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)

Chapter One – Aldehyde- and alcohol dehydrogenases: A literature Review.................................................. 1
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 Yarrowia lipolytica........54
Motivation..............................................................................................57
References............................................................................................59

Chapter Two – An exploratory study into the presence of long chain alcohol and aldehyde dehydrogenases in Yarrowia lipolytica........78
2.1 Introduction......................................................................................78
2.2 Materials and Methods....................................................................79
2.2.1 Growth of organisms.................................................................79
  2.2.1 (a) Pre-cultures............................................................................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 Y. lipolytica.....84
  2.2.5 (a) PCR amplification of FALDH and β-actin genes.....................84
  2.2.5 (b) PCR amplification of FADH gene..........................................86
  2.2.5 (c) PCR amplification of FAOD gene in C. tropicalis OC3.............87
  2.2.5 (d) Southern Hybridization.......................................................87
2.2.6 Expression analysis of ALDH and FADH genes

2.2.6(a) Northern Hybridization

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

2.2.6(c) Preparation of radiolabelled probes for Northern blot analyses

2.3 Results and Discussion

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

2.3.1.1 FAOD, FALDH and FALDH activity in hexadecane and glycerol grown cells

2.3.1.2 Subcellular fractionation of FADH and FALDH activity

2.3.1.3 NAD and NADP dependence of FADH and FALDH activity

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 genes

2.3.2.2 Southern Hybridization to detect FAOD genes in *Y. lipolytica*

2.3.3 Expression of FADH and FALDH genes

2.4 Conclusions

Chapter Three – Gene Disruption and Expression analysis of Fatty Aldehyde Dehydrogenases of *Yarrowia lipolytica*

3.1 Introduction

3.2 Materials and Methods

3.2.1 Plasmids, strains and media

3.2.2 General molecular biology techniques

3.2.3 Construction of disruption cassettes

3.2.4 Ura3 marker rescue by expression of Cre recombinase
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 FALDH-PUT 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

Chapter Four - Putative fatty aldehyde dehydrogenase-encoding genes from the sequenced fungal genomes........135

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

Chapter Five – General Discussion and Conclusions ..........170
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 - γ-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⁺ - Nicotinamide Adenine Dinucleotide Phosphate

NAD⁺ - Nicotinamide Adenine Dinucleotide

N-terminal – amino terminal

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

OD₆₂₀nm – 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

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

Chapter one - Alcohol dehydrogenases

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

Chapter two

Figure 2.1: Induction of FALDH and FADH activities in Y. lipolytica H222 cells
during batch culture. .................................................................90
Figure 2.2: A comparison of FAOD induction in relation to wet cell mass in Y.
lipolytica H222 cells growing in hexadecane. ..................................91
Figure 2.3: Subcellular fractionation of FADH, FALDH and FAOD..............92
Figure 2.4: Comparison of the SFA protein sequence with the partial FADH
amino acid sequence of Y. lipolytica.............................................96
Figure 2.5: PCR product of partial FADH gene sequence of Y. lipolytica.......97
Figure 2.6: Alignment of full FADH protein sequence of Y. lipolytica against
FADHs of other yeasts.................................................................98
Figure 2.7: Southern blot analysis of FAOD gene(s) of Y. lipolytica……….100
Figure 2.8: Northern Hybridization Analysis of FALDH and FADH genes…… 102
Figure 2.9: Expression Analysis of ALDH, FALDH and FADH genes by RT-PCR.........................................................................................................................103

Chapter three

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

Chapter four

Figure 4.1: A phylogenetic tree showing the position of the 25 putative fungal FALDH sequences in relation to other ALDHs.........................146
Figure 4.2: DNAassist 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**

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

**Chapter one - Alcohol dehydrogenases**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.1</td>
<td>The four classes of alcohol dehydrogenases</td>
<td>34</td>
</tr>
<tr>
<td>1.2.2</td>
<td>The new ADH nomenclature for human and mouse</td>
<td>46</td>
</tr>
</tbody>
</table>

**Chapter two**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Oligonucleotide sequence of primers used in chapter 2</td>
<td>85</td>
</tr>
<tr>
<td>2.2</td>
<td>FALDH activity levels in hexadecane and glucose grown cells using NAD and NADP as cofactors</td>
<td>94</td>
</tr>
<tr>
<td>2.3</td>
<td>A comparison of the four putative ALDHs of <em>Y. lipolytica</em> with other ALDHs</td>
<td>99</td>
</tr>
</tbody>
</table>

