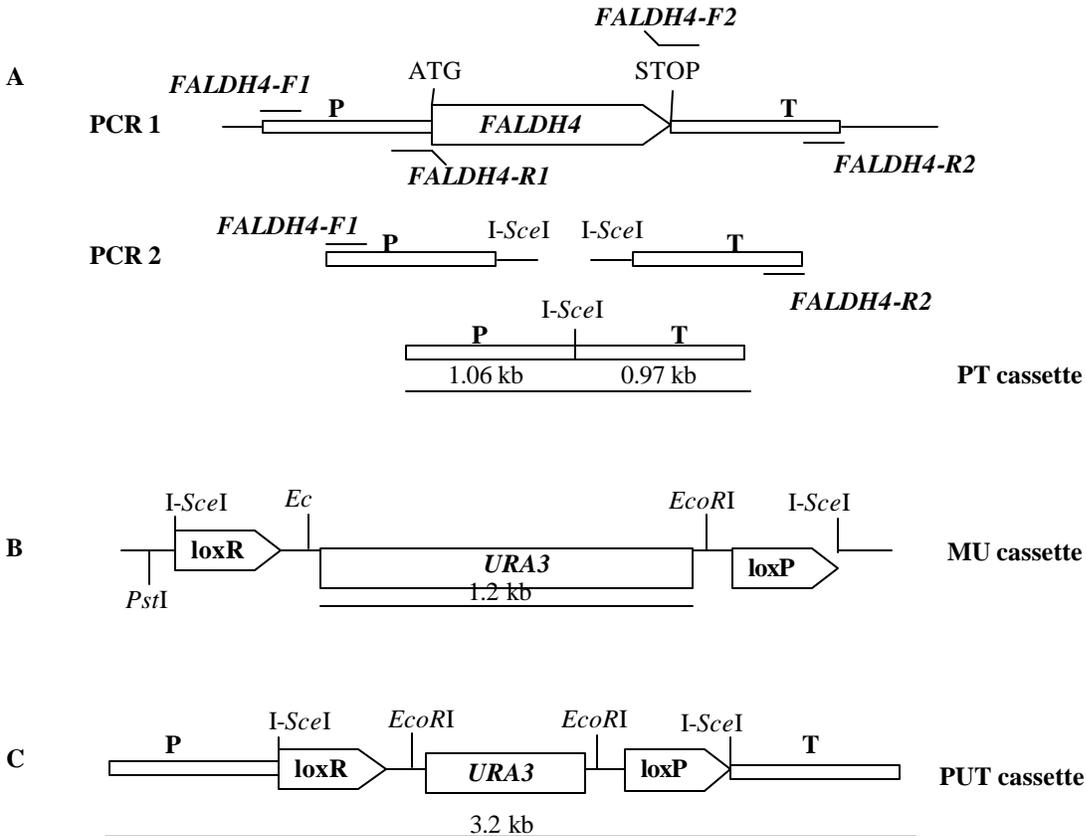
\*Underlined segments represent restriction sites used for cloning

### 3.2.4 *URA3* marker rescue by expression of Cre recombinase

To allow excision of the selectable *URA3* marker gene between the two *loxP* sites, the *FALDH::URA3* strains were transformed with the *cre*-expressing plasmid, *pRRQ2* and selected on YNB leu<sup>-</sup> plates. Loss of the marker followed by loss of the *cre* plasmid was

achieved by growing the transformants in non-selective YPD broth supplemented with 0.5% uracil and 1% leucine in two successive 24h cultures.

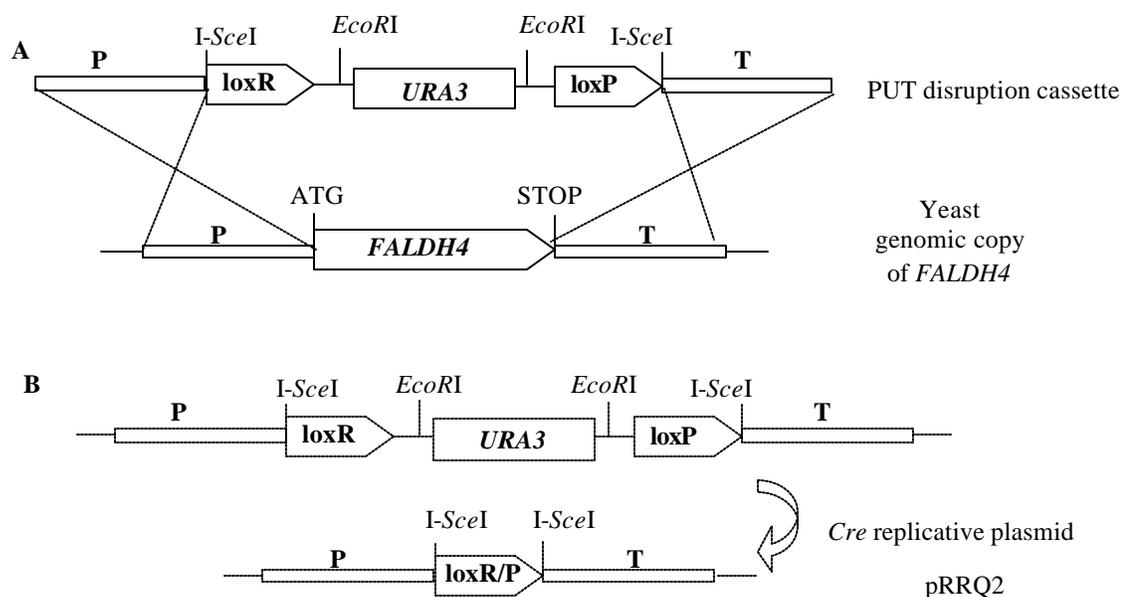
### 3.2.5 Deletion Strategy



**Figure 3.1: Diagrammatic representation of the strategy for construction of gene replacement cassettes for deletion of the four *FALDH* isogenes in *Y. lipolytica* using *FALDH4* gene deletion as an example. (A) The promoter and terminator regions of the *FALDH* genes were amplified in separate PCR reactions incorporating an I-SceI site in the resulting amplicon. The obtained amplicons were used as template in a second PCR reaction resulting in the promoter-terminator (PT) cassette. After cloning of this product into vector pGEM<sup>®</sup>-T Easy the resulting plasmid was linearized using I-Sce-I followed by ligation with the LoxR-URA3-LoxP fragment (B) which was excised using the same restriction enzyme. The resulting construct (C) contained the complete deletion cassette.**

Figure 3.1 is a diagrammatic representation of the strategy used in this study for construction of the gene replacement cassettes for deletion of the four *FALDH* isogenes

in *Y. lipolytica*. The oligonucleotide sequences of primers used in PCR reactions appear in table 3.2. After successful deletion of each *FALDH* gene the *URA3* marker was rescued by transformation with the *cre* plasmid as previously described. The marker rescue process is diagrammatically shown in figure 3.2.



**Figure 3.2:** Diagram showing (A) integration of the deletion cassette into the yeast genome by homologous recombination, resulting in deletion of the *FALDH* gene. (B) The marker gene is rescued by transforming the deletion mutants with the *Cre* plasmid, which recognizes the two loxP sites and cleaves off the intervening *URA3* DNA sequence.

### 3.2.6 Growth analysis of *FALDH* disruption mutants

Growth of *FALDH* deletion mutants on solid media was investigated by growing the strains first on YPD plates. Single colonies were picked from YPD plates and suspended in YPD broth. Serial dilutions of the suspension,  $10^{-3}$ ,  $10^{-6}$ ,  $10^{-9}$  and  $10^{-12}$  were made. Aliquots (2ul) from each dilution were plated on YPD plates and on YNB agar plates containing 0.1% yeast extract and 2% hexadecane. The plates were incubated for 48hrs at 30°C.

Growth of FALDH deletion mutants in liquid media was investigated by monitoring growth in YNB broth containing 0.1% yeast extract and 2% carbon source each of decane (C<sub>10</sub>), dodecane (C<sub>12</sub>), tetradecane (C<sub>14</sub>) and octadecane (C<sub>18</sub>). The turbidity of 200ul samples in duplicate and appropriately diluted with physiological saline solution (FSO; 0.9% NaCl) before transfer to a microtitre plate, were measured at 620nm using a Labsystems iEMS reader MF (Thermo BioAnalysis Co., Helsinki, Finland). Samples were withdrawn from the cultures growing on different carbon sources at 3-hour intervals. The amount of protein in cell-free extracts was determined using the BCA assay kit (Pierce Chemical Co., US) following the manufacturer's protocol.

### **3.2.7 Dry weights determination**

During growth on alkanes duplicate 10ml samples were withdrawn from the culture at 6hr interval. The samples were filtered under vacuum through dry pre-weighed glass microfibre filters (GF/F, Whatman®). The biomass on the filters was washed twice with FSO. The filters with biomass were then dried in an oven at 120<sup>0</sup>C for 24hrs. Mass of dried cells was obtained by subtracting mass of dry filter paper from the total mass. The experiment was repeated twice.

### **3.2.8 FALDH enzyme activity of disruption mutants**

Cell-free enzyme extracts were prepared from 20hr old cells growing in YNB broth containing hexadecane (2%v/v) as described in section 2.2.2(a) of chapter two. FALDH enzyme assays were performed as described in chapter two, section 2.2.3(a) using acetaldehyde, n-decyl aldehyde, and lauraldehyde (all from Fluka) as substrates.

### 3.3 RESULTS AND DISCUSSION

#### 3.3.1 Construction of the *FALDH-PUT* cassettes

Primers for construction of the four FALDH disruption cassettes were based on gene sequences obtained from Dr. J.M Nicaud. The Accession numbers of the four FALDH genes which were temporarily designated FALDH1 to FALDH4 are now available at the GenBank database under the following accession numbers XM\_500179 (FALDH1), XM\_503981 (FALDH2), XM\_500380 (FALDH3), and XM\_505802 (FALDH4). Construction of the four FALDH disruption cassettes was completed as shown in figures 3.3 and 3.4.

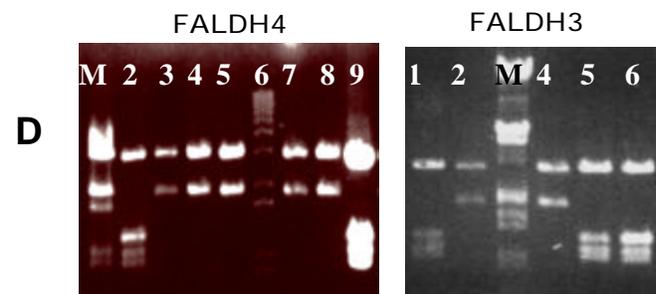
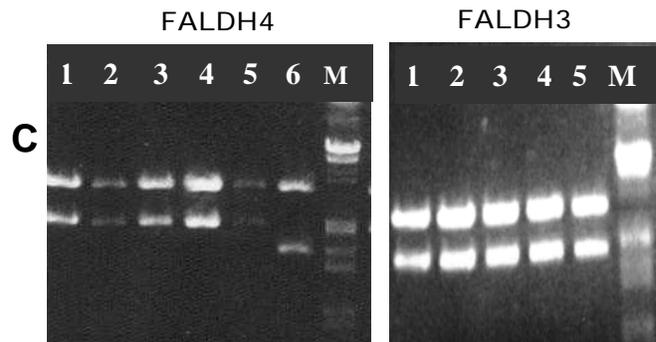
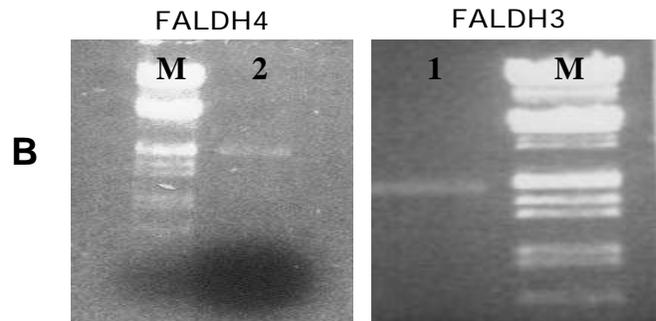
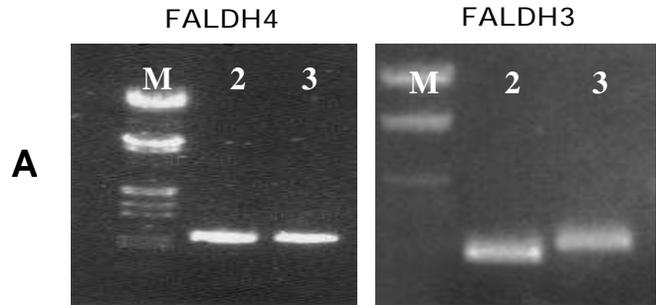
Figure 3.3: Ethidium bromide stained gels depicting the construction of *FALDH3* and 4 deletion cassettes. Lanes indicated by M indicates the molecular weight marker used (λ DNA digested with *EcoRI* and *BamHI*).

(A) Products of the first PCR reaction in which the promoter (lanes 2) and terminator (lanes 3) regions of the *FALDH3* (987 and 1026 bp respectively) and *FALDH4* (880 and 967 bp respectively) were separately amplified.

(B) Products of the second PCR reaction in which the promoter and terminator PCR products were used as template to obtain the combined promoter/terminator (PT) product. In the case of *FALDH3* a product of 2013 bp was expected and in the case of *FALDH4* a product size of 1847 bp.

(C) *EcoRI* digests of the PT PCR products (from B) cloned into pGEM T Easy. Positive ligation was indicated by a single band of approximately 3000 bp representing the pGEM T Easy plasmid and 2013 bp for *FALDH3* (lane 1-5) and 1847 bp for *FALDH4* (lane 1-5) respectively.

(D) The *URA3* marker gene with lox P site at both sides was inserted at the I-SceI site of the PT fragment cloned in pGEM T Easy. Correct clones were identified through digestion with *EcoRI*. Band sizes of 1026, 1200 and 987bp for *FALDH3* (lane 2) and 1200, 967 and 880bp for *FALDH4* (lane 1, 5 and 6) indicated positive clones containing the complete PUT cassette.



Construction of disruption cassettes for FALDH 1 and 2 was carried out as shown for FALDH 3 and 4.

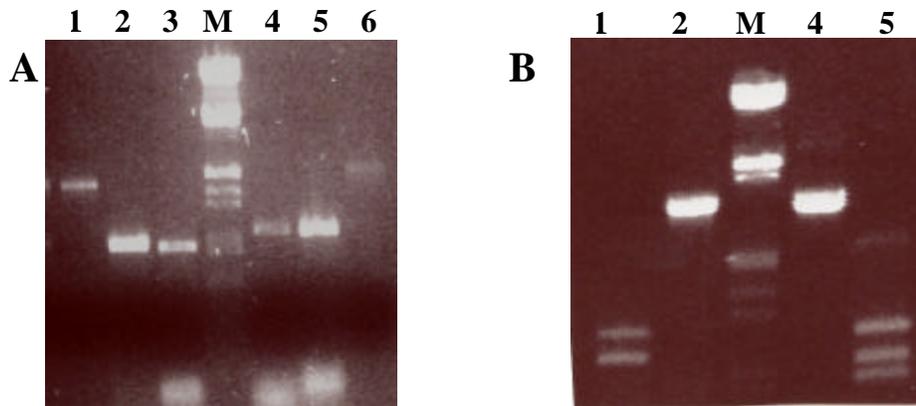


Figure 3.4: Construction of the complete deletion cassette for *FALDH1* and 2.

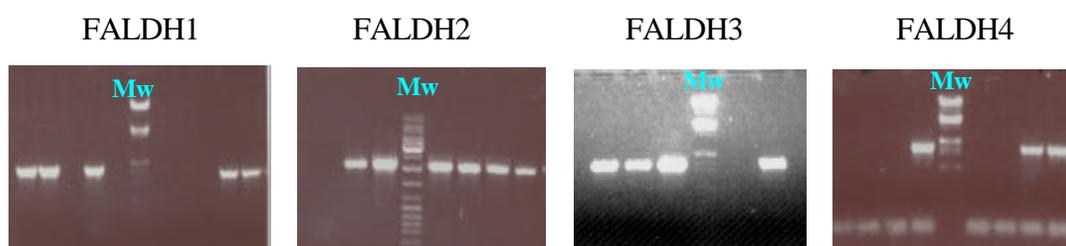
A) PCR products of the promoter region [lane 2, *FALDH1* (expected size 882 bp) and lane 4, *FALDH2* (expected size 1015 bp)] and terminator region [lane 4, *FALDH1* (expected size 989 bp) and lane 5, *FALDH2* (expected size 1008 bp)]. Lane 1 contains the 3071 bp product of *FALDH1* and lane 2 the 3223 bp *FALDH2* PCR products of the second PCR reaction in which the promoter and terminator PCR products were used as template to obtain the combined promoter/terminator (PT) product.

B) After ligation of the PT product in vector pGEM T Easy the *URA3* marker gene with lox P site at both sides was inserted at the I-SceI site of the PT fragment cloned in pGEM T Easy. Correct clones were identified through digestion with *EcoRI*. Band sizes of 1200, 989 and 882bp for *FALDH1* (lane 1) and 1200, 1015 and 1008bp for *FALDH2* (lane 5) indicated positive clones containing the complete PUT cassette.

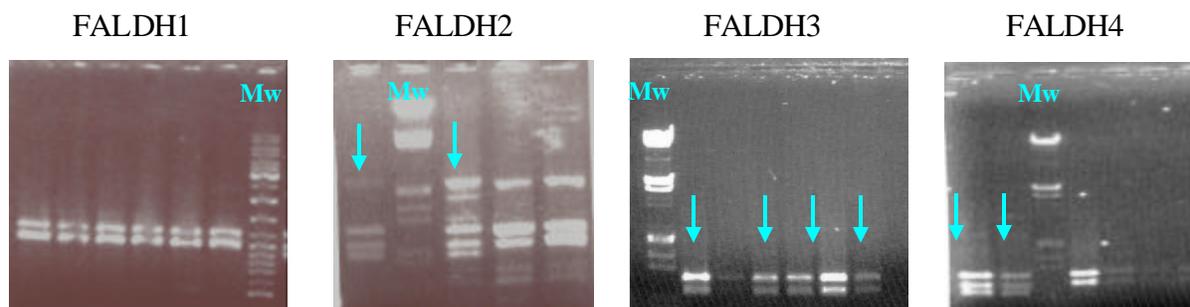
### 3.3.2 Verification of correct disruption of FALDH genes

Positive transformants (i.e. the transformants that were able to grow on ura<sup>-</sup> plates) were isolated and genomic DNA was extracted. Verification of disruption of the relevant FALDH gene was done by PCR on genomic DNA of transformants using primer pair FALDH-F1M/Ura3-R1M. The primer FALDH-F1M was designed to bind to a region upstream (~300bp) from the promoter region of the relevant gene, i.e. to a region that did not form part of the deletion cassette, while Ura3-R1M binds within the

Ura3 marker gene (~250bp). Thus a PCR product could only be obtained if the PUT cassette was integrated at the correct locus and in the right orientation, resulting in a deleted FALDH gene. PCR products obtained for verification of correct disruption of all four FALDHs is shown in figure 3.5. The PCR products were digested with *EcoRI*, and the banding pattern was compared with the linearized original replacement cassette digested with the same enzyme (figure 3.6).



**Figure 3.5: PCR products obtained for verification of correct disruption of FALDHs using primers FALDH- FIM/Ura3-RIM. A PCR product of the correct size indicates integration of the PUT cassette at the correct locus and in the right orientation resulting in a deleted FALDH gene. Expected band size (bp) for FALDH 1, 2, 3 and 4 were 1020, 1490, 1195 and 1200 respectively. Molecular weight marker is indicated by the abbreviation, Mw.**



**Figure 3.6: *EcoRI* digest of PCR products of the deletion cassettes amplified from positive clones. The banding pattern was compared with the linearized original deletion cassette digested with the same enzyme. All clones shown for FALDH 1 are positive and for the other FALDHs arrows indicate positive clones.**

As can be seen in Figure 3.5 not all transformants gave a PCR product, When a PCR was performed on these clones that amplified only the deletion cassette expected size

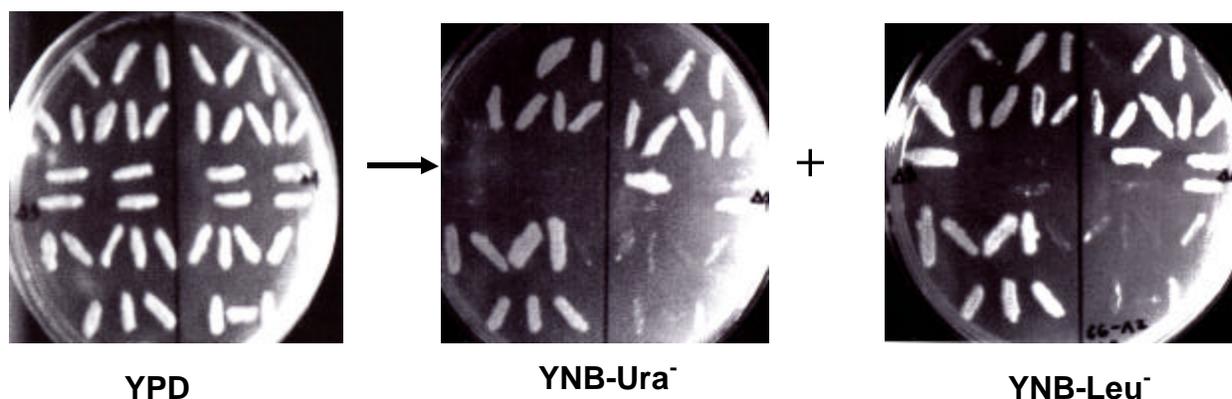
bands were obtained. This indicated that the deletion cassette had been integrated somewhere else in the yeast genome. It has in fact been observed that in most organisms non-homologous integration of introduced DNA is more common than homologous integration (Schiestl *et al.*, 1991). Also we observed that unlike *Saccharomyces cerevisiae*, *Y. lipolytica* has a very poor mechanism of homologous recombination even though it has such high transformation efficiency. We also found that homologous recombination of the deletion cassettes in this yeast was inherently dependent on the method of transformation used and the age of yeast cells. Thus, these two factors together play a critical role when attempting to create deletion mutants in *Y. lipolytica*.

The *Y. lipolytica* parent strain (E150) was used as donor strain for the construction of the deletion cassette and also initially used for deletion of the four *FALDH* genes. As already mentioned, not all the positively selected clones, i.e. selected according to growth on  $ura^-$  media after transformation, had the relevant *FALDH* gene deleted. Even when the method of transformation used was adapted (Barth and Gaillardin, 1996) the age of the yeast cells at the time of transformation was still critical resulting in most clones usually containing the disruption cassette integrated non-specifically. This problem resulted in a great loss of time and money spent on screening for positive clones. Consequently, we resolved to use a different *Y. lipolytica* strain (Po1d) hence all deletions and other experiments thereafter were carried out using this strain. However, it should be noted that as opposed to the E150 strain, Po1d grows significantly slower on alkanes.