**Chapter three**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Organisms, strains and plasmids used in chapter 3</td>
<td>112</td>
</tr>
<tr>
<td>3.2</td>
<td>Oligonucleotide sequence of primers used in chapter 3</td>
<td>114</td>
</tr>
</tbody>
</table>

**Chapter four**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Fungi genome projects report</td>
<td>141</td>
</tr>
<tr>
<td>4.2</td>
<td>BLAST search of fungal genome databases</td>
<td>143</td>
</tr>
<tr>
<td>4.3</td>
<td>The ten most conserved sequence motifs in the 25 putative fungal FALDH sequences</td>
<td>153</td>
</tr>
<tr>
<td>4.4</td>
<td>Table summary of the position and number of transmembrane segments (TMS) in the 25 fungal FALDH/FALH-like proteins</td>
<td>158</td>
</tr>
</tbody>
</table>
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)^+-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 Sjögren-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 2.6\( \text{Å} \). 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 β-strands connected by four α-helices, instead of the usual six β-strands found in other NAD-dependent dehydrogenases. Also, in contrast to the classic β-α-β motif found in other NAD-dependent dehydrogenases a novel mode of NAD-binding, β-α,β motif is observed in ALDH. Moreover, the GxGxxG motif (where x denotes any amino acid) usually found associated with the β1-αA loop is found in ALDH to be in the β4-αD loop, as GxTxxG and associated with the nicotinamide ring. However, the role of β2 in NAD-binding specificity is still retained, resulting in the unique ALDH β-α,β binding motif.

In the ALDH family of enzymes Cys$^{243}$ in class 3 (Cys$^{302}$ 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 β-α-β 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 β-α-β 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](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.](http://www.psc.edu/science/2002/wymore/what_happens_at_the_active_site.html)

The sulphhydryl proton on Cys\textsuperscript{243} is important for initially recognizing the substrate. Substrate binding then causes a shift in the hydrogen bonding partner for Lys\textsuperscript{235} in which it then primarily interacts with the carbonyl oxygen of Thr\textsuperscript{242}, 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\textsuperscript{243} amide backbone supported by interactions with Lys\textsuperscript{235}. 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\textsuperscript{192} 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\textsuperscript{401}) 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\textsuperscript{114}, Thr\textsuperscript{186}, Glu\textsuperscript{209} and Leu\textsuperscript{210} are highly conserved; Gly\textsuperscript{187}, Gly\textsuperscript{211}, Cys\textsuperscript{243}, Glu\textsuperscript{333} and Phe\textsuperscript{335} 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)+ 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 a/β domain with an extended loop between β-1 and a-A, and the coenzyme binds between 2 helices, a-C and a-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 (a-A). Instead the NAD nicotinamide ring is oriented and stabilized by stacking between Gly\textsuperscript{187} and Phe\textsuperscript{335}, as well as by hydrogen bonding with Arg\textsuperscript{292}, from a-11 of the catalytic domain, to NO\textsubscript{3}* of the nicotinamide ribose (Liu et al., 1997). The NAD binding motif in ALDHs involves the loop region between β4, αD and β2 thus giving it the name β4-αD-β2 or β-α,β motif instead of the classic β1-αA-β2 (or β-α-β) motif observed in other NAD-dependent oxidoreductases.

The glycine-rich sequence, G\textsubscript{i}-xxx-G\textsubscript{2} (residues 187-192) equivalent to the NAD-binding motif G\textsubscript{i}xG\textsubscript{2}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\(^{140} \), 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 \textit{Vibrio harveyi} the polar residues are Thr\(^{175} \) (Zhang et al., 2001) and Thr\(^{180} \) in GAPDH (Cobessi et al., 1999).

Class 3 ALDHs however, are able to bind both NAD and NADP though Glu\(^{140} \) is still present. This may therefore imply that the presence of Glu\(^{140} \) 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

<table>
<thead>
<tr>
<th>Reaction No.</th>
<th>Reaction catalyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aldehyde + NAD(P) + H₂O → Acid + NAD(P)H + H⁺</td>
</tr>
<tr>
<td>2</td>
<td>Aldehyde + NAD(P) + CoA → Acyl-CoA + NAD(P) + H⁺</td>
</tr>
<tr>
<td>3</td>
<td>Aldehyde + orthophosphate + NAD(P) → Acyl phosphate + NAD(P)H + H⁺</td>
</tr>
<tr>
<td>4</td>
<td>Formaldehyde + Glutathione + NAD → S-formylglutathione + NADH + H⁺</td>
</tr>
<tr>
<td>5</td>
<td>Aryl aldehyde + NADP + AMP + pyrophosphate + H₂O → Aromatic acid + NADPH + ATP + H⁺</td>
</tr>
</tbody>
</table>

Substrate specificity-based classification

While no one has disputed the reaction-chemistry-directed ALDH classification approach, it seems b 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.

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

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-phosphatate 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 cyanellices, 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 γ-glutamyl semialdehyde dehydrogenase domain fused to a proline residue. The yeast PUT2 gene, encoding Δ1-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 rf2 gene known as the nuclear restorer and antiquitin, which may function in regulation of turgor pressure and/or general stress response.

Finally, Ω, η and ϖ-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.

<table>
<thead>
<tr>
<th>Mouse, Drosophila</th>
<th>All other species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>cDNA, mRNA</td>
</tr>
<tr>
<td>Aldh1a1</td>
<td>ALDH1A1</td>
</tr>
<tr>
<td>ALDH1A1</td>
<td>ALDH1A1</td>
</tr>
</tbody>
</table>

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 ALDHI 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.
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, class 1 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).

<table>
<thead>
<tr>
<th>Family</th>
<th>Examples</th>
<th>Pathway(s)</th>
<th>Substrate specificity</th>
<th>Coenzyme preference</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 3 ALDH</td>
<td>ALDH3A1</td>
<td>Metabolism of lipid peroxidation products, long chain “fatty” aldehydes and certain anti-cancer drugs</td>
<td>Variable, aromatics</td>
<td>NAD or NADP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH3A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-phosphorylating</td>
<td>ALDH11A3</td>
<td>Glycolysis in the dark to generate NADPH for photosynthetic reactions</td>
<td>Highly specific</td>
<td>NADP</td>
<td>Glyceraldehyde-3-phosphate to 3-phosphoglycerate</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate Dehydrogenase (GAPDH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aromatic ALDH</td>
<td>FeaB (phenyl acetaldehyde dehydrogenase)</td>
<td>Catabolism of aromatic aldehydes by microbes</td>
<td>Each aromatic ALDH oxidizes a specific aromatic aldehyde</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Succinic Semialdehyde</td>
<td>ALDH5A1</td>
<td>Metabolism of γ-aminobutyraldehyde (GABA)</td>
<td>Specific</td>
<td>NAD in animals, NADP in bacteria</td>
<td>Succinic semialdehyde to succinate</td>
</tr>
<tr>
<td>Dehydrogenase (SSALDH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turgor ALDH</td>
<td>ALDH3I1</td>
<td>Response to dehydration and osmotic turgor</td>
<td>Not yet determined</td>
<td>Not yet determined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH7B4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH7B6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH21A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ-Glutamyl Semialdehyde</td>
<td>ALDH4A1</td>
<td>Proline metabolism</td>
<td>Specific</td>
<td>NAD</td>
<td>γ-Glutamyl Semialdehyde to glutamate</td>
</tr>
<tr>
<td>Dehydrogenase (GGSALDH)</td>
<td>ALDH4B1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylmalonyl Semialdehyde</td>
<td>ALDH6A1</td>
<td>Valine and pyrimidine metabolism</td>
<td>Specific</td>
<td>NAD</td>
<td>Methylmalonyl semialdehyde to acetyl-CoA</td>
</tr>
<tr>
<td>Dehydrogenase (MMSALDH)</td>
<td>ALDH6B1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betaine ALDH</td>
<td>ALDH1H1</td>
<td>Resistance to dehydration and osmotic turgar</td>
<td>Specific</td>
<td>NAD</td>
<td>Betaine aldehyde to betaine</td>
</tr>
<tr>
<td>(BALDH)</td>
<td>ALDH10A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1 ALDH</td>
<td>ALDH1A1</td>
<td>Metabolism of ethanol, retinaldehyde, 11-hydroxy thromboxane B2, and certain anti-cancer drugs, structural crystallins in shrews and cephalopods</td>
<td>Variable, prefer aliphatic aldehydes</td>
<td>NAD</td>
<td></td>
</tr>
<tr>
<td>Class 2 ALDH</td>
<td>ALDH2B1</td>
<td>Metabolism of ethanol, pollen maturation</td>
<td>Variable, prefers aliphatic aldehydes</td>
<td>NAD</td>
<td></td>
</tr>
<tr>
<td>Fungal ALDH</td>
<td>ALDH1D3</td>
<td></td>
<td>Variable</td>
<td>NAD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH1E3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALDH17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-Formyltetrahydrofolate Dehydrogenase (FTDH)</td>
<td>ALDH1L1</td>
<td>Folate metabolism</td>
<td>Specific</td>
<td>NADP</td>
<td>10-Formyltetrahydrofolate to tetrahydrofolate and CO₂</td>
</tr>
<tr>
<td>2-Hydroxy muconic Semialdehyde Dehydrogenase (HMSALDH)</td>
<td>ALDH12A1</td>
<td>Meta-fission pathway for catechols</td>
<td>Specific for substituted 2-hydroxy muconic semi-aldehydes</td>
<td>NAD</td>
<td>2-Hydroxymuconic semi-aldehyde to 2-hydroxyhexa-2,4-diene-1,6-dicarboxylic acid</td>
</tr>
<tr>
<td>Group X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>Superkingdom</th>
<th>Taxon</th>
<th>Number of genes</th>
<th>Total number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>Crenarchaeae</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Euryarchaeota</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Bacteria</td>
<td>Aquificales</td>
<td>2</td>
<td>331</td>
</tr>
<tr>
<td></td>
<td>Cyanobacteria</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proteobacteria</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spirochaetales</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thermotogales</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thermus/Deinococcus</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryota</td>
<td>Diplomonadida</td>
<td>1</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>Euglenozoa</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Entamoebidae</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fungi</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metazoa</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Viridiplantae</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Total number</td>
<td></td>
<td></td>
<td>529</td>
</tr>
</tbody>
</table>
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$^{187}$ and Phe$^{335}$ – these two residues form an integral part of the coenzyme binding Rossmann fold by interacting with the nicotinamide portion of NAD(P).
Gly$^{240}$ – 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$^{333}$ – 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$^{25}$, Gly$^{105}$, Asn$^{114}$, Pro$^{116}$, Gly$^{131}$, Lys$^{137}$, Gly$^{211}$, Cys$^{243}$, Pro$^{337}$, Gly$^{383}$, Asn$^{388}$, and Gly$^{403}$. 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$^{211}$ is the first residue in the Gly-Gly dipeptide part of the boundary between coenzyme binding and catalytic domains, whereas Gly$^{403}$ is part of the “U-turn” region. Glu$^{209}$, 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$^{243}$ is the catalytic thiol, and together with Arg$^{25}$ 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$^{114}$ 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)