### **3.3.3 Ura3 marker rescue by expression of Cre recombinase**

To allow excision of the selectable *URA3* marker gene between the two *loxP* sites, the *FALDH::URA3* strains were transformed with the *cre*-expressing plasmid, *pRRQ2* and selected on YNB  $leu^-$  plates. Cells that were successfully transformed became  $leu^+$  and therefore could grow on YNB  $leu^-$  media. The *Cre* plasmid recognizes the two *loxP* sites on either side of the *Ura3* marker gene of the PUT replacement cassette, and cleaves out the intervening *URA3* DNA sequence. Cells that have thus lost the marker can no

longer grow on YNB  $ura^-$  media. After several generations of growth on non-selective media (e.g. YPD) the yeast cells also lose the *Cre* plasmid and thus can no longer grow on YNB  $leu^-$  plates. Thus a combination of cell phenotypes is seen on YNB  $ura^-$  and YNB  $leu^-$  plates as shown in figure 3.7.



**Figure 3.7: Recovery of the *URA3* and *LEU2* markers.** Following transformation with the *FALDH* deletion cassette, positive selection was performed on YNB  $ura^-$  media. After confirmation of the deletions a second transformation was performed with the *Cre* plasmid, containing *LEU2* marker gene, resulting in  $ura^+$  and  $leu^+$ . To select clones where the *URA3* marker gene was excised through the action of *cre* recombinase, cells were grown in YPD media for a number and single colonies were plated on YPD media followed by replica plating on YNB  $ura^-$  and YNB  $leu^-$  plates.

To study the role of each of the four *FALDH* genes in alkane metabolism four triple deletion mutants were constructed, in which each contained only one intact *FALDH* gene. A quadruple deletion mutant with all four *FALDH* genes deleted, was also constructed. A diagrammatic representation of the sequential disruption of the *FALDH* isogenes for all 15 different deletion combinations created is shown in figure 3.8.

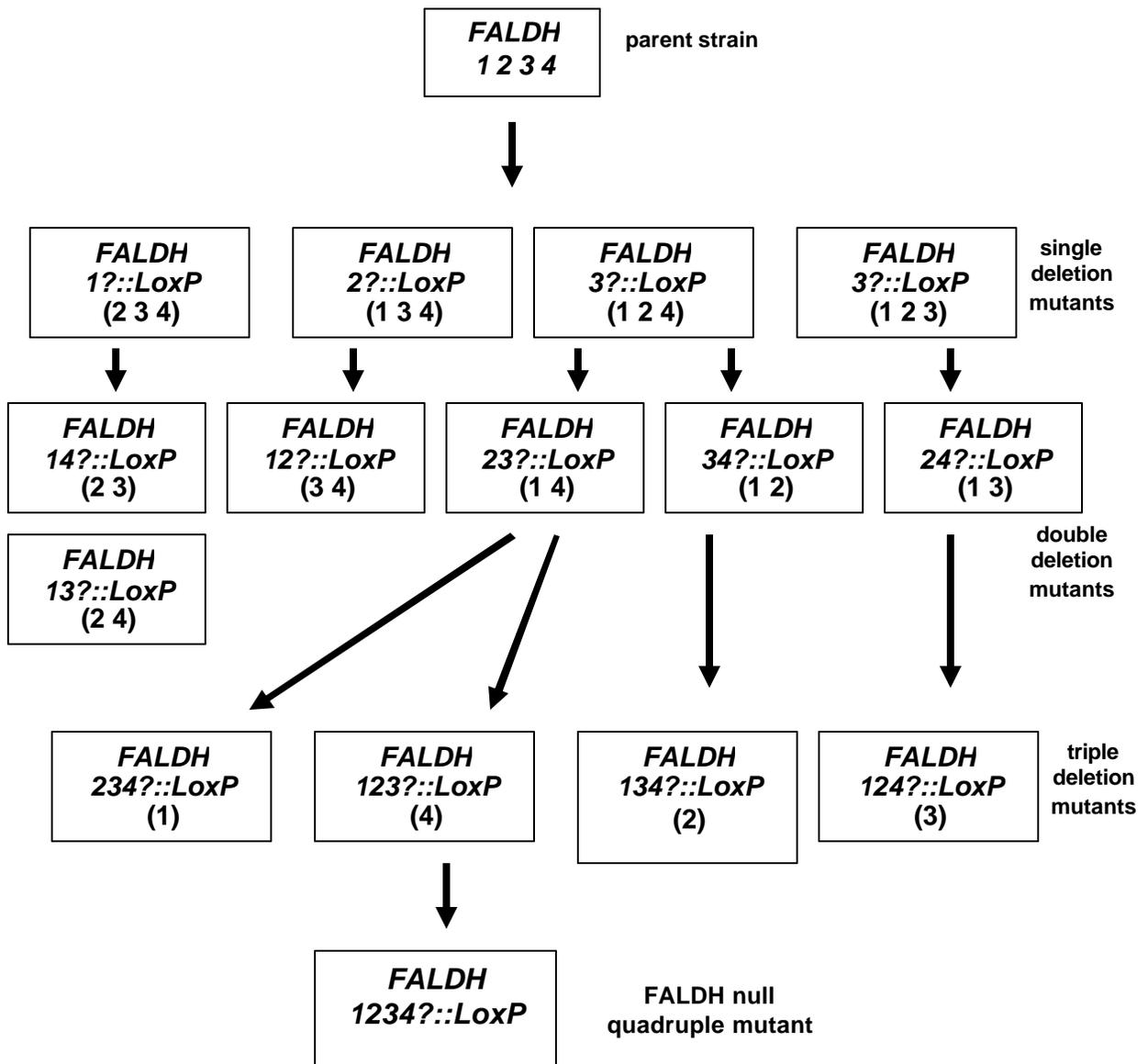
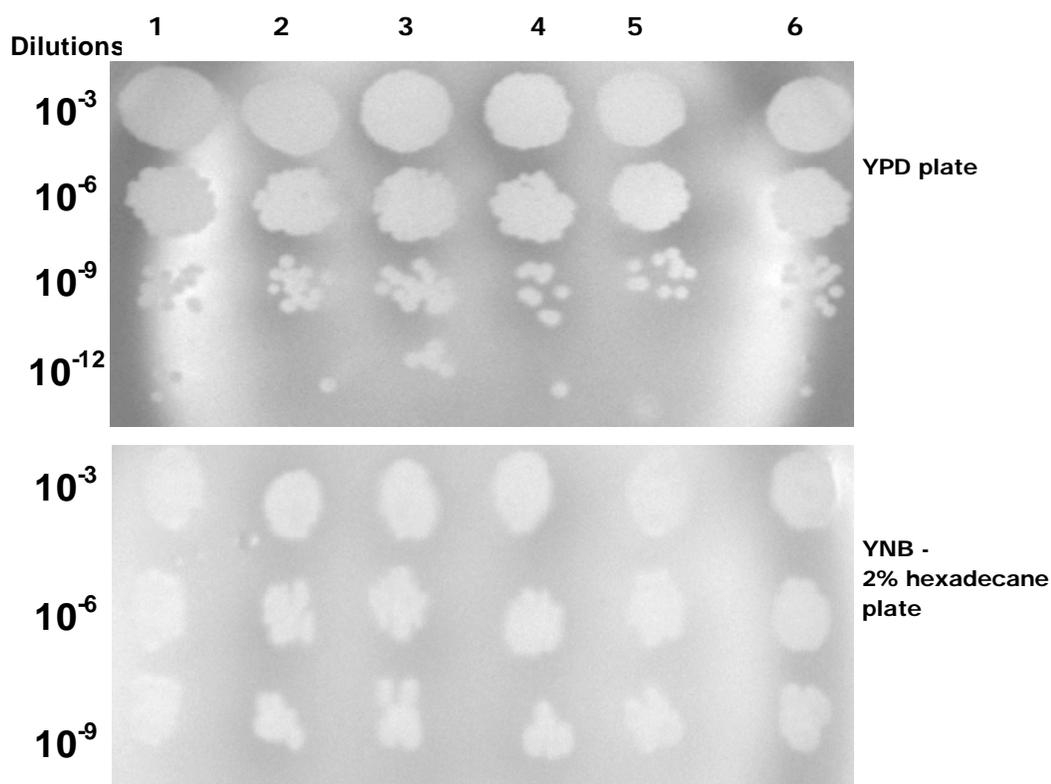


Figure 3.8: A scheme of sequential gene disruption of the *Y. lipolytica* FALDH genes. Numbers in brackets indicate which FALDHs are remaining after the deletion. Arrows indicate the routes followed to create all the 15 different mutant strains.

Growth of the parent strain and the FALDH deletion mutants was compared on solid media (figure 3.9). All strains grew equally well on YPD plates and none seemed to have been negatively affected by the deletions. On YNB agar plates containing 2% hexadecane all strains grew slowly as was usually the case with the parent strain.

However, only one strain, the triple deletion mutant with intact *FALDH2* gene, designated *FALDH134?::loxP* initially seemed to exhibit arrested growth on the alkane plates. In later experiments this was however no longer the case. With four *FALDH* genes deleted one would have expected to see a pronounced change of phenotype from the *FALDH1234?::loxP* strain, but this was not the case. It thus appeared that the *FALDH* deletions had not affected the growth on alkanes on solid media.



**Figure 3.9:** Growth analysis of *Y. lipolytica* *FALDH* triple and quadruple deletion mutants against the wild-type strain. Lane 1 is the wild-type strain whereas lanes 2 to 5 respectively represent triple deletion mutants with *FALDH* 1, 2, 3 and 4 remaining and in lane 6 is the quadruple deletion mutant. Single colonies picked from YPD cultures were grown on YPD broth. Each culture was diluted  $10^{-3}$ ,  $10^{-6}$ ,  $10^{-9}$  and  $10^{-12}$ . Aliquots (5ul) were transferred to a YPD plate and a YNB-hexadecane(2%) plate. The plates were incubated at 30°C for 48h.

### 3.3.4 Growth analysis of mutants

Growth of all triple *FALDH* deletion mutants, the quadruple deletion mutant and the parent strain was investigated in YPD broth. No difference in phenotypes was observed when these strains were grown in YPD broth. All strains grew rapidly and none could be distinguished from the parent strain (figure 3.10 (a)). Thus *FALDH* deletion did not cause a phenotype change as regards utilization of glucose, leading one to believe that *FALDH* genes are not involved in metabolism of this substrate.

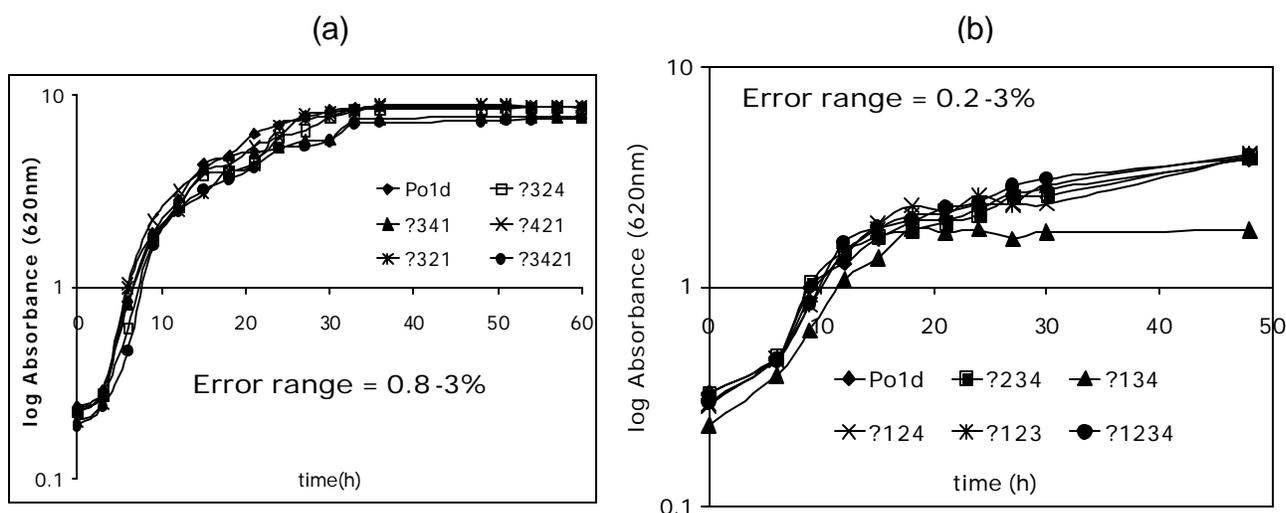
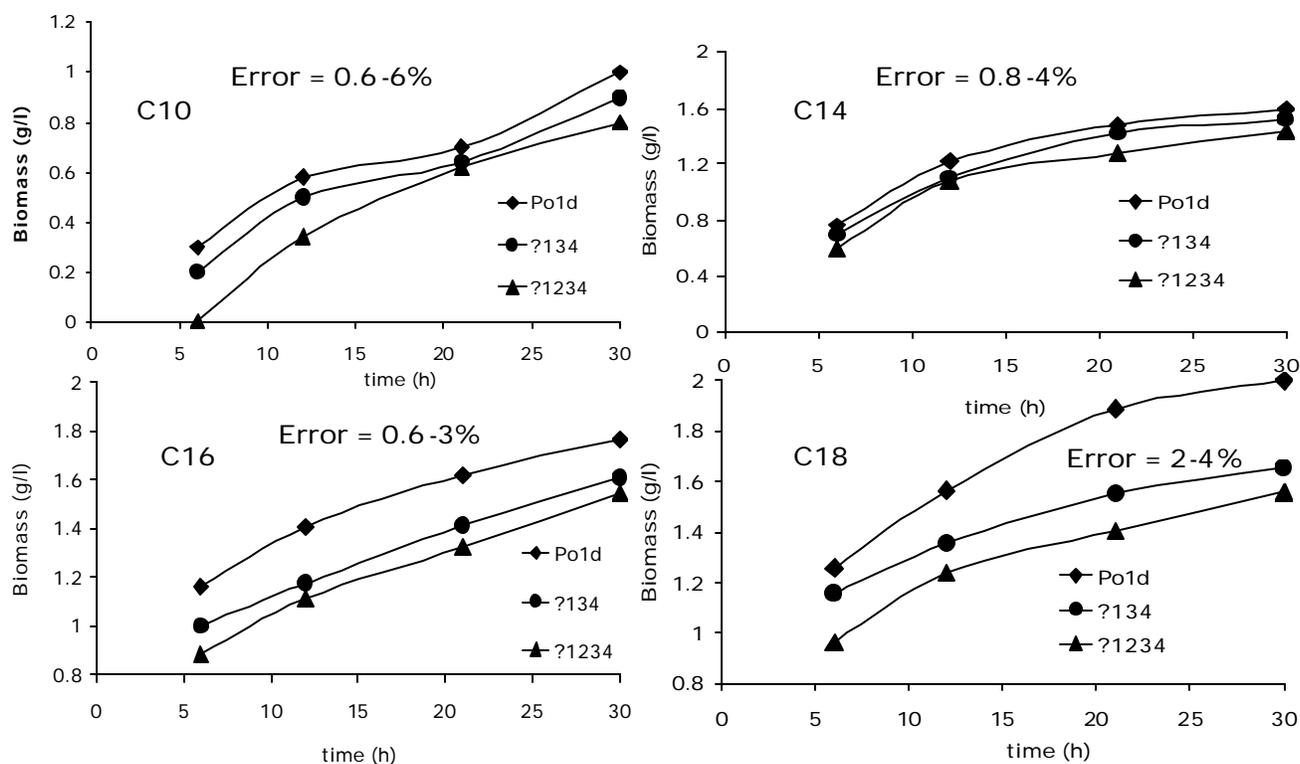


Figure 3.10: A comparison of growth of the parent strain (Po1d) with the *FALDH* triple and quadruple deletion mutants in (a) YPD broth containing 2% glucose and (b) YNB containing 2% octadecane.

Growth of Po1d (parent strain), quadruple and the four triple deletion mutants was also compared in YNB broth supplemented with 2% octadecane ( $C_{18}$ ). In this experiment it again appeared as if only growth of the triple deletion mutant *FALDH134?::loxP* was slightly affected (figure 3.10(b)). Following growth of *Y. lipolytica* on alkanes by OD measurement is often not very reliable. Thus in a subsequent experiment growth of Po1d, *FALDH134?::loxP*, and the quadruple deletion mutant *FALDH1234?::loxP* in

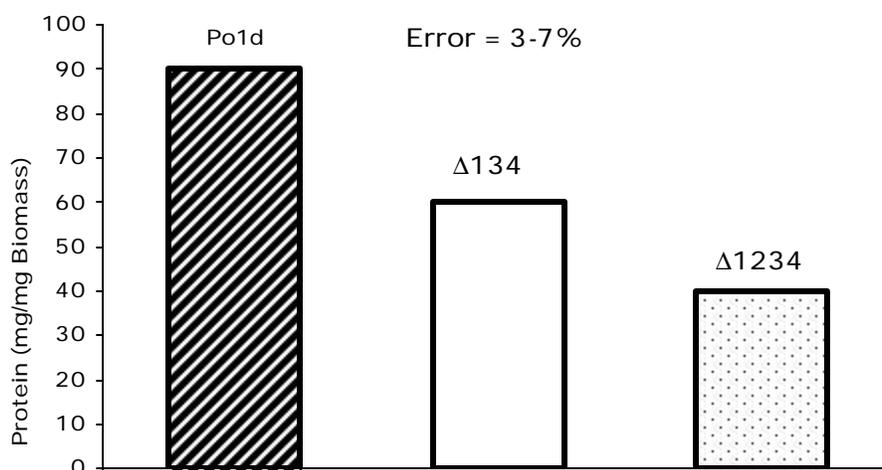
YNB supplemented with decane, tetradecane, hexadecane and octadecane was followed by determining dry weights of filtered and properly washed cells (figure 3.11).



**Figure 3.11: Growth of FALDH deletion mutants *FALDH134?::loxP* (?134) and *FALDH1234?::loxP* (?1234) in YNB broth containing 2% each of decane (C<sub>10</sub>), tetradecane (C<sub>14</sub>), hexadecane (C<sub>16</sub>) and octadecane (C<sub>18</sub>) was compared with the parent strain (Po1d) by monitoring dry cell biomass over time.**

All three strains grew very poorly on decane, and on tetradecane growth of the three strains was very similar. However on hexadecane and particularly octadecane the wild-type strain Po1d produced significantly more biomass than the two deletion mutants. In both cases the quadruple deletion strain produced slightly less biomass than the triple deletion strain. However, OD measurements did not reflect these differences (figure 3.10(b)).

*Y. lipolytica* cells are well known for their hydrophobicity when growing on alkanes, but the *FALDH* deletion mutants showed increased hydrophobicity, with the quadruple *FALDH* deletion mutant cells exhibiting the highest degree of hydrophobicity. Due to this cells from this strain were very difficult to harvest under the normal conditions of centrifuging (3000 x g, 10min) but were successfully harvested at a higher speed for increased length of time (5000 x g, 20 min). In comparison to the mutants, the parent strain exhibited a negligible amount of hydrophobicity, thus enabling easy pelleting of the cells during centrifugation. At this point it is not clear how deletion of the *FALDH* genes has contributed to the increased hydrophobicity, the degree of which seems to be increasing in proportion to the number of *FALDH* genes disrupted in the mutant strains.



**Figure 3.12: (a) The amount of protein in the two *FALDH* deletion mutant strains (Δ134 and Δ1234) in comparison to the parent strain (Po1d).**

A comparison of protein levels from the same amount of wet cell biomass showed that there was less protein in the triple deletion mutant and the least amount of protein in the quadruple deletion mutant (figure 3.12). One may account for this by suggesting that the mutants were accumulating some form of lipid or hydrocarbon based storage material during growth on alkane, which resulted in their increased hydrophobicity.

This probably resulted in poor pelleting, which in turn resulted in the wet biomass containing relatively more water.

Experiments to compare FALDH activity of the wild-type strain (Po1d) with the triple deletion mutant ( $\Delta 134$ ) and the quadruple deletion mutant ( $\Delta 1234$ ) failed, because FALDH activity for all strains was too low to detect significant differences ( $< 0.01\text{U/mg}$  protein). Previous FALDH activity measurements had been carried out on a different strain (H222), and time did not allow optimization of growth and assay conditions for these strains. However, these experiments revealed two other significant differences between the wild-type strain Po1d and the deletion mutants during growth on alkane. These are the increased hydrophobicity, the degree of which seemed to increase in proportion to the number of FALDH genes deleted as well as lower protein content of mutants' cells.

### 3.4 CONCLUSIONS

Experiments described in this report were aimed at understanding the role in alkane assimilation of each of the four *FALDH* genes identified in the genome of the yeast, *Y. lipolytica*. Consequently, 15 FALDH deletion mutants with the four FALDH genes deleted in all possible combinations were constructed. Due to limited time however phenotypic evaluation of all mutant strains was not possible. Results in this regard are therefore from only a few strains and are consequently very preliminary. However it became clear from OD measurements that all four FALDH genes had very little effect on glucose growth and did not totally abolish growth on *n*-alkanes.

The fact that despite loss of the four *FALDH* genes the organism was still able to slowly grow in *n*-alkane media suggests presence of some other mechanism of survival. Presence of other FALDH genes in *Y. lipolytica* is thus strongly suggested. Previously, it appeared that any other aldehyde dehydrogenase genes identified in the genome of this yeast were only short chain ALDHs (Genolevures consortium, personal communication, <http://cbi.labri.u-bordeaux.fr/Genolevures/index.php>), even though due to lack of information these genes have neither been classified with certainty nor shown

to be incapable of fatty aldehyde oxidation. Presently, 28 protein sequences from *Y. lipolytica* are registered in the NCBI database as aldehyde dehydrogenases. Probably more than four of these are FALDHs, some of which may be normally dormant but in an emergency situation as is the case with the described FALDH deletions are activated to participate during alkane assimilation to enable survival of the organism, and this in turn masks the effect of the deletions.

The increased hydrophobicity and lower protein content of the deletion mutants indicate that FALDH deletions have subtle effects that will require in-depth study of all mutants. Further work will have to consider determination of dry weights, protein content and lipid content of filtered biomass. Reliable FALDH enzyme activity levels will also be possible to obtain after optimization of growth and assay conditions. Biotransformation experiments using different alcohols, aldehydes, diols and hydroxyl acids as substrates will also enable thorough assessment of the oxidation capabilities of all mutant strains.

## References

1. **Arakawa H, Lodygin D, Buerstedde JM.** (2001) Mutant loxP vectors for selectable marker recycle and conditional knockouts. *BMC Biotechnology*, 1(1): 7.
2. **Artiguenave F, Wincker P, Brottier P, Duprat S, Jovelin F, Scarpelli C, Verdier J, Vico V, Weissenbach J, Saurin W.** (2000) Genomic exploration of the hemiascomycetous yeasts: 2. Data generation and processing. *FEBS Letters*, 487(1): 13-6.
3. **Barth G, Gaillardin C.** (1996) *Yarrowia lipolytica*. In: K. Wolf, ed. *Non-conventional Yeasts in Biotechnology*. Berlin, Germany: Springer Verlag, pp 313- 388.
4. **Barth G, Gaillardin C.** (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiology Reviews*, 19(4): 219-37.
5. **Casaregola S, Feynerol C, Diez M, Fournier P, Gaillardin C.** (1997) Genomic organization of the yeast *Yarrowia lipolytica*. *Chromosoma*, 106(6): 380-90.
6. **Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA.** (1998) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology*, 9(3): 1316-23.
7. **Delneri D, Tomlin GC, Wixon JL, Hutter A, Sefton M, Louis EJ, Oliver SG.** (2000) Exploring redundancy in the yeast genome: an improved strategy for use of the *cre-loxP* system. *Gene*, 252: 127-135.
8. **Fickers P, Le Dall MT, Gaillardin C, Thornart P, Nicaud J-M.** (2003) New disruption cassettes for gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *Journal of Microbiological Methods*, 55: 727-737.
9. **Gorski JA, Jones KR.** (1999) Efficient bicistronic expression of cre in mammalian cells. *Nucleic Acids Research*, 27(9): 2059-61.
10. **Gueldener U, Heinisch J, Koehler GJ, Voss D, Hegemann JH.** (2002) A second set of *loxP* marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Research*, 30(6): e23.

- 11. Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH.** (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Research*, 24(13): 2519-23.
- 12. Iida T, Sumita T, Ohta A, Takagi M.** (2000) The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, *Yarrowia lipolytica*: cloning and characterization of genes coding for new CYP52 family members. *Yeast*, 16(12): 1077-87.
- 13. Juretzek T, Le Dall M, Mauersberger S, Gaillardin C, Barth G, Nicaud JM.** (2001) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast*, 18(2): 97-113.
- 14. Kujau M, Weber H, Barth G.** (1992) Characterization of mutants of the yeast *Yarrowia lipolytica* defective in acetyl-coenzyme A synthetase. *Yeast*, 8(3): 193-203.
- 15. Madzak C, Treton B, Blanchin-Roland S.** (2000) Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *Journal of Molecular Microbiology and Biotechnology*, 2(2): 207-16.
- 16. Mamoun CB, Beckerich JM, Gaillardin C, Kepes F.** (1999) Disruption of YHC8, a member of the *TSR1* gene family, reveals its direct involvement in yeast protein translocation. *The Journal of Biological Chemistry*, 274(16): 11296-302.
- 17. Mauersberger S, Wang HJ, Gaillardin C, Barth G, Nicaud JM.** (2001) Insertional mutagenesis in the n-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. *Journal of Bacteriology*, 183(17): 5102-9.
- 18. Nagy A.** (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis*, 26(2): 99-109.
- 19. Ohkuma M, Zimmer T, Iida T, Schunck WH, Ohta A, Takagi M.** (1998) Isozyme function of n-alkane-inducible cytochromes P450 in *Candida maltosa* revealed by sequential gene disruption. *The Journal of Biological Chemistry*, 273(7): 3948-53.

20. **Patterton HG, Graves S.** (2000) DNAssist: the integrated editing and analysis of molecular biology sequences in windows. *Bioinformatics*, 16(7): 652-3.
21. **Picataggio S, Deanda K, Mielenz J.** (1991) Determination of *Candida tropicalis* acyl coenzyme A oxidase isozyme function by sequential gene disruption. *Molecular and Cellular Biology*, 11(9): 4333-9.
22. **Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C.** (2001) Tagging morphogenetic genes by insertional mutagenesis in the yeast *Yarrowia lipolytica*. *Journal of Bacteriology*, 183(10): 3098-107.
23. **Sambrook J, Fritsch EF, Maniatis T.** (1989) *Molecular cloning: A laboratory manual*, 2<sup>nd</sup> ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. **Sauer B.** (1987) Functional expression of the *cre-lox* site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 7(6): 2087-96.
25. **Schiestl RH, Petes TD.** (1991) Integration of DNA fragments by illegitimate recombination in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences U S A*, 88(17): 7585-9.
26. **Sherman D, Durrens P, Beyne E, Nikolski M, Souciet JL;** Genolevures Consortium. (2004) Genolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. 1: *Nucleic Acids Research*, 32 Database issue: D315-8.
27. **Thompson JD, Higgins DG, Gibson TJ.** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22): 4673-80 (available at [www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw)).
28. **van Beilen JB, Panke S, Lucchini S, Franchini AG, Rothlisberger M, Witholt B.** (2001) Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the *alk* genes. *Microbiology*, 147(Pt 6): 1621-30.

- 29. Wang H, Le Clainche A, Le Dall MT, Wache Y, Pagot Y, Belin JM, Gaillardin C, Nicaud JM.** (1998) Cloning and characterization of the peroxisomal acyl CoA oxidase ACO3 gene from the alkane-utilizing yeast *Yarrowia lipolytica*. *Yeast*, 14(15): 1373-86.
- 30. Wang HJ, Le Dall MT, Wach Y, Laroche C, Belin JM, Gaillardin C, Nicaud JM.** (1999) Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast *Yarrowia lipolytica*. *Journal of Bacteriology*, 181(17): 5140-8.
- 31. Wiczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP, Boles E.** (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Letters*, 464(3): 123-8.
- 32. Zhao W-N, McAlister-Henn L.** (1996) Expression and gene disruption analysis of the isocitrate dehydrogenase family in yeast. *Biochemistry*, 35: 7873-7878.

## CHAPTER 4

### Putative fatty aldehyde dehydrogenase-encoding genes from the sequenced fungal genomes

PR Matatiele<sup>1</sup>, MS Smit<sup>1</sup>, J Albertyn<sup>1</sup> and J-M Nicaud<sup>2</sup>.

<sup>1</sup> University of the Free State, Department of Microbial, Biochemical and Food Biotechnology, P.O. Box 339, Bloemfontein, 9300, Republic of South Africa

<sup>2</sup> Microbiologie et Génétique Moléculaire, INRA Centre de Grignon, 78850 Thiverval-Grignon, France

#### Abstract

Although fatty aldehyde dehydrogenase (FALDH) activity has been detected in fungi no FALDH genes have been cloned, sequenced and expressed. This work aimed at fishing, through BLAST searches using the human FALDH sequence as query, for fungal FALDH/FALDH-like gene sequences from all the presently available (complete and incomplete) fungal genome sequences. We have identified 25 fungal FALDH/FALDH-like protein sequences from both molds and yeasts. These deduced fungal proteins, which fall into two subfamilies, show approximately 24-90% identity amongst themselves and 26-44% similarity with the human FALDH protein. The proteins are potentially active class 3 ALDHs with long C- and N-termini, and most are probably integral membrane proteins bearing 1-3 transmembrane domains.

#### Introduction

Aldehydes are a class of ubiquitous molecules involved in different physiological processes. Being highly reactive and often toxic molecules, aldehydes can cause lipid peroxidation and modification of proteins and nucleic acids. However, aldehydes do not only represent toxic molecules since they are also found as intermediates in the synthesis of amino acids and osmoprotective osmolytes required for protection against osmotic stress (Kirch *et al.*, 2001; Kotchini *et al.*, 2003). Aldehydes are also found as structural components of lens and corneal crystallins (Cooper *et al.*, 1993; Zinovieva *et al.*, 1993) and they are also found as

pheromone components in many insects (Tasayco and Prestwich, 1990). Also, alkane-utilizing microorganisms are able to use fatty aldehydes resulting from degradation of alkanes (C<sub>6</sub> up to C<sub>40</sub>) as carbon source (Mauersberger *et al.*, 1996; Ueda and Tanaka, 1990).

Toxic aldehydes can be converted to the corresponding less toxic carboxylic acids by the large family of NAD(P)<sup>+</sup>-dependent aldehyde dehydrogenases (ALDHs) [aldehyde: NAD(P)<sup>+</sup> oxidoreductases, EC 1.2.1]. The ALDH group of enzymes consists of diverse subfamilies with four main functions: detoxification, intermediary metabolism, osmotic protection, and NADPH regeneration (Perozich *et al.*, 1999). Various distinct ALDHs have been studied and characterized in detail especially in humans and yeasts (Yoshida *et al.*, 1998; Navarro-Avino *et al.*, 1999).

ALDH enzymes are now broadly grouped into four categories based on substrate specificity namely; semialdehyde dehydrogenases, non-specific ALDHs, ALDH-like proteins and other ALDHs. The non-specific ALDHs comprise classes 1, 2 and 3 ALDHs (Yoshida *et al.*, 1998). Class 1 and 2 ALDHs are cytosolic or mitochondrial tetrameric enzymes involved in detoxification and metabolism of acetaldehyde and other dietary aldehydes, xenobiotics, lipid peroxidation products and certain anti-cancer drugs. Malfunction of these enzymes is associated with susceptibility to ethanol-related diseases. The class 3 ALDHs are cytosolic or microsomal dimeric enzymes associated with oxidation of aromatic aldehydes and fatty aldehydes (medium-chain aliphatic aldehydes) and are also associated with carcinogenesis and severe genetic disorders. Fatty aldehyde dehydrogenase (FALDH) is a member of class 3 ALDHs and mammalian FALDHs as well as several bacterial FALDHs have been extensively studied (Miyauchi *et al.*, 1991; Masaki *et al.* 1994, 1996; Rizzo *et al.*, 2001; Singer *et al.*, 1985; Zhang *et al.*, 2001).

Initially no uniform nomenclature for ALDH genes and enzymes existed resulting in confusion (Navarro-Avino *et al.* 1999). This necessitated establishment of a uniform numbering system. Since 1998 a standardized ALDH nomenclature system has been established based on divergent evolution (Nebert *et al.*, 1991, 2000; Nelson *et al.*, 1996; Vasiliou *et al.*, 1997, 1999). First, the gene and/or protein sequences are divided into families and subfamilies. Protein sequences with amino acid residues that are distinctive and form an individual subset of sequences comprise a family, and the family itself is a member of a larger group of paralogous sequences called an extended or a super-family (Hempel *et al.*, 2003). A superfamily is divided into gene families and subfamilies based on percent identity of each protein as compared with others. Any two ALDH protein sequences exhibiting a = 40% similarity are described as belonging to two different families, whereas sequences that are = 60% identical are considered to belong to the same subfamily (Sophos *et al.*, 2001). Details of the ALDH nomenclature system and names of members of the nomenclature committee are available on the ALDH website (<http://www.uchsc.edu/sp/sp/alcdbase/aldh-nomencl.html>). According to this system when naming an ALDH gene, root symbol "**ALDH**" denoting "aldehyde dehydrogenase" 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. For example, the human fatty aldehyde dehydrogenase gene, formerly known as *ALDH10* or *FALDH* is now denoted as *ALDH3A2*. It is recommended that human *ALDH* variant alleles be given numbers (or number plus a capital letter) following an asterisk (e.g. "*ALDH3A2\*2*, *ALDH2\*4C*").

Among the FALDHs only the human FALDH has thus far been the focus of extensive study due to its involvement in the genetic disorder Sjörger-Larsson syndrome (SLS), an inborn neurologic impairment, characterized by mental retardation and spasticity resulting from a mutation in the human fatty aldehyde dehydrogenase gene (Rizzo *et al.*, 2001). Although it is well-known that fungi, especially the majority of non-conventional yeasts grow very well on fatty

alcohols and aldehydes, very little information is available about genes responsible for metabolism of these substrates. The first three genes involved in fatty alcohol utilization (fatty alcohol oxidases) to be cloned were from *Candida cloacae* and *Candida tropicalis* (Vanhanen *et al.*, 2000) and only recently, three more fatty alcohol oxidase genes have been cloned from *C. tropicalis* (Eirich *et al.*, 2004). However, the majority of yeast ALDH genes sequenced and studied are from *Saccharomyces cerevisiae* [Grzybowska *et al.*, 1993; *Saccharomyces* genome database, (<http://genome-www.stanford.edu/Saccharomyces/>)] and *Schizosaccharomyces pombe* [Wood *et al.*, 2002; (<http://genomebiology.com/2002/3/3/comment/2003>)] whose genomes were the first fungi to be completely sequenced. Most of these ALDH genes have been shown to be involved mainly in ethanol metabolism, with only a few exceptions, which have been implicated in amino acid and 1-pyrroline-5-carboxylate syntheses as well as GABA degradation (<http://www.uchsc.edu/sop/alcdbase/fun-aldh.html>). None have been implicated in FALDH activity. However, enzyme assays using decanal as substrate and monitoring NAD consumption have demonstrated fatty aldehyde dehydrogenase activity in a number of yeasts grown on *n*-alkanes (Yamada *et al.*, 1980; Ueda and Tanaka, 1990; Mauersberger *et al.*, 1996).

There is a drastic increase in whole genome sequence information that is becoming available. More than 50 complete genome sequences are currently publicly available (NCBI's Entrez Genomic site). Summaries of available fungal ALDH/ALDH-like gene sequences were included in the ALDH sequence reviews published by Sophos *et al.*, (2001, 2003). The study of complete genome sequences has revealed that the number of ALDH genes found per organism ranges from 1-5 in archaeal species, 1-26 in eubacteria, and 8-17 genes in eukaryotic species (Sophos *et al.*, 2003). A database of all these genes is available at the ALDH website (<http://www.uchsc.edu/sop/alcdbase>). In this database FALDH genes are described for bacteria and mammals only. The database also gives a neighbour-joining tree of the ALDH superfamily, which

shows the presence of a yeast ALDH gene, *ALDH14* from *S. cerevisiae*, which clusters together with mammalian class 3 ALDHs, of which FALDH is a member. This work aims at fishing for more fungal class 3 or FALDH/FALDH-like gene sequences from all the available (complete and incomplete) fungal genomes.

## **METHODS**

### **BLAST search of fungal genomes for putative FALDHs**

The human FALDH protein sequence (GenBank accession no. gi|1706379|sp|P51648 or U46689) was used as query in BLASTP and TBLASTN searches (Altschul *et al.*, 1997) of fungal genomes for putative FALDH/FALDH-like gene sequences. Both searches were performed using default parameters.

### **Multiple sequence alignment and phylogenetic analysis**

The fungal FALDH/FALDH-like protein sequences were aligned using either DNAssist ver.2.0 (Patterton and Graves, 2000) or the ClustalW program (<http://www.ebi.ac.uk/clustalw/>) available at the European Bioinformatics Institute website (Thompson *et al.*, 1994) and a phylogenetic tree was constructed using the program TreeExplorer (<http://evolgen.biol.metro-u.ac.jp/>).

### **Identification of conserved ALDH regions from BLAST hits**

Analysis tools PROSITE (<http://kr.expasy.org/cgi-bin/prosite-search-ac>) (Gasteiger *et al.*, 2003) and SMART (<http://smart.embl-heidelberg.de/>) (Schultz *et al.*, 1998) were used to search for ALDH conserved regions obtained through the BLAST searches. A protein was identified as an ALDH if it contained at least one of the two ALDH diagnostic amino acid motifs; (i) the ALDH glutamic acid active site signature sequence LELGGKS for mammalian class 3 ALDHs and (ii) the cysteine active site signature sequence (Kirch *et al.*, 2004). In addition, it was expected that the protein must contain the distinguishing feature of FALDH, which is the presence of a hydrophobic domain at the carboxy-terminal, which is made up of about 35 amino acids (Masaki *et al.*, 1994).

## **Domain analysis of identified fungal FALDHs**

Pfam (<http://pfam.wustl.edu/>) (Sonnhammer *et al.*, 1998) and SMART (<http://smart.embl-heidelberg.de/>) (Schultz *et al.*, 1998) domain analysis programs were used to predict the domain architecture of identified FALDHs while PSORT III [<http://psort.ims.u-tokyo.ac.jp/form2.html>] (Horton and Nakai, 1997) was used to predict cell localization as well as for further domain analysis. The TMHMM program (<http://www.cbs.dtu.dk/cgi-bin/nph-webface>) was used to analyze each protein sequence for the presence of transmembrane segments (TMS).

## Results and Discussion

### BLAST search of fungal genomes for putative fungal FALDHs

A summary of the current fungal genome projects available at <http://www.ncbi.nlm.nih.gov/genomes/FUNGI/funtab.html> is given in table 4.1.

**Table 4.1: Fungi genome projects report, adapted from the NCBI website (33 genome projects: 8 complete, 20 WGS assembly, 5 in progress)**  
(<http://www.ncbi.nlm.nih.gov/genomes/FUNGI/funtab.html>).

Organisms				
<i>Ascomycota</i> (28)		<i>Basidiomycota</i> (4)	<i>Microsporidia</i> (1)	
<i>Pezizomycotina</i> (10)	<i>Saccharomycotina</i> (17)	<i>Schizosaccharomycetes</i> (1)		
<i>Aspergillus fumigatus</i>	<i>Candida albicans</i> – WGS	<i>Schizosaccharomyces pombe</i> - complete	<i>Coprinopsis cinerea okayama7#130</i> WGS	<i>Encephalitozoon cuniculi</i> - complete
<i>Aspergillus nidulans</i> FGSC A4 - WGS	<i>Candida glabrata</i> CBS138 – complete		<i>Cryptococcus neoformans var. grubii</i> H99- WGS	
<i>Aspergillus parasiticus</i>	<i>Candida tropicalis</i>		<i>Phanerochaete chrysosporium</i> - WGS	
<i>Aspergillus terreus</i> - WGS	<i>Debaryomyces hansenii</i> CBS767 - complete		<i>Ustilago maydis</i> 521 - WGS	
<i>Coccidioides immitis</i> RS - WGS	<i>Eremothecium gossypii</i> - complete			
<i>Coccidioides posadasii</i> C735	<i>Kluyveromyces lactis</i> NRRL Y-1140 – complete			
<i>Fusarium sporotrichioides</i>	<i>Kluyveromyces waltii</i> NCYC 2644 - WGS			
<i>Gibberella zeae</i> PH-1- WGS	<i>Saccharomyces cerevisiae</i> - complete			
<i>Magnaporthe grisea</i> 70-15- WGS	<i>Saccharomyces bayanus</i> 623-6C - WGS			
<i>Neurospora crassa</i> strain OR74A- WGS	<i>Saccharomyces bayanus</i> MCYC 623 - WGS			
	<i>Naumovia castelli</i> NRRL Y-12630 - WGS			
	<i>Saccharomyces kluyveri</i> NRRL Y-12651- WGS			
	<i>Saccharomyces kudriavzevii</i> IFO 1802- WGS			
	<i>Saccharomyces mikatae</i> IFO 1815- WGS			
	<i>Saccharomyces mikatae</i> IFO 1815- WGS			
	<i>Saccharomyces paradoxus</i> NRRL Y-17217- WGS			
	<i>Yarrowia lipolytica</i> CLIB99- complete			

\*WGS - whole-genome shotgun, unmarked implies sequencing in progress

There are currently (October 13, 2004 update) 33 fungal genome projects, of which 8 are complete, 20 are whole-genome shotgun (WGS) assemblies and 5 are in progress (Table 4.1). The human FALDH protein sequence (ALDH3A2, GenBank accession no. U46689) was used as query in BLASTP and TBLASTN searches of the fungal genomes for putative FALDH/FALDH-like gene sequences.

More than 100 hits were obtained in each BLAST search, however only hits with a score bit of no less than 200 were used, because these corresponded to an amino acid identity of at least 30% (Table 4.2). Twenty-eight putative FALDH/FALDH-like gene sequences were identified from the presently available fungal genomes. These FALDH/FALDH-like protein sequences show approximately 31-43 % identity and 51-61% similarity with the human FALDH gene. Three of these sequences,, though within the required score value, could not be used for further analysis because their available sequences either had incomplete open reading frames or were in contigs.

**Table 4.2: Results of BLAST search of fungal genome databases using human FALDH protein sequence as query. Only results with a score bit of at least 200 are shown.**

Gene sequence No.	Description		TBLASTN			Amino acid Length	% Similarity to other FALDHs (ClustalW scores)		
	Protein sequence No.	Organism	Score	% Identity	% Similarity		Mammalian ALDH3A2	Plants ALDH3's	Mammalian ALDH3A1
XM_404038	EAK87196.1	<i>Ustilago maydis</i> 521	384	43	61	640	40-44	34-40	40-43
AACQ01000035.1	EAL00022.1	<i>Candida albicans</i> SC5314	317	39	58	542	36-37	31-37	36-41
AACQ01000020.1	EAL02020.1	<i>Candida albicans</i> SC5314	312	36	59	661	35-36	31-36	36-41
XM_323331.1	EAA49061.1	<i>Magnaporthe grisea</i> 70-15	309	37	58	523	35-36	30-38	37-38
AACU01000968.1	EAA53613.1	<i>Magnaporthe grisea</i> 70-15	303	37	57	527	35-36	30-34	36-38
CR382139	CAG90160	<i>Debaryomyces hansenii</i> CBS767	300	35	57	578	33-34	30-34	36-38
AABX01000718.1	EAA28392.1	<i>Neurospora crassa</i> OR74A	280	33	51	533	34-35	31-36	34-37
*TIGR222929 contig:1140:c posadasii		<i>Coccidioides posadasii</i> C735	261	38	57				
*TIGR5085 contig:5231:a fumigatus		<i>Aspergillus fumigatus</i>	249	37	54				
AACD1000098.1	EAA62737.1	<i>Aspergillus nidulans</i> FGSC A4	211	41	59	517	34-36	31-38	35-38
AACD01000168.1	EAA64317.1	<i>Aspergillus nidulans</i> FGSC A4	214	31	54	530	32-34	30-35	33-35
AACM01000412.1	EAA67698.1	<i>Gibberella zeae</i> PH-1	230	33	56	532	34-35	31-33	35-39
AACU01000237.1	EAA56915.1	<i>Magnaporthe grisea</i> 70-15	214	33	54	597	33-34	27-34	33-34
NC_001145.2	NP_013828.1	<i>Saccharomyces cerevisiae</i>	248	32	53	532	30-31	27-31	30-34
AADM01000092.1		<i>Kluyveromyces waltii</i> NCYC2644	267	35	54	524	32-34	28-32	30-34
AABZ01000665.1		<i>Saccharomyces mikatae</i> IFO 1815	256	35	54	525	30-32	30-33	31-33
*AACH01001154.1		<i>Saccharomyces mikatae</i> IFO 1815	251	34	55	(525)			
AACA01000400.1		<i>Saccharomyces bayanus</i> MCYC 623	255	34	53	534	32-33	27-33	32-35
AABY01000215.1		<i>Saccharomyces paradoxus</i> NRRL Y-17217	247	33	53	532	30-31	26-31	31-35
AACI01000788.1		<i>Saccharomyces kudriavzevii</i> IFO 1802	236	32	53	532	29-30	26-32	30-34
AACF01000006.1		<i>Saccharomyces castellii</i> NRRL Y-12630	219	30	51	522	27-28	27-31	28-32
CR382128.1/XM_500380	XP_500380 (FALDH3)	<i>Yarrowia lipolytica</i> CLIB99	342	40	60	529	38-39	33-37	37-43
CR382132.1/XM_505802	XP_505802 (FALDH4)	<i>Yarrowia lipolytica</i> CLIB99	340	38	59	519	38-40	34-36	38-40
CR382131.1/XM_503981	XP_503981 (FALDH2)	<i>Yarrowia lipolytica</i> CLIB99	327	38	58	521	35-37	32-36	37-39
CR382127.1/XM_500179	XP_500179 (FALDH1)	<i>Yarrowia lipolytica</i> CLIB99	307	38	57	533	33-36	32-38	35-38
AAEY01000059	EAL17607.1	<i>Cryptococcus neoformans</i> var. <i>neoformans</i> B-3501A				535	36	28-33	34-36
XP_452089.1	CR382122.1	<i>Kluyveromyces lactis</i> NRRL Y-1140	248	32	54	521	30-32	28-32	31-34
CR380957.1	CAG61338.1	<i>Candida glabrata</i> strain CBS138	239	32	53	527	29-31	27-30	29-31

*Candida*, *Yarrowia*, *Cryptococcus*, *Debaryomyces* (all yeasts) and *Aspergillus* (filamentous fungi) are well known genera of microorganisms capable of aliphatic hydrocarbon assimilation, whereas *Kluyveromyces* and *Saccharomyces* lack hydrocarbon-assimilating representatives (Markovetz *et al.*, 1968; Bos and de Bruyn, 1973; Watkinson and Morgan, 1990). Thus, with the exception of *Saccharomyces* and *Kluyveromyces*, all yeast strains identified in this study containing FALDH/FALDH-like gene sequences are able to grow on alkanes (Bos and de Bruyn, 1973). However, *S. cerevisiae* (and possibly all the other *Saccharomyces* and *Kluyveromyces* yeast strains) have been shown to be able to produce long chain alcohols through metabolism of glucose and exogenously provided lipids (White *et al.*, 1987). It is therefore possible that ALDH genes responsible for metabolism of fatty aldehydes arising from oxidation of fatty alcohols might be present in these yeasts as shown by the BLAST search results in this study.