<table>
<thead>
<tr>
<th>Motif number&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Length</th>
<th>Motif&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>[Past]-[WFy]-[Ne]-[FYgalv]-[Ptl]</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>[Apnci]-[Liamv]-[AvSlcimm]-[AClimgf]-[G]-[Ncdl]-[Tvcspg]-[Vaimcltg]-[Vil]-[Lmiwahfchcy]-[Kh]-[Ptvghms]-[ASdhp]-[Epsadqgilt]</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>[Grkpwhsay]-[FLeivqarmmhk]-[Pg]-[Plakdiesvrf]-[Gnede]-[Vliat]-[VLifyac]-[Nglqshat]-[Vllyaqgfst]-[IVlms]</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>[IVlgfy]-[SATmnflthq]-[Fyla]-[Tvil]-[G]-[Sgen]-[Tvrsindepaqk]-[EAprqgktnldh]-[VTiasgm]-[Gafi]</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>[Lamfgs]-[Enlqt]-[Lmcagi]-[Gs]-[Ga]-[Knlmqshvl]-[SNade]-[Pahftswv]-[cnlmgivahst]-[IVlfa]-[Viamt]-[Fdlmhcanyv]-[Daeskpm]-[Dsntaev]-[Acvistey]-[Dnlera]</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>[Fylma]-[Fylrmdaqetwsvikp]-[Nhstyfaci]-[QAsnhtcmg]-[G]-[Qe]-[crvitksand]-[Cr]</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>[Gdtskae]-[Yfnarthelswv]-[FYLwvis]-[IVlfym]-[Qeapkgmrnhlsywv]-[Pa]-[Tachmy]-[Vil]-[FLivwn]</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>[Ektqrqs]-[E]-[IVlnsp]-[F]-[Ga]-[Ps]-[Vilcf]</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>[Nrst]-[Dnaseqtkregi]-[TSrvnacqgik]-[Epdtgqikrvfshync]-[Yflqvm]-[Gpa]-[Lnmv]-[Astgvqcfl]-[Agslfte]-[AGysct]-[Vllfams]-[Fhwyivlem]-[TSag]-[KRnsqteahdp]-[DNSileakt]</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>[Pasw]-[Fwahwv]-[Gtqs]-[G]-[Fvenesimtawrq]-[Kgm]-[mqarelnshgdpt]-[Stm]-[Gfls]-[lfntlygshrvq]-[Gdhhrsy]-[Rdpsagkte]</td>
</tr>
</tbody>
</table>