It has also been shown that the ability of microorganisms to degrade alkanes is not necessarily a collective function of a species but rather an individual property of strains and this is more prominent in molds (Nyns *et al.*, 1968). Consequently, no direct growth on *n*-alkanes has been reported in *Aspergillus nidulans* even though *Aspergillus* is a known genus of alkane-degraders. In addition no direct growth on *n*-alkanes has been reported for *Magnaporthe grisea*, *Ustilago maydis* (both phytopathogens) and *Neurospora crassa* (a saprophyte). All four organisms (*A. nidulans*, *M. grisea*, *U. maydis* and *N. crassa*) are however known for exhibiting a multi-drug resistance against a wide spectrum of fungicides, including aromatic and chlorinated hydrocarbon fungicides (Grindle and Temple, 1982; Andrade *et al.*, 2000; Ramesh *et al.*, 2001). Resistance against these hydrocarbon fungicides suggests that the organisms have found a way to detoxify these chemicals, perhaps through the use of some usually dormant genes like FALDHs. Moreover, the ability of these four and other microorganisms (including *S. cerevisiae*) to degrade aromatic hydrocarbons and chlorinated aliphatics can be exploited (Sutherland, 1992). In fact, these microorganisms are being used in bioremediation

of wastewater and soil contaminated with polycyclic aromatic hydrocarbons (PAHs) and BTEX (benzene, toluene, ethyl benzene and xylene) resulting from a variety of human activities such as fossil fuel combustion, industrial processing and pesticide overuse (Sutherland, 1992; Roland, 2002). Overall, PAHs and BTEX are mutagenic and carcinogenic compounds, and thus fungicide resistance and/or ability to metabolize these compounds by fungi is aimed at reducing their mutagenicity and hence detoxification of these compounds. Involvement of ALDH enzymes in detoxification in living cells cannot be overemphasized at this point.

Thus, presence of FALDHs in these fungi may not only be for the obvious aliphatic hydrocarbon assimilation but rather for other functions such as degradation of any other high molecular weight hydrocarbons such as PAHs and BTEX. However this may be depending on the organism's ability to call on the use of these dormant genes whenever such a need arises, i.e. adaptation. It is therefore interesting to observe the similarity of the putative FALDH/ FALDH-like proteins of these organisms with those of other well-known *n*-alkane degraders in order to be able to establish a relationship, if any, among these ALDH genes.

### **Multiple sequence alignment and phylogenetic analysis**

The phylogenetic tree of all ALDHs constructed by Perozich *et al.* (1999) consists of two main trunks, the "Class1/2" trunk and the "Class 3" trunk, with the latter containing in addition to others (aromatic ALDHs, succinic semialdehyde dehydrogenases and glyceraldehyde-3-phosphate dehydrogenases) ALDH3 family itself. The ALDH3 family is the only one in the Class 3 trunk which consists of variable substrate ALDHs. The other families consist of substrate specific ALDHs. In the phylogenetic tree, fungal ALDHs only appear in the "Class1/2" trunk which consists of substrate specific ALDHs. It is further suggested that fungal ALDHs apparently diverged much later in evolution to facilitate adaptation readily to new environments and new evolutionary niches (Hempel *et al.*, 1999).

A phylogenetic tree showing the 25 putative fungal FALDH protein sequences obtained in this study in relation to other ALDHs belonging to the “Class 3” branch of the ALDH tree (Perozich *et al.*, 1999) is shown in figure 4.1.

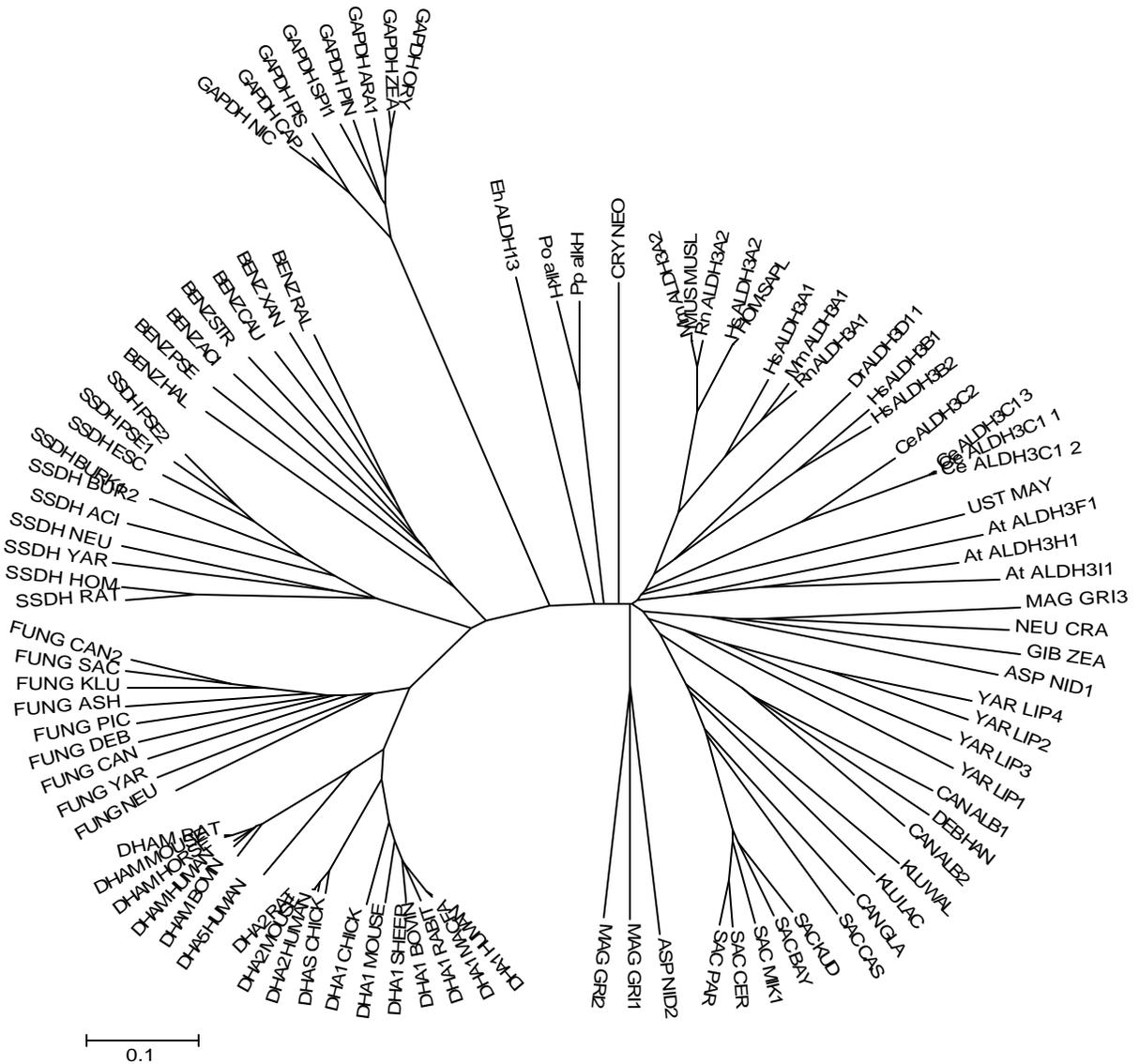


Figure 4.1: Phylogenetic relationships of putative fungal FALDH sequences in relation to fatty aldehyde dehydrogenases (designated ALDH) and short chain aldehyde dehydrogenases (designated DHA) from various organisms including fungi (designated FUNG). The tree was constructed using the program TreeExplorer (<http://evolgen.biol.metro-u.ac.jp/>) using default parameters. The tree was constructed based on 99 amino acid sequences by the neighbor-joining (NJ) method using Kimura distances. Branch lengths are proportional to Kimura distances, which are indicated by the scale bar below the tree. Dehydrogenases from various other organisms were included as references. Abbreviations used are as follows; BENZ - benzaldehyde dehydrogenases, SSDH - succinic semialdehyde dehydrogenases, GAPDH - glyceraldehyde-3-phosphate dehydrogenases and DHA - aldehyde dehydrogenases. Accession numbers for all sequences used are listed in appendix A.

This tree shows that the putative fungal FALDHs are more closely related to class 3 ALDHs than they are to the other members of the “Class 3” branch, which leads us to conclude that these proteins are definitely class 3 ALDHs. The tree also shows that these protein sequences fall into two distinct groups. One larger group comprises all the yeast and some mold ALDH sequences, while the other smaller group consists of mold ALDH sequences only. Only five fungal sequences, namely *Ustilago maydis* (UST\_MAY), *C. albicans* isozymes (CAN\_ALB1 and CAN\_ALB2) and the two *Y. lipolytica* isozymes (YAR\_LIP3 and YAR\_LIP4) showed amino acid identities =40% to some of the previously described ALDH3 sequences. However CAN\_ALB1 sequence shows 41% identity to SAC\_MIK1 (*Saccharomyces mikatae*) sequence, and this protein belongs to the same subfamily as the *S. cerevisiae* (SAC\_CER) which has been designated ALDH14 by Perozich *et al.*(1999). The CAN\_ALB1 sequence also shows 41 and 40% amino acid identity with the YAR\_LIP2 and YAR\_LIP4 sequences. Moreover, *Ustilago maydis* (UST\_MAY) and *Cryptococcus neoformans* (CRY\_NEO) sequences both of which are basidiomycetes as opposed to the ascomycetes (the yeasts) are found in the ALDH3 cluster even though they are at the extreme ends of the group. We therefore propose that all the yeast ALDHs as well as the UST\_MAY and CRY\_NEO sequences belong to the ALDH3 family and represent five subfamilies of the ALDH3 family.

Analysis through sequence alignment indicates that the FALDH/FALDH-like protein sequences show approximately 24-90% identity amongst themselves, 27-40 % similarity with the mammalian FALDH proteins (ALDH3A2) and 27-43% similarity with other (plant and mammalian) class 3 ALDHs. It is also noticeable that in some organisms (e.g. *C. albicans* and *Y. lipolytica*) the isozymes exhibit the same percentage similarity amongst themselves as well as with the reference FALDH protein. In *A. nidulans*, the two isozymes fall into two completely different branches of the phylogenetic tree yet exhibit the same percentage similarity with the

reference FALDH protein. As previously stated, any two ALDH protein sequences exhibiting a  $\approx 40\%$  similarity are described as belonging to two different families, whereas sequences that are  $\approx 60\%$  identical are considered to belong to the same subfamily (Sophos *et al.*, 2001). Based on all these, we conclude that the 25 fungal ALDH sequences belong to several fungal class 3 ALDH families, which can be further divided into two different subfamilies. Thus we have confirmed presence of fungal class 3 ALDH families, and further suggest presence of two new subclasses of fungal FALDHs. However at this stage it is difficult to conclude whether these are two different types of fungal FALDHs or whether one group presents fungal FALDHs while the other is just ordinary class 3 ALDHs. This will however only be elucidated when more fungal ALDH sequences and their activity data become available.

C-albicans1	1	MS-P-----	-----P	4	
c-albicans2	1	MSKPSSIKKSKASAIKPSANSKSKTPKIEETPKLQOVETRLGEVPTTKVSIKRNISITTESVKASEDKSTPQSTNTPAAAVAKSNPNTINAE	P	91	
s-bayanus	1			0	
s-kudriavzevii	1			0	
s-cerevisiae	1			0	
s-paradoxus	1			0	
s-mikatae	1			0	
s-castellii	1			0	
C-glabrata	1			0	
K-lactis	1			0	
k-waltii	1			0	
y-lipolytica2	1			0	
y-lipolytica4	1			0	
y-lipolytica3	1			0	
y-lipolytical	1			0	
m-grisea3	1			0	
n-crassa	1			0	
G-zeae	1			0	
a-nidulans1	1			0	
hs-FALDH	1			0	
Hs-ALDH3A1	1			0	
U-maydis	1			0	
m-griseal	1		MREWLGNLRFAAIFIRIQSRPDHSLRFTVLDPLPVHYRLSPSS	45	
m-grisea2	1			0	
a-nidulans2	1			0	
Sc-ALDH1	1			0	
D-hanseni	1			0	
At-ALDH3	1			M 1	
C-neoformans	1			0	
C-albicans1	52	SKID-----ES-----FTDIG-NTSTSWY-----	TEIKSIEPGVKHLTDSFFT	41	
c-albicans2	9	AKIPNEKSLKTESPQQNGATTTEKESDVLETKSTSSSTTVSNNNNSVLQY-----	TELSEIPIGVERITKAFHS	162	
s-bayanus	1	MSNTDC--KI-----	LRVTLSEIDQKIQA-----	SGDCFFD--KQLKLSHES	39
s-kudriavzevii	1	MPNCDA--MI-----	LPYPLPEIDQKVQA-----	SRDFFFD--KQLKLSHES	39
s-cerevisiae	1	MSNDGS--KI-----	LNYPVSKIDEIVEI-----	SRNFFFE--KQLKLSHEN	39
s-paradoxus	1	MSNSGS--KI-----	LNYPVSKIDEKIET-----	SRNFFFE--KQLKLSHEN	39
s-mikatae	1	M-N-----	LHYTSLSEIDEKVQI-----	SKTFFFE--KQLKLSHES	33
s-castellii	1	M--AT-----	LNYPVSEIDTKIQS-----	SRNFYLA--RQLQLSHEK	34
C-glabrata	1	MA--EI-----	LNYTELSEIDAKISI-----	SRNYFQQ--NRKELANQK	35
K-lactis	1	M--SK-----	FKFTDLKSIPIKIEL-----	AKQFYRS--RQARLAQSK	34
k-waltii	1	ME-----	ANYTPVEEIDAIVAS-----	CKNSYLK--RQKEVANSR	33
y-lipolytica2	1		MSEFDWESILPA-----	TPLGEIEKDIQTLRQGFPS	31
y-lipolytica4	1		MSTFDWESIVPA-----	TPLDQIPGDIQRLRKGFRS	31
y-lipolytica3	1	MTTTATETPTTNTVPTTSLPKETASPGGTASVNTSPDWESICGK-----	TPLEEIESDISRLKKTFRS	63	
y-lipolytical	1		MSWETITPP-----	TPIDTFDSNLQRLRDSFET	28
m-grisea3	1	MWKGGPSHMLMSWKKEIELPVKLYNHNGKPIVMNGSKAGISSIPAFEA-----	TALDSIRPTCDKVLMPFAS	67	
n-crassa	1		MAASKVEIAPFEV-----	TPLDALPAVCSTARATFAS	32
G-zeae	1		MADNTSYTTPGLEH-----	TPLDEINAKVDLVRKTFRS	34
a-nidulans1	1		MGAIIDPELQF-----	TPIEEIQERVSRVVKTFLE	30
hs-FALDH	1		MELEVR-----	VRQAFLS	14
Hs-ALDH3A1	1		MSKISEAVKR-----	ARAAFFS	17
U-maydis	46	TLARLHQPRALTSFAHLALLPSSISPSPTTLRLVCLVNSHHSSVSSSLKRQSPIMAAAAATAEAGLQYTPIDDIPISVISDLRAAFLT	136		
m-griseal	1		MTSTTTTIVTSVTPIKTTY-----	STPAEVDDAHSHTLHATFRT	37
m-grisea2	1		MDTHSY-----	TSPETFSEIHKTLNATFAS	25
a-nidulans2	1		MAEY-----	TSDADVLEAYTTLQTTFKS	23
Sc-ALDH1	1	MTKLH-----	FDTAEPVKITLPEGLTYEQP-----	TGLFINNKFMKAQDGKTYPVDEP	48
D-hanseni	1	MFKNSIAKQARSNVAYAKSFIRMS-----	SNLPLOVSIKLPNGTYYEQP-----	TGMFINNEYVYVYKQKKTDFVISP	67
At-ALDH3	2	ASRRVSSLLSRFSMSSRSIFSLRGMNRGAQRYSNLAAAVENTITPPVKVEH-----	TQLLIGGRFVDAVSGKTFPTPLD	76	
C-neoformans	1		MDMYPDPDHTP-----	TPLPQIEQIYDVTNKRKFRS	29
C-albicans1	42	KQKTHDIQFRLNQLRNYAYADNVDAKCE--ALSDFNSASSETKNLIVGGLNELVHTIASLHEWV-----	KPEK	111	
c-albicans2	163	G-KTHSLQFRLKQLRNYAYTKDNQEAACD--ALQKDFHLPSESTRNYEFATGLNELVFIQSLHKWS-----	KPQP	231	
s-bayanus	40	DPRKEDLQFRQLKQKYYAKDHEEEDID--AMYQDFHTKLESIANSTTKLMNDILHLIEILPKIM-----	KPKR	109	
s-kudriavzevii	40	DPRKKDLQSRQPOLKQKYYAKDHEEEDID--AMYQDFHTKLESVFSSTTKLMNDILHLIEILPKMM-----	KPKR	109	
s-cerevisiae	40	NPRKKDLQFRQLKQKYYAKDHEEEDID--AMYKDFHFNKIESVLNSTTKLMNDILHLIEILPKLI-----	KPKR	109	
s-paradoxus	40	NPRKKDLQFRQLKQKYYAKDHEEEDID--AMYKDFHTKVESVLNSTTKLMNDILHLIEILPKMM-----	KTRK	109	
s-mikatae	34	NPRKKDLQFRQLKQKYYAKDHEEEDID--AMYQDFHTKLESVLSSTTKLMNDILHLIEILPKMM-----	KPQK	103	
s-castellii	35	SPAKADLFRQTLKQKYYNSLKHEENLIQ--AMPQDYHARQESIGTINPLLNATLHIINHLKSKWM-----	KPKR	104	
C-glabrata	36	NVRSADNQRQFQWLKQKYYAKDHEEEDID--ALEKDFHNRQESISLTVVYKLLGDLHLIEIGIPKWN-----	KPKK	105	
K-lactis	35	SPRATDINERIRSLQKQKYYAKDNEERLIA--AASKDFHARSETLTIILVPLVYNNIYLILKMKKEI-----	EPKK	104	
k-waltii	34	DPLKDLLELRKALKSYYGKDRDEEDIQ--ALQKDFHNRQESVLMMLIVLYNHIHLAIEQLPNWI-----	KPEK	103	
y-lipolytica2	32	G-KTLDLNFRLDQIRKQKYYADNVDAKCE--AIHKDLGQPVFETELCISFQWGEFNNVSNLKKWA-----	ADET	100	
y-lipolytica4	32	G-KTLDLNFRLDQIRKQKYYADNVDAKCE--AVYKDLGQPKHETDLCVGLFWGEFNNVSNLKKWA-----	ADED	100	
y-lipolytica3	64	G-KTLDLNFRLDQIRKQKYYADNVDAKCE--AIKADLGPDPFETMAAFSVQMGEFNYVVKNLPKWV-----	KDEK	132	
y-lipolytical	29	G-KLDSVDYRLEQLRQKYYADNVDAKCE--AVTKDLHPRPFETELTIVLFRVDFSTVILKRLKVV-----	KEEK	97	
m-grisea3	68	H-KTKDLWRKQVLRKQKYYADNVDAKCE--AMQLDFGQPAYEVLLSISFVKNDCMFMLSNLNTFA-----	KDEY	136	
n-crassa	35	H-KTKDLWRKQVLRKQKYYADNVDAKCE--ALQKDLGQGGYEDFTVVDVKNDCMLHMNNLNTFA-----	KTEK	101	
G-zeae	33	G-RTKDLEFRMRQIRKQKYYADNVDAKCE--ALLKDLGQCKYEAVLAIDWCKQCECLDMNMEKWL-----	RDEP	103	
a-nidulans1	31	H-KTRDVEFRLVQLRQKYYAKDHEEEDID--ALRSDLGQPEFTEVSVVLENDIVFISKNLHKWV-----	KDEK	99	
hs-FALDH	15	G-RSRPLRFRLQQLKQKYYADNVDAKCE--AIAADLQKSEFNVSQSVITVLEGEIDFMLENLPEWV-----	TAKP	83	
Hs-ALDH3A1	18	G-RTRPLQFRIQQLKQKYYADNVDAKCE--ALAADLQKSEFNWYVYVYVLEIEIYMIQKLPWA-----	ADEP	86	
U-maydis	137	G-KTRSVEYRKNQQLKQKYYADNVDAKCE--SLRKLKDRHSPESIFALMGTTNEIEVAATKIDKWA-----	KPAK	205	
m-griseal	38	G-LTKDLAWRRWQLKQKYYADNVDAKCE--ALKADLNHRPFESLFTIRSVKADISHLKNELDWT-----	STKP	106	
m-grisea2	26	G-KTKSEWRKQVLRKQKYYADNVDAKCE--ALQKDLGQHESMRAALSGLKSIDILHLIKHVEEWA-----	ATEP	94	
a-nidulans2	24	G-KTKELAWRKQVLRKQKYYADNVDAKCE--ALKADLNHRPFETFTFCANVKGDVIEHLKNIKDKWT-----	ADQ	91	
Sc-ALDH1	49	STENTVCEVSSATFDEYALACADRAHDTETWATQDPEFGRLLSKLAIELESQIDLVSSIEALDNKG--TLALARGDVTIAINCLRDAAA	138		
D-hanseni	77	STEEHIDVVEALEEDDTAEEAASAAKRS--SWSTGDPVATVSRKLLKLAALVDANADTLAHIEALDNKG--SLMCSRGDVALTMYFRFYCAG	156		
At-ALDH3	68	RNGEVIQVSEGDADENRAAAARKADGEPWPKMAYERSKLLFRFAALIEKHNDIEAALLETWNGKPYEQSAQIEVPMPLARVFRYAG	167		
C-neoformans	30	G-VTCSLAYRFLNFKQKYYADNVDAKCE--AVENDQKQPPDVTIGLWLPALNEIDLAVKNVENWMMK-----	DESR	99	



	VIII	IX	X		
C-albicans1	358	WDDSSKGEIFGCPFLP	LTYYDKLTSLSRDTIRNNDTFLAQYVFTSG	STSRKYNRQLDQILTTIRSGGLIVNDVLMHVALINAPFGGGVQGS	447
C-albicans2	477	WDDSSKGEIFGCPFLP	LTYYDLEKSCRETIANNDTFLAQYVFTSG	PTSRQYNSQINTITTLVRSGLVINDVLMHVALHNAAPFGGGVGTG	566
s-bayanus	361	WDDPLMQENFAPFLP	LEYDNLDETDIDELKKHDTFLVQYIFSDN	-----QTEINRILTRLRSGGCVVGDVTVHVAITDAAPFGGIGS	444
s-kudriavzevii	361	WDDPLMQENFAPFLP	LEYDNLDETDIDELKKHDTFLVQYIFSDS	-----QTEINRILTRLRSGGCVVGDVTVHVAITDAAPFGGIGS	444
s-cerevisiae	361	WDDPLMQENFAPFLP	LEYDNLDETDIDELKKHDTFLVQYIFSDS	-----QTEINRILTRLRSGGCVVGDVTVHVAITDAAPFGGIGS	444
s-paradoxus	361	WDDPLMQENFAPFLP	LEYDNLDETDIDELKKHDTFLVQYIFSDS	-----QTEINRILTRLRSGGCVVGDVTVHVAITDAAPFGGIGS	444
s-mikatae	355	WDDSLMQENFAPFLP	LEYDNLDETDIDELKKHDTFLVQYIFSDS	-----TTEIEHLTRLRSGGCIVGDVTVHVAITDAAPFGGIGS	438
s-castellii	352	WDDSTMQENFAPFLP	LKYSNLDLDTLDTKTIATNNELVQYIFSDS	-----NEEISHLRLRSGGCIVNDTILHVGITDAAPFGGIGS	435
C-glabrata	356	WDDALMQENFAPFLP	LKYTDIDATIDKLSKEDTFLVQYIFSEK	-----QKEIDHILARLRSGDCICDFTTIVHGIKDAAPFGGIGS	439
K-lactis	351	WDDPSMQENFAPFLP	LMRYDLDLDETDIDELKKHDTFLVQYIFSDS	-----KDEIDHILTRLRSGACVVDNLLHVGITDAAPFGGIGS	434
k-waltii	350	WDDALMQENFAPFLP	VTYSNLDLSAISKVELLNLYLAIYFSDN	-----RREDIQIMFQIKSGGCVIDGTIVHVAAMDAPFGGIGS	433
y-lipolytica2	345	DGDSLWDEIFGCPFLP	LTARSVDEGIRYVHENNDTFLAMVYFTDN	-----ASEGEYIQSQINSGGLIFNDLTVHVGCVQAPFGGGVGS	428
y-lipolytica4	344	DNDSLWDEIFGCPFLP	LTARSVDEGIRYVHENNDTFLAMVYFTDK	-----ASEGDYIQSQINSGGLIFNDLTVHVGCVQAPFGGGVGS	428
y-lipolytica3	375	ANDSLWDEIFGCPFLP	TKAMSTEAAACNFVLEHHPTEAEYIFSDN	-----NSEIDYIRDRVSSGLVINDLTVHVGCVQAPFGGGVGS	458
y-lipolytical	344	DDDLWDEIFGCPFLP	IKPRTLGGQVDYVTRNNDTFLAMVYFSD	-----PKEVDWLQTRIRAGSVNINEVIEQVGLASLPLSGVGS	427
m-grisea3	381	ADDSMVEEESFCPLFA	MAFDLNAIAIKLNSVKTFLALYAFGKD	-----ETNKVINEMTSGGASINDSFAHAKLDAAPFGGIGS	463
n-crassa	344	LDDPMVEEESFCPLFS	YVPVDLQALSIANNVHTFLALMAGDKS	-----ETNRILDEMTSGGACINDSYFHGAVHTVAPFGGGVGS	426
G-zeae	346	IEDSMVDEAFCPFLFA	MAIDSLDQAIIDNSVDPTFLSLSTFGSK	-----DENKKGATCNDAFVHVAITDAAPFGGGVGS	421
a-nidulans1	332	PDDSMVQESFCEPLIP	LPVDNIDEAININNSIQSTLGLYFPGSK	-----DTEKILSQTRSGCVSVNDAALH-IPTLFGGVEG	412
hs-FALDH	323	PKTKVWDEIFGCPFLP	VPVKNVDEAINFINERKK-FLALYVFSHN	-----HKLIKRMIDETSSGGVTGNDVIMHFTLNSPFGGGVGS	405
Hs-ALDH3A1	326	PQSPVWDEIFGCPFLP	IVCVRSLEEAIQFINQRK-FLALYMFSSN	-----DKVIKMIETSSGGVAANDVIVHITLHSPFGGGVGS	408
U-maydis	450	PDDAIVGEGIFGCPFLP	VPVVRDVEAAVDLINSRQ-FLALYLPAGD	-----NRVKNFPPDNTSRGACVQGDVTVHVAITDAAPFGGGTGP	532
m-griseal	348	LSDPLVSEELFCPLP	ITAT-PDEAIRTSSGDS-FLALYMFARD	-----DAFVHVLDRTLSSGGVTNNIAVHVAITDAAPFGGGVGS	429
m-grisea2	333	MSDSLVSEELFCPLP	ITAT-PNEAIAAINSLEL-FLALYVFSND	-----QKEIDHVLSTISGGVTINDALMHAAPVNAAPFGGGVGS	414
a-nidulans2	331	MNDPLVSEELFCPLCP	IKGS-TDDAIAKINSLEL-FLALYVFSQD	-----QNEVEHISLSTSSGGCVNGLVHVAITDAAPFGGGVGS	412
Sc-ALDH1	396	EDMRIEIEEIFGCVVT	AKFKTLEEGVEMNSSEF-FLGSGIETES	-----LSTGLKAKMLKATVWINTYNDP-STRVFGGVEG	476
D-hanseni	413	EDMQIIEEIFGCVVT	TKFKTIDEVVDLNSDF-FLAAGIHTEN	-----INTAIGVANRLQATVWINTYNDP-PMVFGGGYQS	493
At-ALDH3	424	DDMLIATDEIFGCPFLP	LKFKDLDEVIARANNRSY-FLAAGIHTQN	-----LDTAHLRMLRSLRVTVWVINCVDVLD-ASLFGGGVGS	504
C-neoformans	353	REEGGVWDEIFGCPFLA	IPVDEDIDGAIAYINAREK-FLALYVCAKG	-----RSIFEKVISQTSSTATWNDPAMATFARLAPFGGGVGS	436
C-albicans1	448	GGCSYHKKFSFRSFTHERTTMEQK	-----LWDFMVKVRYPPYNSKNDLVRVSSQQYNGKVVFSRTGDVP	-----VAG-----PGRF	520
C-albicans2	567	GNCAYHGEFSYRAFTHERTVLEQH	-----LWDFMVKVRYPPYANKKDKLIASSQQYGGRRVFNREGNVR	-----IGG-----PPLL	639
s-bayanus	445	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
s-kudriavzevii	445	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
s-cerevisiae	445	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
s-paradoxus	445	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
s-mikatae	439	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
s-castellii	436	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
C-glabrata	440	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
K-lactis	435	GYGNYGGLHGFNTFSHERTVFKQP	-----YWNDFTLFMRYPNNERKEKLLRFALE	-----RKPWFDRNGDN-----KWG-----LRQY	513
k-waltii	434	GSNNYHKKWSEFTPTHERTVVRQP	-----FWDFLLKRLRYPPPTAKKVKLLKATE	-----RQPWFDRNGDN-----HRS-----YKLI	508
y-lipolytica2	429	GYGNYHKKWSEFTPTHERTVVRQP	-----FWDFLLKRLRYPPPTAKKVKLLKATE	-----RQPWFDRNGDN-----HRS-----YKLI	508
y-lipolytica4	428	GYGNYHKKWSEFTPTHERTVVRQP	-----FWDFLLKRLRYPPPTAKKVKLLKATE	-----RQPWFDRNGDN-----HRS-----YKLI	508
y-lipolytica3	459	GNCAYHGEFSYRAFTHERTVLEQH	-----LWDFMVKVRYPPYANKKDKLIASSQQYGGRRVFNREGNVR	-----IGG-----PPLL	639
y-lipolytical	428	GYGNYHKKWSEFTPTHERTVVRQP	-----FWDFLLKRLRYPPPTAKKVKLLKATE	-----RQPWFDRNGDN-----HRS-----YKLI	508
m-grisea3	464	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
n-crassa	427	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
G-zeae	422	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
a-nidulans1	413	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
hs-FALDH	406	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
Hs-ALDH3A1	409	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
U-maydis	533	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
m-griseal	430	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
m-grisea2	415	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
a-nidulans2	413	GCAYRHKASFDQFDSHRRSCLVRPLM	-----NDEGLKRVYPP-SPAKMTQH	-----	478
Sc-ALDH1	477	GYGREGVEEVYHAYTEVKAVRIKL	-----	-----AIKRGLL--LTKS	488
D-hanseni	494	GYGREGVEEVYHAYTEVKAVRIKL	-----	-----AIKRGLL--LTKS	488
At-ALDH3	505	GYGREGVEEVYHAYTEVKAVRIKL	-----	-----AIKRGLL--LTKS	488
C-neoformans	437	GYGREGVEEVYHAYTEVKAVRIKL	-----	-----AIKRGLL--LTKS	488
C-albicans1	521	FSSWNSF-----AGVIGLLETFITNKQ	-----	-----EESL	505
C-albicans2	640	FSAWNSF-----AGVIGLLETFITNKQ	-----	-----EESL	505
s-bayanus	514	LSLSAVV-----VLVSGICACYSSLF	-----	-----EESL	505
s-kudriavzevii	514	FSLSAVF-----MLISAICAYRSS	-----	-----EESL	505
s-cerevisiae	514	FSLSAAV-----ILISTIYAHCSS	-----	-----EESL	505
s-paradoxus	514	FSLSAAV-----ILISTIYAHCSS	-----	-----EESL	505
s-mikatae	508	FVSVVMI-----VLVGAICAHFY	-----	-----EESL	505
s-castellii	505	FPLSFLI-----IISGMLYKALV	-----	-----EESL	505
C-glabrata	509	AMFSVTC-----ALLAVGVKSPFV	-----	-----EESL	505
K-lactis	504	GIILTSG-----VLLGFLINAMM	-----	-----EESL	505
k-waltii	503	LGRATFG-----VLLAVGSKALYDAYA	-----	-----EESL	505
y-lipolytica2	498	FGKLFWF-----VIVLGLGAASLKSIGFL	-----	-----EESL	505
y-lipolytica4	498	FRKIWFV-----VLIVAIGAAGAKALL	-----	-----EESL	505
y-lipolytica3	528	FN	-----	-----EESL	505
y-lipolytical	499	LGNKLLW-----IIIAALVAYAKRNEILLITIAQIMSVFIK	-----	-----EESL	505
m-grisea3	537	FGMGAAGPKVLRDRERASALEACISRDWYGLHSLGSSIRANAWPGDASWALIASKDWLVD	-----	-----EESL	505
n-crassa	500	LGLGKTK-----GVKGALMRWLVVAVGYLVSAYMKARRA	-----	-----EESL	505
G-zeae	495	FGLGSKK-----AKGALLRWAFVVAADVLEAKKGTLSQLLTR	-----	-----EESL	505
a-nidulans1	483	LTLGGGS-----AKAGAGRAAVVAAYVFMKIIEERSKL	-----	-----EESL	505
hs-FALDH	479	LVKYYQA-----VLRKALLIFLVHRLRWSKQR	-----	-----EESL	505
Hs-ALDH3A1	454		-----	-----EESL	505
U-maydis	613	LSHSRPS-----PFLSMSQSLFPMVHYNMLPTQ	-----	-----EESL	505
m-griseal	496	AGRGRLWYAT-----AVGVGVVAAVGARYLQA	-----	-----EESL	505
m-grisea2	498	IGAGLYWIDD-----IRSAKLAVLGFVRRVGGEL	-----	-----EESL	505
a-nidulans2	489	LKSARFLVYL-----VLLGLADVGLDRRLGVLSKVRDLLEVEVRSFL	-----	-----EESL	505
Sc-ALDH1	501		-----	-----EESL	505
D-hanseni	522		-----	-----EESL	505
At-ALDH3	535		-----	-----EESL	505
C-neoformans	506	RRRKRWR-----VIVLGLFMLGALVGGKRLGNGQ	-----	-----EESL	505

Figure 4.2: DNAssist multiple sequence alignment of the 25 FALDH/FALDH-like protein sequences and other class 3 ALDH proteins for comparison. The parts containing the 10 conserved ALDH sequence motifs are shown in boxes and numbered in Roman numerals.

## Identification of conserved residues and motifs

### (a) Conserved Motifs

Catalytically important residues and segments in the ALDH structure are highly conserved. These conserved residues are essential for maintaining critical turns and loops in the tertiary structure of the ALDH protein, which in turn has direct bearing on functional definition of the protein. Sequence comparisons among ALDH genes from bacteria, plants and animals demonstrating ALDH enzymatic activity have shown at a glance three diagnostic amino acid motifs; (i) the ALDH glutamic acid active site signature sequence MELGGNA, (ii) the Rossmann fold GXGXXG coenzyme binding site and (iii) the catalytic thiol (Kirch *et al.*, 2004). However, a closer inspection has revealed that there are ten conserved sequence motifs in the ALDH family (Perozich *et al.*, 1999). The ten conserved motifs in ALDHs are stretches of sequences ranging from five up to 14 or 15 amino acids (Perozich *et al.*, 1999). They are spread along the entire ALDH sequence, but fold back together and come into contact with each other in the 3-D structure (Perozich *et al.*, 1998). Overall all these motifs reside at or near the active site of the enzyme, and appear to effect essential ALDH structure/function elements, since most contain a conserved turn or loop (Perozich *et al.*, 1998) with a highly conserved hydrophobic small amino acid such as glycine, proline, aspartic acid or asparagine, which does not take part in enzyme function. Table 4.3 gives a list of the ten conserved ALDH motifs (Perozich *et al.*, 1998) and the frequency of appearance of individual residues in the 25 putative fungal FALDH/FALDH-like sequences.

An alignment of the fungal ALDH sequences (figure 4.2) shows that they all have the ten ALDH conserved sequence motifs. Motif 1 is known to be the most conserved in all ALDHs whereas motif 3 is generally lacking in class 3 ALDHs (Perozich *et al.*, 1999). A closer look at table 4.3 however shows that both motifs 1 and 3 are very variable in the fungal FALDHs, and motif 1 is even absent in one case.

**Table 4.3: Ten most conserved sequence motifs in ALDHs and the frequency of appearance in the 25 putative fungal FALDH sequences**  
(adapted from Perozich *et al.*, 1999).

Motif number	Length	Motif <sup>a</sup> in all ALDHs	Motif <sup>b</sup> in fungal FALDHs
1	5	[Past]-[WFy]-[Ne]-[FYgalv]-[Ptl]	[P <sub>18</sub> a <sub>5</sub> t]-[F <sub>11</sub> W <sub>10</sub> y <sub>3</sub> ]-[N <sub>24</sub> ]-[F <sub>13</sub> Y <sub>11</sub> ]-[P <sub>24</sub> t] (this motif is missing in <i>A-nidulans</i> 1)
2	14	[Apnci]-[Liamv]-[Avslcimng]-[ACtlmvvgf]- <b>G</b> -[Ncdi]-[Tavcspg]- [Vaimfcltgy]-[Vil]-[Lvmiwafhcy]-[ <b>Kh</b> ]-[Ptvghms]-[ASdhp]- [Epsadqgilt]	[A <sub>25</sub> ]-[I <sub>16</sub> l <sub>8</sub> t]-[A <sub>19</sub> t <sub>2</sub> g <sub>2</sub> s <sub>2</sub> ]-[A <sub>19</sub> g <sub>3</sub> sct]-[G <sub>25</sub> ]-[N <sub>18</sub> c <sub>7</sub> ]-[T <sub>14</sub> c <sub>3</sub> a <sub>2</sub> s <sub>4</sub> p <sub>2</sub> ]-[V <sub>15</sub> i <sub>5</sub> a <sub>5</sub> ]- [V <sub>13</sub> i <sub>6</sub> a <sub>4</sub> l <sub>2</sub> ]-[L <sub>19</sub> i <sub>3</sub> fvw]-[K <sub>25</sub> ]-[P <sub>18</sub> f <sub>2</sub> m <sub>2</sub> q <sub>1</sub> t]-[S <sub>24</sub> a]-[E <sub>24</sub> d]
3	10	[Grkpwhsay]-[FLeivqnarmhk]-[Pg]-[Plakdievsrf]-[Gnde]-[Vliat]- [VLifyac]-[Nglqshat]-[Vilyaqgfst]-[IVlms]	[A <sub>8</sub> G <sub>7</sub> y <sub>5</sub> r <sub>2</sub> s <sub>2</sub> q]-[L <sub>17</sub> f <sub>8</sub> ]-[D <sub>15</sub> p <sub>7</sub> egt]-[P <sub>12</sub> s <sub>3</sub> d <sub>2</sub> k <sub>3</sub> ng <sub>1</sub> rte]-[G <sub>10</sub> d <sub>4</sub> e <sub>5</sub> q <sub>8</sub> 4r]- [L <sub>7</sub> a <sub>5</sub> i <sub>4</sub> c <sub>3</sub> v <sub>3</sub> gq <sub>w</sub> ]-[F <sub>9</sub> I <sub>8</sub> v <sub>5</sub> y <sub>2</sub> ]-[Q <sub>9</sub> r <sub>4</sub> a <sub>2</sub> t <sub>2</sub> desl <sub>f</sub> knv]-[V <sub>14</sub> a <sub>4</sub> i <sub>4</sub> yct]-[V <sub>21</sub> f <sub>2</sub> il]
4	10	[IVlgy]-[SATmnlfhq]-[Fyla]-[Tvil]- <b>G</b> -[Sgen]-[Tsvrindepaqk]- [EAprqkktvnlhd]-[VTiasgm]-[Gafi]	[L <sub>24</sub> V]-[F <sub>15</sub> m <sub>6</sub> vac1]-[Y <sub>18</sub> f <sub>7</sub> ]-[T <sub>25</sub> ]-[G <sub>25</sub> ]-[S <sub>17</sub> n <sub>6</sub> g <sub>2</sub> ]-[P <sub>7</sub> t <sub>6</sub> g <sub>4</sub> h <sub>2</sub> a <sub>2</sub> vksn]- [r <sub>6</sub> t <sub>6</sub> k <sub>3</sub> a <sub>3</sub> q <sub>d</sub> lnvph]-[V <sub>22</sub> i <sub>2</sub> t]-[G <sub>23</sub> a <sub>2</sub> ]
5	16	[Lamfgs]-[Enlqf]-[Ltmcagi]-[Gs]-[Ga]-[Knlmqshiv]-[SNadc]- [Pahftswv]-[cnl <sub>f</sub> mgivahst]-[Ivlyfa]-[Viamt]-[Fdlmhcanyv]-[Daeskprnt]- [Dsntaev]-[Acvistey]-[Dnlera]	[L <sub>25</sub> ]-[E <sub>25</sub> ]-[L <sub>25</sub> ]-[G <sub>25</sub> ]-[G <sub>25</sub> ]-[K <sub>18</sub> q <sub>4</sub> lir]-[S <sub>18</sub> n <sub>5</sub> c <sub>2</sub> ]-[P <sub>25</sub> ]-[a <sub>8</sub> v <sub>10</sub> t <sub>7</sub> ]-[F <sub>18</sub> i <sub>5</sub> vy]- [I <sub>13</sub> V <sub>12</sub> ]-[T <sub>17</sub> l <sub>2</sub> ds <sub>2</sub> haf]-[k <sub>9</sub> e <sub>6</sub> d <sub>2</sub> s <sub>3</sub> rg <sub>2</sub> pn]-[N <sub>11</sub> h <sub>3</sub> s <sub>4</sub> t <sub>2</sub> d <sub>4</sub> k]-[A <sub>10</sub> f <sub>7</sub> c <sub>4</sub> iv <sub>2</sub> ]-[k <sub>6</sub> d <sub>6</sub> n <sub>6</sub> ps <sub>3</sub> q <sub>3</sub> ]
6	8	[Fyvlna]-[Fgylrmdaqetwsvikp]-[Nhstyfaci]-[QAsnhtcmg]- <b>G</b> -[Qe]- [crvitsand]-[Cr]	[F <sub>15</sub> v <sub>2</sub> l <sub>2</sub> t <sub>2</sub> ymsq]-[g <sub>8</sub> v <sub>5</sub> l <sub>4</sub> f <sub>2</sub> sytqm <sub>2</sub> ]-[N <sub>24</sub> t]-[A <sub>14</sub> s <sub>7</sub> cg <sub>2</sub> t]-[G <sub>25</sub> ]-[Q <sub>24</sub> e]-[I <sub>12</sub> t <sub>8</sub> v <sub>5</sub> ]-[C <sub>25</sub> ]
7	9	[Gdtskac]-[Yfnarthclswv]-[FYlwvis]-[IVlfym]-[Qeapkgmynhlswyv]- [Pa]-[Tachlmy]-[VII]-[FLivwn]	[T <sub>11</sub> l <sub>5</sub> e <sub>2</sub> nkgsqdp]-[r <sub>8</sub> c <sub>5</sub> l <sub>6</sub> k <sub>3</sub> hiy]-[F <sub>9</sub> y <sub>4</sub> i <sub>3</sub> l <sub>2</sub> c <sub>3</sub> ksmr]-[I <sub>12</sub> l <sub>5</sub> v <sub>5</sub> m <sub>3</sub> ]-[P <sub>12</sub> a <sub>4</sub> e <sub>4</sub> q <sub>2</sub> g <sub>2</sub> k]- [P <sub>22</sub> gvi]-[T <sub>23</sub> s <sub>2</sub> ]-[V <sub>12</sub> i <sub>7</sub> a <sub>3</sub> l <sub>3</sub> ]-[V <sub>17</sub> i <sub>7</sub> f]
8	7	[Ektdrqs]- <b>E</b> -[Ivtnfsp]- <b>F</b> -[Ga]-[Ps]-[Vilef]	[q <sub>7</sub> e <sub>8</sub> d <sub>6</sub> g <sub>4</sub> ]-[E <sub>25</sub> ]-[n <sub>9</sub> i <sub>8</sub> l <sub>4</sub> s <sub>3</sub> a]-[F <sub>25</sub> ]-[G <sub>17</sub> a <sub>8</sub> ]-[P <sub>24</sub> t]-[I <sub>12</sub> v <sub>9</sub> l <sub>4</sub> ]-[
9	15	[Nrst]-[Dnaseqtkregi]-[TSrvnalcqgik]-[Epd <sub>t</sub> gqikvrfshyncl]-[Yfkqvm]- [Gpa]-[Lnmv]-[Astgvqcf]-[Ags <sub>l</sub> tfc]-[AGysct]-[VII <sub>f</sub> ams]-[Fhwyivlem]- [TSag]-[KRnsqteahdp]-[DNsileakt]	[e <sub>5</sub> s <sub>5</sub> n <sub>6</sub> k <sub>4</sub> r <sub>2</sub> a <sub>2</sub> q]-[n <sub>5</sub> s <sub>4</sub> e <sub>4</sub> v <sub>3</sub> k <sub>2</sub> l <sub>3</sub> hida]-[H <sub>15</sub> d <sub>2</sub> l <sub>2</sub> gfsr <sub>2</sub> q]-[D <sub>13</sub> p <sub>6</sub> a <sub>2</sub> ekrs]-[T <sub>17</sub> r <sub>2</sub> skqy <sub>2</sub> n]- [P <sub>24</sub> g]-[L <sub>25</sub> ]-[A <sub>15</sub> v <sub>8</sub> gs]-[L <sub>10</sub> Q <sub>10</sub> m <sub>3</sub> ea]-[Y <sub>22</sub> smg]-[I <sub>14</sub> v <sub>5</sub> a <sub>2</sub> ptlm]-[F <sub>23</sub> ch]- [S <sub>12</sub> g <sub>4</sub> t <sub>6</sub> a <sub>3</sub> ]-[D <sub>12</sub> s <sub>4</sub> kre <sub>2</sub> g <sub>2</sub> n <sub>2</sub> q]-[d <sub>6</sub> s <sub>6</sub> k <sub>5</sub> g <sub>2</sub> ]
10	12	[Pasw]-[Fwyahv]-[Gtqs]- <b>G</b> -[Fvyesnimitawrq]-[Kgrn]-[mqarelnskghdpt]- [Sm]-[Gfls]-[Ifntlmygshrvq]-[Gdnhrsy]-[Rdpsagkte]	[P <sub>24</sub> a]-[F <sub>23</sub> l <sub>2</sub> ]-[G <sub>24</sub> s]-[G <sub>25</sub> ]-[V <sub>14</sub> i <sub>6</sub> ty]-[G <sub>23</sub> rk]-[d <sub>5</sub> n <sub>4</sub> t <sub>4</sub> q <sub>5</sub> 2mpea <sub>2</sub> ]-[S <sub>25</sub> ]-[G <sub>25</sub> ]- [Y <sub>13</sub> n <sub>2</sub> h <sub>2</sub> m <sub>2</sub> qw <sub>2</sub> tfs]-[G <sub>25</sub> ]-[N <sub>10</sub> A <sub>8</sub> s <sub>3</sub> tgry]

a-Motifs are given as ProSite patterns. Capital letters denote predominant residues at each bracketed position.

b- Index numbers denote frequency of appearance of each amino acid in the fungal ALDH sequences.

Motif 5, which bears the ALDH glutamic acid active site, is the most strictly conserved in all the fungal ALDH sequences. The ALDH glutamic acid active site signature sequence is LELGGKS for these fungal sequences. The ALDH cysteine active site, which is less conserved than the Glu active site is found within motif 6. The Rossmann fold coenzyme-binding site is found in motif 4 as either GXGXXG or GXTXXG in these sequences.

### **(b) Conserved residues**

There are five residues that are absolutely conserved in all catalytically active ALDHs, namely Cys<sup>243</sup>, Glu<sup>333</sup>, Phe<sup>335</sup>, Gly<sup>187</sup> and Gly<sup>240</sup> (Perozich *et al.*, 1999). (The rat cytosolic class 3 ALDH numbering has been adopted in describing the positions of any amino acids mentioned in this manuscript, except where otherwise specified). Cys<sup>243</sup>, which serves as the catalytic thiol, is found within motif 6. while Glu<sup>333</sup> and Phe<sup>335</sup> reside in motif 8 (Hempel *et al.*, 2001). Glu<sup>333</sup> serves as the general base, which activates the catalytic thiol through a water molecule. It may also be involved in cofactor binding (Hempel *et al.*, 2001). ALDHs lacking enzymatic activity such as the  $\Omega$ -crystallins and allergens have been found to have other residues in place of Cys<sup>243</sup>, Glu<sup>333</sup> or Phe<sup>335</sup>. Gly<sup>187</sup> together with Phe<sup>335</sup> form an integral part of the coenzyme binding Rossmann fold by interacting with the nicotinamide portion of NAD(P). Gly<sup>240</sup> maintains tertiary ALDH structure by allowing the main chain to twist back on itself thus enabling positioning of the catalytic nucleophile. All five of these absolutely conserved residues are also found conserved in the 25 putative fungal FALDH proteins, which may lead one to conclude that these are probably enzymatically active ALDH proteins.

A further eleven residues (Arg<sup>25</sup>, Gly<sup>105</sup>, Asn<sup>114</sup>, Pro<sup>116</sup>, Gly<sup>131</sup>, Lys<sup>137</sup>, Gly<sup>211</sup>, Pro<sup>337</sup>, Gly<sup>383</sup>, Asn<sup>388</sup>, and Gly<sup>403</sup>) were found conserved in more than 95% of 145 ALDH sequences examined by Perozich *et al* (1999). Some of these residues lie at critical turns and loops in the ALDH structure. For example, Gly<sup>211</sup> is part of the dipeptide, which marks the boundary between the coenzyme binding and catalytic domains. Gly<sup>403</sup> is involved in one of the U-turn regions whereas Asn<sup>114</sup> is essential

for activity since it stabilizes the carbonyl oxygen of the aldehyde. Eight of these eleven conserved residues are conserved in the fungal ALDHs. Asn<sup>388</sup> is replaced by glycine in *U. maydis* and the *Saccharomyces* sequences. Asn<sup>114</sup> is missing in the *A. nidulans* isozyme1 sequence while Pro<sup>337</sup> is replaced by threonine in the *A. nidulans* isozyme2 sequence.

Lys<sup>137</sup> is proposed to hydrogen bond with the adenine ribose of the NADH/NADPH cofactor (Hempel *et al.*, 2001) and is conserved in all the fungal FALDHs. Other residues, which have been identified as involved in coenzyme binding, are Glu<sup>140</sup> and Gly<sup>192</sup>. Glu<sup>140</sup> and Gly<sup>192</sup> are substituted with Asp and Ala respectively in the *M. grisea* sequences. Alanine at position 192 is often found to confer NADP dependence (Hempel *et al.*, 2003).

Another group of residues has been proposed to be specifically important in the ALDH3 family (Hempel *et al.*, 2003). The group consists of Asp<sup>247</sup>, Tyr<sup>248</sup>, Lys<sup>359</sup>, Pro<sup>360</sup>, Glu<sup>69</sup>, Gln<sup>122</sup>, Pro<sup>123</sup>, Lys<sup>235</sup>, and Thr<sup>242</sup>. The first four residues are important in subunit/subunit interaction recognition at the dimer interface. All these residues except Pro<sup>360</sup> and Glu<sup>69</sup> are generally very variable among these fungal FALDHs. Even Glu<sup>69</sup> is replaced by aspartic acid in the *M. grisea* sequences. Glu<sup>69</sup> and Gln<sup>122</sup> lie in the substrate-binding pocket of the ALDH structure. Gln<sup>122</sup> is very variable in the fungal FALDHs, while Pro<sup>123</sup> is relatively conserved (exceptions are *C. albicans* isozyme1, *D. hansenii* and *G. zeae*). Mutation of Pro<sup>123</sup> is implicated in SLS (Rizzo *et al.*, 2001). Asp<sup>247</sup> has been replaced by asparagine in most of the mold sequences. In humans this D-N mutation results in loss of FALDH activity and is associated with SLS in humans (Rizzo *et al.*, 2001). Histidine is found in place of Tyr<sup>248</sup> in the *Magnaporthe* and *Aspergillus* sequences. Lys<sup>235</sup> and Thr<sup>242</sup> are proposed to probably play a role in protonation of the tetrahedral intermediate, but are quite variable in the fungal sequences. Also, findings of a study of SLS patients by Shibaki *et al.* (2004) strongly suggested that Val<sup>363</sup> is important for structural folding of the catalytic domain and is therefore essential for the normal functioning of the FALDH protein. However, the Val<sup>363</sup> residue is only conserved in

the *Saccharomyces* sequences whereas all the other sequences have alanine instead. One *Aspergillus* sequence has glycine at this position. This lack of strict conservation may suggest that Val<sup>363</sup> is not critically essential for catalysis in the fungal ALDH proteins.

### **Domain analysis of the putative fungal FALDHs**

At present crystal structures of ALDHs belonging to several different groups have been solved (Liu *et al.*, 1997; Steinmetz *et al.*, 1997; Cobessi *et al.*, 1999), and ALDHs are now known to exist in two forms as either dimers or tetramers. Despite the difference in oligomeric state an ALDH subunit consists of three distinct domains namely, the catalytic domain, the coenzyme binding domain and the oligomerization domain.

Class 1, 2 and 3 ALDHs are all ALDHs that accept variable substrates. However, class 1 and 2 ALDH monomers form tetramers that possess a 56-amino acid extension at the N-terminus whereas class 3 ALDH monomers have a 17-amino acid extension at the C-termini, which interacts with other monomers to form dimers (Rodriguez-Zavala and Weiner, 2001). These extensions, the 56-amino acid tail at the N-terminus of class 1 and 2 ALDHs and the 17-amino acid tail at the C-termini of class 3 ALDHs, proposedly affect oligomerization. It has been shown that the N-terminal portion (the first 21 amino acids) of ALDH2 is responsible for folding, assembly and stability of both the precursor and mature ALDH2 tetramers (Zhou and Weiner, 2001). This N-terminal portion is also responsible for translocation of the ALDH into the mitochondria. It has also been shown that the C-terminal tail in ALDH3 determines the quaternary structure of ALDH3 by enabling interaction between two monomers thus favouring formation of dimers in this class of ALDHs and apparently preventing formation of tetramers (Rodriguez-Zavala and Weiner, 2001). A property of mammalian FALDHs, which distinguishes it from the other class 3 ALDHs, is that they are membrane-bound microsomal enzymes. This distinguishing property is reflected in the protein sequences by the presence of hydrophobic domains at the carboxy-terminal (Masaki *et al.*, 1994; Lin *et al.*, 2000),

which are preceded by transmembrane domains (residues 463-480). The C-termini of FALDHs are thus significantly longer than those of other members of the ALDH3 family. Sequence analysis shows that the 25 fungal ALDH3 proteins have N-termini that are significantly longer than those of the mammalian class 3 ALDHs but more or less similar to the plant class 3 ALDHs. However, the C-termini of the fungal ALDHs are very much similar in length to the mammalian FALDHs.

Using the TMHMM program (<http://www.cbs.dtu.dk/cgi-bin/nph-webface>) each protein sequence was analyzed for the presence of transmembrane segments or domains (TMS or TMD) (see table 4.4). Of the 25 fungal FALDH/FALDH-like sequences identified, this program identified only seven sequences as possessing at least one TMS. The ALDHs of *N. crassa*, *S. mikatae*, and *Y. lipolytica* isozymes 2 and 4 have this TMS in the C-terminus, at positions similar to that in mammalian FALDH proteins. However, in the ALDHs of *S. cerevisiae* and *S. bayanus* (residues 130-152 for both) the TMS is found very close to the N-terminal, and in *M. grisea* protein there is a TMS at both terminals (residues 127-149 and 499-521). Using the PSORT III program (<http://psort.ims.u-tokyo.ac.jp/form2.html>) (Horton and Nakai, 1997) TMSs were identified even in some of the other remaining ALDHs. With this program only five sequences were identified as not having any TMS. These are the isozyme *Ylipolytica*3, both isozymes *C-albicans*1 and 2, *K-waltii* and *U-maydis* proteins.

The PSORT III program was used to predict subcellular localization for each fungal ALDH protein. With the exception of *C-albicans*2, which seems to be a nuclear protein (55% reliability), all the other proteins are probably cytoplasmic (75-94% reliability). None of the proteins seem to have N-terminal signal peptide but some do have endoplasmic reticulum-retention motifs at the C-terminus, a second peroxisomal targeting signal and/or a possible vacuolar-targeting motif (see table 4.4). The human FALDH protein is similarly a cytoplasmic protein with no targeting signal sequence. The enzyme is synthesized on free polysomes and then inserted post-translationally into the endoplasmic reticulum without undergoing any post-

translational modification (Miyachi *et al.*, 1991; Vasiliou *et al.*, 1996; Rogers *et al.*, 1997; Chang and Yoshida, 1997; Lin *et al.*, 2000). From the results of the prediction of protein localization sites we gather that these fungal ALDH proteins may also be synthesized on free polysomes and then inserted post-translationally into the endoplasmic reticulum possibly also without undergoing any post-translational modification.

**Table 4.4: Position and number of transmembrane segments (TMS) in the 25 fungal FALDH/FALH-like protein sequences identified. The reference human FALDH protein is also included.**

FALDH (organism)	No of TMSs	Position of TMD residues	Length	Cell localization (% Reliability)	Other features
Human ALDH10	1	463-480	18	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 463) Possible peroxisomal targeting signal: AKL at 148
A-nidulans1	1	119-135	17	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 119)
A-nidulans2	3	126-142 381-397 494-510	17 17 17	Cytoplasmic (94.1%)	Membrane topology: type 3a Leucine zipper pattern: LVYLVLLGLADVGLDRRLGVL at 495
C-albicans1	0			Cytoplasmic (94.1%)	ER membrane retention signals: ITNK
C-albicans2	0			Nuclear (55.5%)	Possible vacuolar targeting motif: ILPI at 490
C-glabrata	0	136-152 506-522	17 17	Cytoplasmic (89%)	ER membrane retention signals: KSFF 2nd peroxisomal targeting signal: KLLGDILHL at 85 Membrane topology: type 3a
C-neoformans	0				No special features found
D-hanseni	0			Cytoplasmic (94.1%)	ER membrane retention signals: LTEK possible vacuolar targeting motif: KLPN at 34
G-zeae	2	138-154 505-521	17 17	Cytoplasmic (89%)	Membrane topology: type 3a
K-lactis	2	122-138 503-519	17 17	Cytoplasmic (89%)	Possible vacuolar targeting motif: ILPI at 364 Membrane topology: type 3a
K-waltii	0			Cytoplasmic (89%)	No special features found
M-grisea1	2	127-149 499-521	23 23	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 141)
M-grisea2	1	128-144	17	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 128)
M-grisea3	1	173-189	17	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 173)
N-crassa	2	137-153 510-527	17 18	Cytoplasmic (94.1%)	Membrane topology: type 2 (cytoplasmic tail 1 to 137) ER membrane retention signals: KARR Leucine zipper pattern: LQWRLVQLRKLYWALDDFKASL at 38
S-bayanus	2	130-152 513-529	23 17	Cytoplasmic (76.7%)	Membrane topology: type 3a 2 <sup>nd</sup> peroxisomal targeting signal: KLMNDILHL at 89 Possible vacuolar targeting motif: ILPI at 100
S-castellii	1	127-143	17	Cytoplasmic (76.7%)	Membrane topology: type 2 (cytoplasmic tail 1 to 127) ER membrane retention signals: YKAL Di-leucine motif in the tail: LL at 41 and LL at 85

(Table 4.4 continued)

FALDH (organism)	No of TMSs	Position of TMD residues	Length	Cell localization (% Reliability)	Other features
S-cerevisiae	1	130-152	23	Cytoplasmic (89%)	Discrimination of nuclear localization signals: PRKK (4) at 41, KPRR (4) at 106 and PRKKDLE (5) at 41 2 <sup>nd</sup> peroxisomal targeting signal: KLMNDILHL at 89 Possible vacuolar targeting motif: ILPI at 100
S-kudriavzevii	2	140-156 513-529	17 17	Cytoplasmic (76.7%)	Membrane topology: type 2 (cytoplasmic tail 1 to 130) 2 <sup>nd</sup> peroxisomal targeting signal: KLMNDILHL at 89 Possible vacuolar targeting motif: ILPK at 100
S-mikatae	2	134-150 502-524	17 23	Cytoplasmic (89%)	Membrane topology: type 3a Nuclear localization signal PRKK at 34
S-paradoxus	1	130-146	17	Cytoplasmic (89%)	Membrane topology: type 3a 2 <sup>nd</sup> peroxisomal targeting signal: KLMNDILHL at 89 Possible vacuolar targeting motif: ILPK at 94 and ILPK at 100
U-maydis	0			Cytoplasmic (76.7%)	ER membrane retention signals: REWL
Y-lipolytica1	2	132-148 498-514	17 17	Cytoplasmic (94.1%)	Membrane topology: type 3b
Y-lipolytica2	1	501-520	20	Cytoplasmic (94.1%)	Membrane topology: type Nt (cytoplasmic tail 1 to 504) Possible vacuolar targeting motif: ILPI at 358 Leucine zipper pattern: LERIFGKLWFVIVLGLGAASL at 494
Y-lipolytica3	0			Cytoplasmic (89%)	No special features found
Y-lipolytica4	1	501-518	18	Cytoplasmic (94.1%)	Membrane topology: type Nt (cytoplasmic tail 1 to 502) ER membrane retention signals: AKAL Di-leucine motif in the tail: LL at 357, 453, 476
c-neoformans	0			Cytoplasmic (89%)	No special features found
D-hansenii	0			Cytoplasmic (94.1%)	ER membrane retention signal: L K Possible vacuolar targeting motif: KLPN at 34
C-glabrata	2	136-152 506-522	21 17	Cytoplasmic (89%)	Membrane topology: type 3a 2 <sup>nd</sup> peroxisomal targeting signal: KLLGDILHL at 85
K-lactis	2	122-138 497-519	17 23	Cytoplasmic (89%)	Membrane topology: type 3a Possible vacuolar targeting motif: ILPI at 364

From all these results, it is apparent that the fungal ALDHs obtained in this study fall into two groups, non-membrane ALDHs (no TMS) and tail-anchored ALDH proteins (presence of a TMS at termini). The tail-anchored ALDHs branch into both N- and C-tail anchored proteins as well as tail anchored ALDH proteins with a TMS at both termini. The presence of TMDs is an important feature as it has been shown that in the absence of dominant cytosolic or luminal targeting determinants, proteins may be sorted within the secretory pathway according to the interactions between their TMDs and the surrounding bilayer (Yang *et al.*, 1997). Miyauchi *et al.* (1991) also suggested that the C-terminal hydrophobic domain of rat microsomal aldehyde dehydrogenase (msALDH), which is also a FALDH, facilitates insertion of the protein into the membranes of the endoplasmic reticulum (ER). More studies of msALDH have now shown that the 35-amino acid carboxy-terminal of this protein

and probably other similar C-tail-anchored proteins possess ER-targeting sequences in addition to the hydrophobic transmembrane domain (Masaki *et al.*, 2003). Studies have indeed confirmed that in tail-anchored proteins the signals targeting the protein to the outer membrane are contained within the mature protein (Horie *et al.*, 2002).

## **CONCLUSION**

We have identified 28 FALDH/FALDH-like gene sequences of which nine are from molds and 19 from yeast species. Twenty-five of these proteins were identified phylogenetically as class 3 ALDHs with a unique feature in that they have relatively long N- and C-termini. Moreover we have found that though maintaining the normal ALDH motif patterns there is a tendency among these fungal sequences for preference of some residues over the others. Both ALDH active sites are present in all these putative FALDH/FALDH-like proteins and therefore all these genes are likely to code for enzymatically active ALDHs.

In both bacteria and yeast, FALDH enzyme activity has been detected during growth on alkanes and related compounds, where it participates in the carbon flow from *n*-alkanes to cell constituent synthesis and energy production through  $\beta$ -oxidation. Not all organisms in this study are confirmed alkane-assimilating microorganisms. However, these fungi are able to metabolize other high molecular weight hydrocarbons such as aromatic hydrocarbons and chlorinated aliphatics, which therefore leads us to suggest that these organisms do possess active genes encoding proteins required for metabolism of aldehyde intermediates of the hydrocarbon-assimilation pathway that may result from degradation of these hydrocarbons.

Unique features of the putative fungal FALDHs are the relatively long C- and N-termini, which probably distinguish fungal class 3 ALDHs as a group. These properties may indicate, as stated by Kirch *et al.* (2004), that the ALDH3 gene family members have evolved as a result of functional specialization in different

tissues and subcellular compartments. Moreover, these properties may indicate that of the class 3 ALDHs, fungal ALDHs have diverged much later in evolution with characteristics enabling them to adapt readily to new environments and new evolutionary niches (Hempel *et al.*, 1999).

## References

1. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17): 3389-402.
2. **Andrade AC, Del Sorbo G, Van Nistelrooy JG, Waard MA.** (2000) The ABC transporter AtrB from *Aspergillus nidulans* mediates resistance to all major classes of fungicides and some natural toxic compounds. *Microbiology*, 146 (Pt 8): 1987-97.
3. **Borgese N, Colombo S, Pedrazzini E.** (2003) The tail of tail-anchored proteins: coming from the cytosol and looking for a membrane. *The Journal of Cell Biology*, 161(6): 1013-1019.
4. **Bos P, De Bruyn JC.** (1973) The significance of hydrocarbon assimilation in yeast identification. *Antonie van Leeuwenhoek*, 39: 99-107.
5. **Chang C, Yoshida A.** (1997) Human fatty aldehyde dehydrogenase gene (ALDH10): organization and tissue-dependent expression. *Genomics*, 15; 40(1): 80-5.
6. **Chervitz SA, Hester ET, Ball CA, Dolinski K, Dwight SS, Harris MA, Juvik G, Malekian A, Roberts S, Roe T, Scafe C, Schroeder M, Sherlock G, Weng S, Zhu Y, Cherry JM, Botstein D.** (1999) Using the *Saccharomyces* Genome Database (SGD) for analysis of protein similarities and structure. *Nucleic Acids Research*, 27(1): 74–78.
7. **Cobessi D, Tête-Favier F, Marchal S, Azza S, Branlant G, Aubry A.** (1999) Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *Journal of Molecular Biology*, 290:161-173.

8. **Cooper DL, Isola NR, Stevenson K, Baptist EW.** (1993) Members of the ALDH gene family are lens and corneal crystallins. *Advances in Experimental Medicine and Biology*, 328: 169-179.
9. **Davydov VV, Dobaeva NM, Bozhkov AI.** (2004) Possible role of alteration of aldehyde's scavenger enzymes during aging. *Experimental Gerontology*, 39: 11-16.
10. **Demosay D, Rocchi S, Mas J-C, Grillo S, Pirola L, Chavey C, van Obberghen E.** (2004) Fatty aldehyde dehydrogenase: potential role in oxidative stress protection and regulation of its gene expression by insulin. *The Journal of Biological Chemistry*, 279(8): 6261-6270.
11. **Dolinski K, Balakrishnan R, Christie K, Costanzo M, Dwight S, Engel S, Fisk D, Hong E, Issel-Tarver L, Sethuraman A, Theesfeld C, Binkely G, Lane C, Schroeder M, Dong S, Kaloper M, Weng S, Andrada R, Botstein D and Cherry JM.** SGD - Saccharomyces Genome Database (<http://genome-www.stanford.edu/Saccharomyces/>) (12/8/2004).
12. **Eirich LD, Craft DL, Steinberg L, Asif A, Eschenfeldt WH, Stols L, Donnelly MI, Wilson CR.** (2004) Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain ATCC 20336. *Applied Environmental Microbiology*, 70(8): 4872-9.
13. **Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A.** (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13): 3784-8.
14. **Grindle M, Temple W.** (1982) Fungicide-resistance of os mutants of *Neurospora crassa*. *Neurospora Newsletter*, 29: 16-17.
15. **Grzybowska E, Migdalski A, Zagulski M.** (1993) Sequencing the genome of yeast *Saccharomyces cerevisiae*: Aims, strategies and results. *Postepy Biochemistry*, 39(3): 156-64.
16. **Hempel J, Liu ZJ, Perozich J, Rose J, Lindahl R, Wang BC.** (1997) Conserved residues in the aldehyde dehydrogenase family. Locations in the class 3 tertiary structure. *Advances in Experimental Medicine and Biology*, 414: 9-13.

- 17. Hempel J, Nicholas H, Lindahl R.** (1993) Aldehyde dehydrogenases: Widespread structural and functional diversity within a shared framework. *Protein Science*, 2: 1890-1900.
- 18. Hempel J, Perozich J, Chapman T, Rose J, Boesch JS, Liu ZJ, Lindahl R, Wang BC.** (1999) Aldehyde dehydrogenase catalytic mechanism. A proposal. *Advances in Experimental Medicine and Biology*, 463: 53-9.
- 19. Hempel J, Perozich J, Chapman T, Rose J, Boesch, JS, Liu Z-J, Lindahl R, Wang B-C.** (2001) Beyond the catalytic core of ALDH: a web of important residues begins to emerge. *Chemico-Biological Interactions*, 130-132:39-46.
- 20. Hempel J, Perozich J, Wymore T, Nicholas H.** (2003) Algorithm for identification and ranking of family specific residues, applied to the ALDH3 family. *Chemico-Biological Interactions*, 143-144:23-28.
- 21. Horie C, Suzuki H, Sakaguchi M, Mihara K.** (2002) Characterization of signal that directs C-tail-anchored proteins to mammalian mitochondrial outer membrane. *Molecular Biology of the Cell*, 13:1615-1625.
- 22. Horton P, Nakai K.** (1997) Better prediction of protein cellular localization sites with the *k* nearest neighbors classifier. *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, 5:147-52.
- 23. Kirch H-H, Nair A, Bartels D, Wei Y, Schnable PS, Wood AJ.** (2004) The ALDH gene superfamily of Arabidopsis. *Trends in Plant Science*, 9(8): 371-7.
- 24. Kirch H-H, Nair A, Bartels D.** (2001) Novel ABA- and dehydration-inducible aldehyde dehydrogenase genes isolated from the resurrection plant *Craterostigma plantagineum* and *Arabidopsis thaliana*. *The Plant Journal*, 28(5): 555-567.
- 25. Kotchoni OS, Bartels D.** (2003) Water stress induces the up-regulation of a specific set of genes in plants: Aldehyde dehydrogenases as an example. *Bulgarian Journal of Plant Physiology, Special Issue*: 37-51.
- 26. Kutzenko AS, Lamzin VS, Popov VO.** (1998) Conserved supersecondary structural motif in NAD-dependent dehydrogenases. *FEBS Letters*, 423: 105-109.

- 27. Lin Z, Carney G, Rizzo WB.** (2000) Genomic organization, expression, and alternate splicing of the mouse fatty aldehyde dehydrogenase gene. *Molecular Genetics and Metabolism*, 71(3): 496-505.
- 28. Liu ZJ, Hempel J, Sun J, Rose J, Hsiao D, Chang WR, Chung YJ, Kuo I, Lindahl R, Wang BC.** (1997) Crystal structure of a class 3 aldehyde dehydrogenase at 2.6 Å resolution. *Advances in Experimental Medicine and Biology*, 414:1-7.
- 29. Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang WR, Kuo I, Perozich J, Lindahl R, Hempel J, Wang BC.** (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Structural Biology*, 4(4): 317-26.
- 30. Markovetz AJ Jr, Cazin J, Allen JE.** (1968) Assimilation of alkanes and alkenes by fungi. *Applied Microbiology*, 16(3): 487-9.
- 31. Masaki R, Kameyama K, Yamamoto A.** (2003) Post-translational targeting of a tail-anchored green fluorescent protein to the endoplasmic reticulum. *Journal of Biochemistry (Tokyo)*, 134(3): 415-26.
- 32. Masaki R, Yamamoto A, Tashiro Y.** (1994) Microsomal aldehyde dehydrogenase is localized to the endoplasmic reticulum via its carboxy-terminal 35 amino acids. *The Journal of Cell Biology*, 126(6): 1407-1420.
- 33. Masaki R, Yamamoto A, Tashiro Y.** (1996) Membrane topology and retention of aldehyde dehydrogenase in the endoplasmic reticulum. *The Journal of Biological Chemistry*, 271(28): 16939-16944.
- 34. Mauersberger S, Ohkuma M, Schunck W-H, Takagi M.** (1996) *Candida Maltosa* In: *Non-conventional Yeasts in Biotechnology; A handbook*. K. Wolf, Springer-Verlag, Berlin, Germany, pp437-458.
- 35. Miyauchi K, Masaki R, Taketani S, Yamamoto A, Akayama M, Tashiro Y.** (1991) Molecular cloning, sequencing, and expression of cDNA for rat liver microsomal aldehyde dehydrogenase. *Journal of Biological Chemistry*, 278(5): 3489-3496.
- 36. Navarro-Avino JP, Prasad R, Miralles VJ, Benito RM, Serrano R.** (1999) A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces*

*cerevisiae* and characterization of the stress-inducible ALD2 and ALD3 genes. *Yeast*, 15: 829-842.

37. **Nebert DW, Nelson DR, Coon MJ, Estabrook RW, Feyereisen R, Fujii-Kuriyama Y, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, et al.** (1991) The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA and Cellular Biology*; 10(1): 1-14.
38. **Nebert DW.** (2000) Suggestions for the nomenclature of human alleles: relevance to ecogenetics, pharmacogenetics and molecular epidemiology. *Pharmacogenetics*; 10(4): 279-90
39. **Nebert DW.** Current status of the IUBMB-recommended nomenclature program for several gene superfamilies (<http://www.uchsc.edu/sp/sp/alcdbase/aldhcov.html>) (10/5/2005).
40. **Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW.** (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, 6(1): 1-42.
41. **Nyns EJ, Auquiere JP, Wiaux AL.** (1968) Taxonomic value of the property of fungi to assimilate hydrocarbons. *Antonie Van Leeuwenhoek*, 34(4): 441-5.
42. **Patterton HG, Graves S.** (2000) DNAssist: the integrated editing and analysis of molecular biology sequences in windows. *Bioinformatics*, 16(7): 652-2.
43. **Perozich J, Hempel J, Morris SM Jr.** (1998) Roles of conserved residues in the arginase family. *Biochimica et Biophysica Acta*, 15; 1382(1): 23-37.
44. **Perozich J, Nicholas H, Lindahl R, Hempel J.** (1999) The big book of aldehyde dehydrogenase sequences. An overview of the extended family. *Advances in Experimental Medicine and Biology*, 463:1-7. Review.
45. **Perozich J, Nicholas H, Wang BC, Lindahl R, Hempel J.** (1999) Relationships within the aldehyde dehydrogenase extended family. *Protein Science*, 8(1): 137-46.
46. **Ramesh MA, Laidlaw RD, Durrenburger F, Orth AB, Kronstad JW.** (2001) The cAMP signal transduction pathway mediates resistance to dicarboximide

and aromatic hydrocarbon fungicides in *Ustilago maydis*. *Fungal Genetics and Biology*, 32: 183-193.

47. **Rayner JC, Pelham HRB.** (1997) Transmembrane domain-dependent sorting of proteins to the ER and plasma membrane in yeast. *The EMBO Journal*, 16(8): 1832-1841.
48. **Reggiori F, Black MW, Pelham HRB.** (2000) Polar transmembrane domains target proteins to the interior of the yeast vacuole. *Molecular Biology of the Cell*, 11:3737-3749.
49. **Rizzo WB, Lin Z, Carney G.** (2001) Fatty aldehyde dehydrogenase: structure, expression and mutation analysis in Sjögren-Larsson syndrome. *Chemico-Biological Interactions*, 130-132:297-307.
50. **Rodriguez-Zavala J, Weiner H.** (2001) Role of the C-terminal tail on the quaternary structure of aldehyde dehydrogenases. *Chemico-Biological Interactions*, 130-132:151-160.
51. **Rogers GR, Markova NG, De Laurenzi V, Rizzo WB, Compton JG.** (1997) Genomic organization and expression of the human fatty aldehyde dehydrogenase gene (FALDH). *Genomics*, 15; 39(2): 127-35.
52. **Roland JG.** (2002) Selected Abstracts and Bibliography of International Oil Spill Research. Prepared by the Louisiana Oil Spill Coordinator, Office of the Governor, Baton Rouge: Louisiana Applied Oil Spill Research and Development Program, 1996. pp. 727 (<http://www.osradp.lsu.edu/Bibliographie.html>) (3/9/2005).
53. **Rost B.** (2002) Enzyme function less conserved than anticipated. *Journal of Molecular Biology*, 318: 595-608.
54. **Schultz J, Milpetz F, Bork P, Ponting CP.** (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of U S A.*, 95(11): 5857-64.
55. **Shibaki A, Akiyama M, Shimizu H.** (2004) Novel ALDH3A2 heterozygous mutations are associated with defective lamellar granule formation in a Japanese family of Sjogren-Larsson Syndrome. *The Journal of Investigative Dermatology*, 123: 1197-1199.

- 56. Singer ME, Finnerty WR.** (1985) Fatty aldehyde dehydrogenases in *Acinetobacter* sp. strain HO1-N: role in hexadecanol metabolism. *Journal of Bacteriology*, 164(3): 1011-6.
- 57. Skibbe DS, Liu F, Wen T-J, Yandeu MD, Cui X, Cao J, Simmons CR, Schnable PS.** (2002) Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and *Arabidopsis*. *Plant Molecular Biology*, 48: 751-764.
- 58. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R.** (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Research*, 26(1): 320-2.
- 59. Sophos NA, Pappa A, Ziegler TL, Vasiliou V.** (2001) Aldehyde dehydrogenase gene superfamily: the 2000 update. *Chemico-Biological Interactions*, 130-132(1-3): 323-37.
- 60. Sophos NA, Vasiliou V.** (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chemico-Biological Interactions*, 143-144: 5-22.
- 61. Steel GJ, Brownsword J, Stirling CJ** (2002) Tail-anchored protein insertion into yeast ER requires a novel posttranslational mechanism that is independent of the SEC machinery. *Biochemistry*, 41: 11914-11920.
- 62. Steinmetz CG, Xie P, Weiner H, Hurley TD.** (1997) Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure*, 5(5): 701-11.
- 63. Sutherland JB.** (1992) Detoxification of polycyclic aromatic hydrocarbons by fungi. *Journal of Industrial Microbiology*, 9(1): 53-61.
- 64. Tasayco ML, Prestwich GD.** (1990) A specific affinity reagent to distinguish aldehyde dehydrogenases and oxidases. Enzymes catalyzing aldehyde oxidation in an adult moth. *Journal of Biological Chemistry*, 265(6): 3094-101.
- 65. Thompson JD, Higgins DG, Gibson TJ.** (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.
- 66. Ueda M, Tanaka A.** (1990) A Long-chain aldehyde dehydrogenase of *Candida* yeast. *Methods in Enzymology*, 188:176-8.

- 67. Vanhanen S, West M, Kroon JT, Lindner N, Casey J, Cheng Q, Elborough KM, Slabas AR.** (2000) A consensus sequence for long-chain fatty-acid alcohol oxidases from *Candida* identifies a family of genes involved in lipid omega-oxidation in yeast with homologues in plants and bacteria. *Journal of Biological Chemistry*, 275(6): 4445-52.
- 68. Vasiliou V, Bairoch A, Tipton KF, Nebert DW.** (1999) Eukaryotic aldehyde dehydrogenase (ALDH) genes: human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping. *Pharmacogenetics*, 9(4): 421-34.
- 69. Vasiliou V, Kozak CA, Lindahl R, Nebert DW.** (1996) Mouse microsomal Class 3 aldehyde dehydrogenase: AHD3 cDNA sequence, inducibility by dioxin and clofibrate, and genetic mapping. *DNA and Cell Biology*, 15(3): 235-45.
- 70. Vasiliou V.** (1997) Aldehyde dehydrogenase genes. *Advances in Experimental Medicine and Biology*, 414:595-600.
- 71. Watkinson RJ, Morgan P.** (1990) Physiology of aliphatic hydrocarbon-degrading microorganisms. *Biodegradation*, 1(2-3): 79-92. Review.
- 72. Weiner H, Flynn TG.** (1989) Nomenclature of mammalian aldehyde dehydrogenases, p. xix-xxi. *In* H. Weiner, and T. G. Flynn (ed.), *Progress in clinical and biological research*, vol. 290. Alan R. Liss, Inc., New York, N.Y.
- 73. White MJ, Hammond RC, Rose AH.** (1987) Production of long chain alcohols by yeasts. *Journal of general Microbiology*, 133: 2181-2190.
- 74. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, etc.** (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 415(6874): 871-80.
- 75. Yabal M, Brambillasca S, Soffientini P, Pedrazzini E, Borgese N, Makarow M.** (2003) Translocation of the C terminus of a tail-anchored protein across the endoplasmic reticulum membrane in yeast mutants defective in signal peptide-driven translocation. *Journal of Biological Chemistry*, 278(5): 3489-3496.
- 76. Yamada T, Nawa H, Kawamoto S, Tanaka A, Fukui S.** (1980) Subcellular localization of long-chain alcohol dehydrogenase and aldehyde dehydrogenase in alkane-grown *Candida tropicalis*. *Archives of Microbiology*, 128:145-151.

- 77. Yang AS, Honig B.** (2000) An integrated approach to the analysis and modeling of protein sequences and structures. II. On the relationship between sequence and structural similarity for proteins that are not obviously related in sequence. *Journal of Molecular Biology*, 301: 679-689.
- 78. Yang M, Ellenberg J, Bonifacino JS, Weissman AM.** (1997) The transmembrane domain of a carboxyl-terminal anchored protein determines localization to the endoplasmic reticulum. *The Journal of Biological Chemistry*, 272(3): 19770-19775.
- 79. Yoshida A, Rzhetsky A, Hsu LC, Chang C.** (1998) Human aldehyde dehydrogenase gene family. *European Journal of Biochemistry*, 251:549-557
- 80. Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E.** (2001) Differences in nucleotide specificity and catalytic mechanism between *Vibrio harveyi* aldehyde dehydrogenase and other members of the aldehyde dehydrogenase superfamily. *Chemico-Biological Interactions*, 130-132(1-3): 29-38.
- 81. Zhou J, Weiner H.** (2001) The N-terminal portion of mature aldehyde dehydrogenase affects protein folding and assembly. *Protein Science*, 10: 1490-1497.
- 82. Zinovieva RD, Tomarev SI, Piatigorsky J.** (1993) Aldehyde dehydrogenase-derived omega-crystallins of squid and octopus: Specialization for lens expression. *Journal of Biological Chemistry*, 268(15): 11449-11455.

## CHAPTER 5

### GENERAL DISCUSSION AND CONCLUSIONS

Studies intended to elucidate the biochemistry and genetics of the alkane-degradation pathway have mainly focused on cytochrome P450 monooxygenases and the  $\beta$ -oxidation system. Consequently, information on the other enzymes involved in alkane-utilization pathway is limited. Moreover, although fatty alcohol dehydrogenase (FADH) and fatty aldehyde dehydrogenase (FALDH) activity have been demonstrated in yeasts growing on alkanes, only fatty alcohol oxidases (FAOD) have been purified from these yeasts (Dickinson and Wadforth, 1992; I'chenko *et al.*, 1994; Hommel *et al.*, 1994), while no FADH or FALDH has yet been purified to homogeneity. Moreover enzyme activity of reported FADH and FALDH encoding genes has not been proven. On the other hand two FAOD encoding genes in *C. cloacae* and four in *C. tropicalis* have been cloned and sequenced (Vanhanen *et al.*, 2000; Eirich *et al.*, 2004) and FAOD enzyme activity during growth on alkanes has also been confirmed in these yeasts. The question of involvement of FAODs, FADHs and FALDHs in alkane utilization is further complicated by reports that the P450 monooxygenase system is able to solely carry out complete oxidation of *n*-alkanes to the corresponding fatty acids (Scheller *et al.*, 1998). Therefore, this study initially aimed at clarifying the involvement of FAOD, FADH and FALDH in alkane degradation by the yeast *Y. lipolytica* using enzyme assays, RT-PCR and Northern hybridization analysis. Following this study was the disruption of the FALDH genes with the aim to create *Y. lipolytica* strains capable of accumulating fatty alcohols.

Activity assays using decanal and dodecanol as substrates only detected significant induction of FALDH activity, whereas FADH and FAOD activities were very low during growth on hexadecane. The FALDH activity (~1.4 U/mg protein) was also low compared to the activity reported by Yamada *et al.* (1980) (98 U/mg protein). FADH (~1 U/mg protein) and FAOD activities (~0.1 U/mg protein) were

also very low compared to literature values for FADH (~80 U/mg; Yamada *et al.*, 1980) and FAOD (~20 U/mg; Il'chenko *et al.*, 1994). However, the enzyme assays have proven coexistence of FADH and FAOD enzymes in this strain of *Y. lipolytica*, as was suggested in literature (Barth and Gaillardin, 1996). The very low levels of both enzymes made it difficult to establish as to which of the two enzymes is mainly responsible for fatty alcohol oxidation during growth on alkanes. The FAOD enzyme has previously been purified (Ilchenko *et al.*, 1994) and shown to be highly photo- and thermo-labile (Kemp *et al.*, 1990). We believe that it was due to this sensitivity that crude cell-free extracts initially having some FAOD activity could sustain this activity for only a very short time (<30 min) even though the samples were kept on ice and protected from light. Kemp *et al.* (1994) also reported detection of no FAOD activity in two strains of *Y. lipolytica*, which they also attributed to the photolability of this enzyme. It is thus clear that for any experimental work using this enzyme to succeed the issue of heat and light sensitivity must first be dealt with. Since no issues could be found leading to error in determination of FADH enzyme activity it was thus concluded that the observed low FADH levels showed that this enzyme was not induced by growth in hexadecane. Moreover, similar FADH levels were obtained for cells grown in glucose and glycerol, which are known not to induce this enzyme. On the contrary, the low FAOD levels already mentioned were only observed in hexadecane-grown cells and not in glucose- or glycerol- grown cells, which leads one to believe that FAOD is induced by growth of *Y. lipolytica* in alkane and is thus responsible for fatty alcohol oxidation in this yeast. The constitutive expression of FADH as it seemed may be required under conditions where FAOD is not expressed.

Differential centrifugation experiments showed that both FALDH and FADH activities were present in the peroxisomal/mitochondrial fraction as well as in the fraction that contained microsomal and soluble enzymes. An attempt to separate microsomal and soluble enzymes through the use of Y-PER reagent (Pierce Chemical Co., US), which should release soluble proteins from the cells while maintaining membrane associated enzymes, failed. In fact reduced levels of both

enzymes in presence of the Y-PER reagent suggest that this chemical may be inhibitory on the proteins or interfering with the enzyme assay.

BLAST searches of the *Y. lipolytica* genome database initially showed the presence of two putative FALDH encoding genes and one FADH encoding gene. After sequencing and annotation of the *Y. lipolytica* genome was complete (Dujon *et al.*, 2004), the number of putative FALDH encoding genes increased to four. No FAOD encoding gene was found in the sequenced *Y. lipolytica* genome. The FAOD gene(s) could not be detected by PCR either. However, our Southern hybridization experiments using the FAOD gene from *C. tropicalis* as a probe indicated presence of several FAOD homologues in the genome of *Y. lipolytica*. Presence of several FAOD enzyme activities in *Y. lipolytica* has previously been reported (Ilchenko *et al.*, 1994), thus implying presence of several FAOD isogenes in this yeast. It is for this reason that one would believe that the Southern hybridization experiments were correct. Detection of FAOD gene(s) in *Y. lipolytica* genome through BLAST search and PCR were based on known FAOD gene sequences from other yeasts (Vanhanen *et al.*, 2000; Eirich *et al.*, 2004). As already stated in literature phylogenetic trees constructed using known well-conserved gene sequences showed that *Y. lipolytica* is different from other known filamentous yeasts in many ways (Barth and Gaillardin, 1996). Genetic and molecular biology studies have also revealed many features of this yeast not observed in other known yeasts (Barth and Gaillardin, 1997; Muller *et al.*, 1998). It is thus quite likely that even though FAOD gene(s) may be present in *Y. lipolytica* it might escape detection due their uncommon sequence(s). The possibility of novel FAOD gene(s) in *Y. lipolytica* is therefore strongly suggested by these findings. Alternatively it is possible that there are no FAOD gene present in *Y. lipolytica* E150 (CLIB 99), which was sequenced, while it might be present in *Y. lipolytica* H222, which was used for the FAOD enzyme determinations (by previous authors and in this work) and for the Southern hybridization experiments.

Northern blot analysis and RT-PCR, to determine induction of the two aldehyde dehydrogenases and two fatty aldehyde dehydrogenases, indicated that one of the aldehyde dehydrogenases, labelled ALDH1, was constitutively expressed during growth on glucose and alkanes, while one of the fatty aldehyde dehydrogenases, labelled FALDH4 was only induced during growth on alkanes. Northern blot analysis showed that the FADH was weakly induced under all conditions.

In order to establish which of the four FALDH encoding genes coded for FALDHs that are necessary for alkane degradation, the four genes were deleted in all possible combinations using the Cre-Lox P recyclable tools, which enable marker rescue allowing simultaneous deletion of several members of a gene family by use of a recyclable single marker (Fickers *et al.*, 2003). Time did not allow thorough phenotypic analysis of the fifteen mutant strains that were created in this way. However, the triple deleted mutants (each with only one functional *FALDH* gene) and the quadruple deleted mutant (with all four *FALDH* genes deleted) were tested for growth on alkanes. Results from these growth studies indicated that all these strains could still grow on alkanes with chain lengths ranging between C10 and C18. However growth of one of the triple deletion strains (with only *FALDH2* intact) and the quadruple deleted strain was slightly reduced. Efforts to do FALDH activity assays were unsuccessful, because time did not allow optimization of the growth and assay conditions. The strain used for the deletions (Po1d) was unfortunately different from the one used for the initial enzyme assays (H222), because attempts to use the Cre-Lox P system on H222 had been unsuccessful. However, cell harvesting by centrifugation and protein assays on the same amounts of wet biomass indicated that the cells of the FALDH deficient mutants were more hydrophobic than the cells of the wild type strain and contained less protein. This was an interesting observation, which should be investigated for all the FALDH deletion mutants. *Y. lipolytica* mutants with *POX* genes coding for acyl-coenzyme A oxidases deleted also displayed changes in lipid content and cell surface properties (Mlícková *et al.* 2004)

The fact that despite loss of the four *FALDH* genes the organism was still able to grow on *n*-alkanes suggests presence of some other mechanism of alkanol oxidation. In fact presence of other ALDH genes (at least nine more) in the genome of *Y. lipolytica* has been reported (NCBI and UniProt databases). However at present enzyme activities and the capability to oxidize fatty aldehydes for all these genes have not yet been determined. Moreover, the P450 monooxygenase system has been shown to be capable to solely carry out complete oxidation of *n*-alkanes through to the corresponding fatty acids (Scheller *et al.*, 1998). This therefore suggests that disruption of all existing ALDH genes may not facilitate accumulation of fatty alcohols since the P450 monooxygenase effect would enable bypass of the FALDH activity route. In humans FALDHs function, among other things, in detoxification and synthesis of secondary metabolites (Demosay *et al.*, 2004; Davydov *et al.*, 2004). So it is possible that the putative FALDHs identified in the genome of *Y. lipolytica*, which seem not to play any role in alkane metabolism, may play the same role in this yeast. In fact, three alcohol dehydrogenase genes in this yeast have been indicated as encoding enzymes necessary for an ethanol-acetaldehyde shuttle (Kerscher *et al.*, 2001)

BLAST searches for the FALDH encoding genes in *Y. lipolytica* brought to our attention the presence of similar sequences in other fungal genomes. These sequences had not yet been described in literature or formally classified. No activity information is also available about their gene products. Through BLAST searches using the human FALDH sequence as query we subsequently identified 28 FALDH/FALDH-like gene sequences of which nine are from molds and 19 from yeast species. These fungal proteins, which fall into at least two subfamilies, show approximately 24-90% identity amongst themselves, and 26-44% similarity with the human FALDH protein. The proteins are potentially active class 3 ALDHs with long C- and N-termini, and most are expected to be integral membrane proteins bearing 1-3 transmembrane domains.

Not all organisms identified in this study are alkane-assimilating microorganisms. However, many of the non-alkane degrading microorganisms are able to metabolize aromatic hydrocarbons and chlorinated aliphatics (Sutherland, 1992; Barth and Gaillardin, 1996), which therefore leads us to suggest that these organisms do possess active genes encoding proteins required for metabolism of intermediates of the hydrocarbon-assimilation pathway that may result from degradation of these hydrocarbons. The unique features described for these fungal protein sequences leads one to believe that perhaps this is what distinguishes fungal class 3 ALDH as a group. Perhaps also as suggested by other authors, these properties may indicate that the ALDH3 gene family members have evolved as a result of functional specialization in different tissues and subcellular compartments (Kirch *et al.*, 2004) and this group has also diverged much later in evolution with characteristics enabling them to adapt readily to new environments and new evolutionary niches (Hempel *et al.*, 1999).

Finally results from the present study indicated: (i) induction of the one FALDH encoding gene (*FALDH4*) during growth on alkanes (ii) that the four putative FALDHs in *Y. lipolytica* together play a role in alkane utilization but (iii) that disruption of all four putative *FALDH* genes did not abolish growth on alkanes. It is therefore probable that many of the fungal class 3 ALDH encoding genes, which we uncovered during this study, encode functional FALDHs, but the gene products are not necessarily involved in alkane utilization. Cloning of the putative FALDHs in a suitable host, such as the FALDH quadruple deletion strain of *Y. lipolytica* created during this study, might eventually indicate which of these genes really encode functional FALDHs.

## References

1. **Barth G, Gaillardin C.** (1996) *Yarrowia lipolytica*. In: K. Wolf, ed. Non-conventional Yeasts in Biotechnology. Berlin, Germany: Springer Verlag, pp 313-388.
2. **Barth G, Gaillardin C.** (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiology Reviews, 19(4): 219-37.
3. **Dickinson FM, Wadforth C.** (1992) Purification and some properties of alcohol oxidase from alkane-grown *Candida tropicalis*. Biochemical Journal, 282 ( Pt 2):325-31.
4. **Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, etc.** (2004) Genome evolution in yeasts. Nature, 430(6995): 35-44.
5. **Eirich LD, Craft DL, Steinberg L, Asif A, Eschenfeldt WH, Stols L, Donnelly MI, Wilson CR.** (2004) Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain ATCC 20336. Applied Environmental Microbiology, 70(8): 4872-9.
6. **Fickers P, Le Dall MT, Gaillardin C, Thornart P, Nicaud J-M.** (2003) New disruption cassettes for gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. Journal of Microbiological Methods, 55: 727-737.
7. **Hempel J, Perozich J, Chapman T, Rose J, Boesch JS, Liu ZJ, Lindahl R, Wang BC.** (1999) Aldehyde dehydrogenase catalytic mechanism. A proposal. Advances in Experimental Medicine and Biology, 463: 53-9.
8. **Hommel RK, Lassner D, Weiss J, Kleber H-P.** (1994) The inducible microsomal fatty alcohol oxidase from alkane-grown *Candida tropicalis* and *Yarrowia lipolytica*. Applied Microbiology and Biotechnology, 40: 729-734.
9. **Il'chenko AP, Morgunov IG, Honeck H, Mauersberger S, Vasil'kova NN, Müller H-G** (1994) Purification and properties of alcohol oxidase from the yeast *Yarrowia lipolytica* H-222. Biochemistry (Moscow), 59(9): 969-974.
10. **Kemp GD, Dickinson FM, Ratledge C.** (1990) Inducible long chain alcohol oxidase from alkane-grown *Candida tropicalis* and *Yarrowia lipolytica*. Applied Microbiology and Biotechnology, 32: 461-464.

11. **Kemp GD, Dickinson FM, Ratledge C.** (1994) Occurrence of fatty alcohol oxidase in fatty-acid-utilizing yeasts and molds. *Applied Microbiology and Biotechnology*, 40: 873-875.
12. **Kerscher SJ, Eschemann A, Okun PM, Brandt U.** (2001) External alternative NADH:ubiquinone oxidoreductase redirected to the internal face of the mitochondrial inner membrane rescues complex I deficiency in *Yarrowia lipolytica*. *Journal of Cell Science*, 114: 3915-3921
13. **Kirch H-H, Nair A, Bartels D, Wei Y, Schnable PS, Wood AJ.** (2004) The ALDH gene superfamily of Arabidopsis. *Trends in Plant Science*, 9(8): 371-7.
14. **Mlickova K, Roux E, Athenstaedt K, d'Andrea S, Daum G, Chardot T, Nicaud JM.** (2004) Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast *Yarrowia lipolytica*. *Applied and Environmental Microbiology*, 70(7):3918-24.
15. **Muller S, Sandal T, Kamp-Hansen P, 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(14):1267-83.
16. **Scheller U, Zimmer T, Becher D, Schauer F, Schunck WH.** (1998) Oxygenation cascade in conversion of *n*-alkanes to alpha, omega-dioic acids catalyzed by cytochrome P450 52A2. *Journal of Biological Chemistry*, 273(49):32528-34.
17. **Sutherland JB.** (1992) Detoxification of polycyclic aromatic hydrocarbons by fungi. *Journal of Industrial Microbiology*, 9(1): 53-61.
18. **Vanhanen S, West M, Kroon JT, Lindner N, Casey J, Cheng Q, Elborough KM, Slabas AR.** (2000) A consensus sequence for long-chain fatty-acid alcohol oxidases from *Candida* identifies a family of genes involved in lipid omega-oxidation in yeast with homologues in plants and bacteria. *Journal of Biological Chemistry*, 275(6): 4445-52.

- 19. Yamada T, Nawa H, Kawamoto S, Tanaka A, Fukui S.** (1980) Subcellular localization of long-chain alcohol dehydrogenase and aldehyde dehydrogenase inn-alkane-grown *Candida tropicalis*. Archives of Microbiology, 128:145-151.

## SUMMARY

**Key words: *Yarrowia lipolytica*, alkanes, fatty alcohol dehydrogenase, fatty aldehyde dehydrogenase, Southern blot, northern blot, RT-PCR, gene disruptions, Cre-loxP, fungal FALDHs**

The cytochrome P-450 monooxygenase and  $\beta$ -oxidation systems of alkane-utilizing yeasts have been studied extensively, whereas very little is known about the fatty (long chain) alcohol and fatty aldehyde oxidizing enzymes. With the recent completion of sequencing of the genome of *Yarrowia lipolytica*, an alkane-degrading yeast, several putative aldehyde dehydrogenases (ALDHs) have been identified. Four of these were identified as fatty ALDHs (FALDHs). Northern blot analysis and RT-PCR showed that one of the *FALDH* genes, labelled *FALDH4*, is induced during growth of *Y. lipolytica* on alkanes, whereas another aldehyde dehydrogenase gene, labelled *ALDH1*, was constitutively expressed. Functional analysis of the four *FALDH* isogenes was initiated by single gene deletion of the four fatty aldehyde dehydrogenase isogenes in all possible combinations. The Cre-loxP recyclable tools system was used for gene disruption.

Growth properties of the triple and quadruple deletion strains on alkanes were investigated. A slightly arrested growth in hexadecane was observed in two strains, the triple deletion mutant with intact *FALDH2* isogene and the quadruple deletion mutant with all four *FALDH* isogenes deleted. Very strong hydrophobicity during growth of these mutants in hexadecane was also observed. At this stage one can only say that disruption of FALDH isogenes had a slight negative effect on growth of this yeast on alkanes.; However, it is not yet clear which individual isogenes are the most important for alkane metabolism in this organism.

Although fatty aldehyde dehydrogenase (FALDH) activity has been detected in fungi no FALDH genes have yet been cloned, sequenced and expressed. Through BLAST searches using the human FALDH sequence as query we have identified

28 FALDH/FALDH-like gene sequences of which nine are from molds and 19 from yeast species. A comparative study of these sequences showed that fungal FALDH sequences may fall into several different subclasses of the ALDH3 family. Unique features of these proteins included presence of several transmembrane domains and in particular relatively long C- and N-termini.

Searches of the sequenced *Y. lipolytica* genome for fatty alcohol oxidase (FAOD) and fatty alcohol dehydrogenase (FADH) encoding genes, which could be involved in the oxidation of fatty alcohols to aldehydes, yielded only one putative FADH encoding gene. However, FADH activity during growth on n-alkanes was very low and Northern-blot analyses showed that this gene was only weakly expressed during growth on hydrocarbon and non-hydrocarbon substrates.

## APPENDIX A

ID	Accession No.	Organism
At_ALDH3F1	CAE48163.1	<i>Arabidopsis thaliana</i>
At_ALDH3H1	CAE51203.1	<i>Arabidopsis thaliana</i>
At_ALDH3I1	CAC84903.1	<i>Arabidopsis thaliana</i>
BENZ_ACI	Q6FCB6	<i>Acinetobacter sp.</i> (strain ADP1)
BENZ_CAU	Q9A5Q0	<i>Caulobacter crescentus</i>
BENZ_HAL	Q9HMJ6	<i>Halobacterium sp.</i> (NRC-1)
BENZ_PSE	P43503	<i>Pseudomonas putida</i>
BENZ_RAL	Q8XT86	<i>Ralstonia solanacearum</i>
BENZ_STR	Q9L124	<i>Streptomyces coelicolor</i>
BENZ_XAN	Q8PDI4	<i>Xanthomonas campestris (pv. campestris)</i>
Ce_ALDH3C1_1	Q60WY8	<i>Caenorhabditis briggsae</i>
Ce_ALDH3C1_2	O16518	<i>Caenorhabditis elegans</i>
Ce_ALDH3C1_3	Q86S57	<i>Caenorhabditis elegans</i>
Ce_ALDH3C2	Q8MXJ7	<i>Caenorhabditis elegans</i>
DHA1_BOVIN	P48644	<i>Bos taurus</i>
DHA1_CHICK	P27463	<i>Gallus gallus</i>
DHA1_HUMAN	P00352	<i>Homo sapiens</i>
DHA1_MACFA	Q8HYE4	<i>Macaca fascicularis</i>
DHA1_MOUSE	P24549	<i>Mus musculus</i>
DHA1_RABIT	Q8MI17	<i>Oryctolagus cuniculus</i>
DHA1_SHEEP	P51977	<i>Ovis aries</i>
DHA2_HUMAN	O94788	<i>Homo sapiens</i>
DHA2_MOUSE	Q62148	<i>Mus musculus</i>
DHA2_RAT	Q63639	<i>Rattus norvegicus</i>
DHA5_HUMAN	P30837	<i>Homo sapiens</i>
DHAM_BOVIN	P20000	<i>Bos taurus</i>
DHAM_HORSE	P12762	<i>Equus caballus</i>

APPENDIX A (continued)

<b>ID</b>	<b>Accession No.</b>	<b>Organism</b>
DHAM_HUMAN	P05091	<i>Homo sapiens</i>
DHAM_MOUSE	P47738	<i>Mus musculus</i>
DHAM_RAT	P11884	<i>Rattus norvegicus</i>
DHAS_CHICK	O93344	<i>Gallus gallus</i>
Dr_ALDH3D1_1	NP_775328.2	<i>Danio rerio</i>
Eh_ALDH13	AAA19741.1	<i>Entamoeba histolytica</i>
FUNG_ASH	Q758W1	<i>Ashbya gossypii</i>
FUNG_CAN	Q6FVP8	<i>Candida glabrata</i> CBS138
FUNG_CAN2	Q6FPK0	<i>Candida glabrata</i> CBS138
FUNG_DEB	Q6BJB3	<i>Debaryomyces hansenii</i> CBS767
FUNG_KLU	Q6CLU0	<i>Kluyveromyces lactis</i> NRRL Y-1140
FUNG_NEU	Q8X0L4	<i>Neurospora crassa</i>
FUNG_PIC	Q12648	<i>Pichia angusta</i>
FUNG_SAC	P46367	<i>Saccharomyces cerevisiae</i>
FUNG_YAR	Q6CD79	<i>Yarrowia lipolytica</i> CLIB99
GAPDH_ARA1	P25857	<i>Arabidopsis thaliana</i>
GAPDH_CAP	Q8VWP2	<i>Capsicum annuum</i>
GAPDH_NIC	P09044	<i>Nicotiana tabacum</i>
GAPDH_ORY	Q7X8A1	<i>Oryza sativa</i> (japonica cultivar-group)
GAPDH_PIN	P12859	<i>Pisum sativum</i>
GAPDH_PIS	Q41019	<i>Pinus sylvestris</i>
GAPDH_SPI1	P12860	<i>Spinacia oleracea</i>
GAPDH_ZEA	Q6LBU9	<i>Zea mays</i>
HOM_SAPL	P51648	<i>Homo sapiens</i>
Hs_ALDH3A1	Q6PKA6	<i>Homo sapiens</i>
Hs_ALDH3A2	P51648	<i>Homo sapiens</i>
Hs_ALDH3B1	Q8N515	<i>Homo sapiens</i>
Hs_ALDH3B2	P43353	<i>Homo sapiens</i>
Hs_ALDH3B2	P48448	<i>Homo sapiens</i>

APPENDIX A (continued)

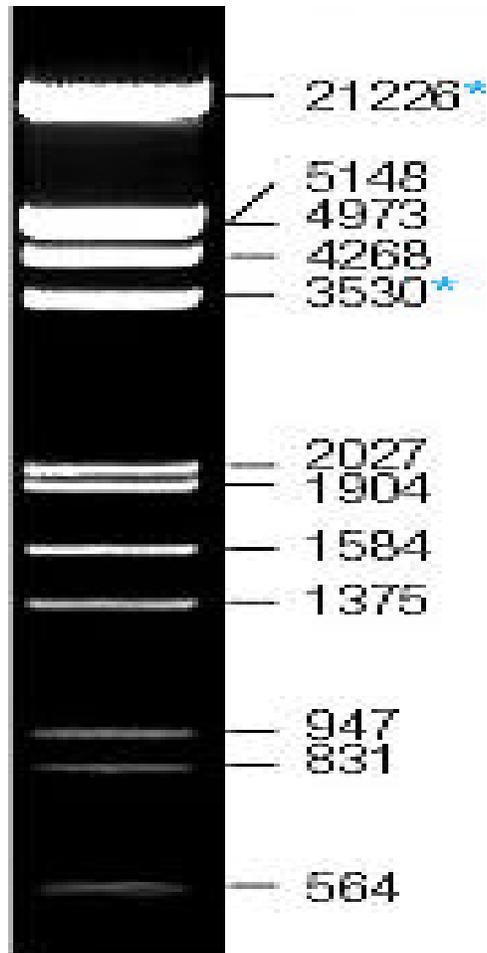
ID	Accession No.	Organism
Mm_ALDH3A1	P47739	<i>Mus musculus</i>
Mm_ALDH3A2	P47740	<i>Mus musculus</i>
Mm_ALDH3B1	Q8VHW0	<i>Mus musculus</i>
MmALDH3A2	Q99L64	<i>Mus musculus</i>
MUS_MUSL	AAK01551.1	<i>Mus musculus</i>
Po_alkH	P12693	<i>Pseudomonas oleovorans</i>
Pp_alkH	CAB51050.1	<i>Pseudomonas putida</i>
Rn_ALDH3A1	P11883	<i>Rattus norvegicus</i>
Rn_ALDH3A2	P30839	<i>Rattus norvegicus</i>
Rn_ALDH3A2	P30839	<i>Rattus norvegicus</i>
SSDH_ACI	Q6F9G0	<i>Acinetobacter sp.</i> (strain ADP1)
SSDH_BUR2	Q62B48	<i>Burkholderia mallei</i> ATCC 23344
SSDH_BURK1	Q63NL9	<i>Burkholderia pseudomallei</i>
SSDH_ESC	Q8X950	<i>Escherichia coli</i> O157:H7
SSDH_HOM	Q8N3W6	<i>Homo sapiens</i>
SSDH_NEU	Q7SFB1	<i>Neurospora crassa</i>
SSDH_PSE1	Q9I6M5	<i>Pseudomonas aeruginosa</i>
SSDH_PSE2	Q88AT6	<i>Pseudomonas syringae</i> (pv. tomato)
SSDH_RAT	P51650	<i>Rattus norvegicus</i>
SSDH_YAR	Q6C0B4	<i>Yarrowia lipolytica</i> CLIB99
ASP_NID1	EAA62737.1	<i>Aspergillus nidulans</i> FGSC A4
ASP_NID2	EAA64317.1	<i>Aspergillus nidulans</i> FGSC A4
CAN_ALB1	EAL00022.1	<i>Candida albicans</i> SC5314
CAN_ALB2	EAL02020.1	<i>Candida albicans</i> SC5314
CAN_GLA	CAG61338.1	<i>Candida glabrata</i> strain CBS138
CRY_NEO	EAL17607.1	<i>Cryptococcus neoformans</i> var. <i>neoformans</i> B-3501A
DEB_HAN	XP_461954.1	<i>Debaryomyces hansenii</i> CBS767
GIB_ZEA	EAA67698.1	<i>Gibberella zeae</i> PH-1

APPENDIX A (continued)

<b>ID</b>	<b>Accession No.</b>	<b>Organism</b>
KLU_LAC	CR382122.1	<i>Kluyveromyces lactis</i> NRRLY-1140
KLU_WAL	AADM01000092.1	<i>Kluyveromyces waltii</i> NCYC2644
MAG_GRI1	EAA49061.1	<i>Magnaporthe grisea</i> 70-15
MAG_GRI2	EAA53613.1	<i>Magnaporthe grisea</i> 70-15
MAG_GRI3	CAG90160	<i>Debaryomyces hansenii</i> CBS767
NEU_CRA	EAA28392.1	<i>Neurospora crassa</i> OR74A
SAC_BAY	AACA01000400.1	<i>Saccharomyces bayanus</i> MCYC 623
SAC_CAS	AACF01000006.1	<i>Saccharomyces castellii</i> NRRL Y-12630
SAC_CER	NP_013828.1	<i>Saccharomyces cerevisiae</i>
SAC_KUD	AACI01000788.1	<i>Saccharomyces kudriavzevii</i> IFO 1802
SAC_MIK1	AABZ01000665.1	<i>Saccharomyces mikatae</i> IFO 1815
SAC_PAR	AABY01000215.1	<i>Saccharomyces paradoxus</i> NRRL Y-17217
UST_MAY	EAK87196.1	<i>Ustilago maydis</i> 521
YAR_LIP1	Q6CG32	<i>Yarrowia lipolytica</i> CLIB99
YAR_LIP2	Q6C0L0	<i>Yarrowia lipolytica</i> CLIB99
YAR_LIP3	Q6CGN3	<i>Yarrowia lipolytica</i> CLIB99
YAR_LIP4	Q6C5T1	<i>Yarrowia lipolytica</i> CLIB99

## APPENDIX B

The  $\lambda$ -DNA/ HindIII+EcoRI Marker



## APPENDIX C

Copies of pages from website references used are attached.