<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.
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

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](http://www.uchsc.edu/sp/sp/alcdbase/aldh-nomencl.html)). The numbers 1-10 in bold, indicate position of each of the ten motifs.
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\(^+\) to NADP\(^+\) as nucleotide cofactor (Kelson \textit{et al.}, 1997; Rizzo \textit{et al.}, 2001). However, bacterial FALDHs from \textit{Vibrio harveyi} and \textit{Acinetobacter spp.} were found to have a higher affinity for NADP\(^+\) than NAD\(^+\) (Singer and Finnerty, 1985; Zhang \textit{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 \textit{Candida}, \textit{Pseudomonas} and \textit{Acinetobacter} (Ueda and Tanaka, 1990; Singer and Finnerty, 1985; Fox \textit{et al.}, 1992; Zhang \textit{et al.}, 2001). The mammalian FALDH protein has a subunit molecular weight of about 54kDa, similar to most ALDHs (Kelson \textit{et al.}, 1997; Rizzo \textit{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-B4. 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 (Miyauchi \textit{et al.}, 1991; Vasiliou \textit{et al.}, 1996; Rogers \textit{et al.}, 1997; Chang and Yoshida, 1997; Lin \textit{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).

Exon 10

Mouse

\[\begin{array}{c}
\text{...VCLVA\textcolor{red}{VAAVI\textcolor{red}{V}}K^{481}} \\
\text{KYQALPRGKALLASLIVHR}{\textcolor{red}{RWSS Knot}^{507}} \\
\end{array}\]

Human

\[\begin{array}{c}
\text{...TFLGIV\textcolor{red}{VAAV}{\textcolor{red}{LV}}K^{481}} \\
\text{KYQAVLRR\textcolor{red}{KALLIF}VHR}{\textcolor{red}{LVHRWSSQR}^{508}} \\
\end{array}\]

Figure 1.1.5: Carboxy-terminal sequences of mouse and human major FALDH protein (exons 9-10) and minor FALDH? protein (exons 9-9'-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 (exon 9') 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 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 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).

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

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 isoymes. In addition, type I γ-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$^{373}$ 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)+.

\[
\text{ADH} \quad \text{RCH}_2\text{OH} \quad \text{RCHO} \quad \text{NAD(P)}^+ \quad \text{NAD(P)H}
\]

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)$^+$ 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/alcdehynrt.html).

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

<table>
<thead>
<tr>
<th>Class</th>
<th>Characteristics</th>
<th>Examples, EC numbers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short chain</td>
<td>~250-300 residues: Drosophila Adh (EC 1.1.1.1), steroid reductases (EC 1.1.1.51)</td>
<td>Jörnvall et al., 1995</td>
<td></td>
</tr>
<tr>
<td>Medium chain</td>
<td>dimeric forms</td>
<td>~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)</td>
<td>Jörnvall et al., Jörnvall et al., Persson et al., 1994</td>
</tr>
<tr>
<td></td>
<td>tetrameric forms</td>
<td>S. cerevisiae ADH (EC 1.1.1.1), sheep sorbitol dehydrogenase (EC1.1..1.14)</td>
<td>Jörnvall et al., Magonet et al., 1992</td>
</tr>
<tr>
<td></td>
<td>Non-zinc-</td>
<td>quinone reductase/β-crystallins (EC 1.6.5.5), enoyl reductase (EC 1.3.1.10)</td>
<td>Borrás et al., 1989</td>
</tr>
<tr>
<td></td>
<td>containing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long chain</td>
<td>600-750 residues, bacterial ethanol dehydrogenases, e.g. Acetobacter acetii; pyrroloquinoline quinine as cofactor (EC 1.1.99.8)</td>
<td>Mackintosh and Fewson, 1987 Inoue et al., 1989</td>
<td></td>
</tr>
<tr>
<td>Iron-activated</td>
<td>1,2-propanediol dehydrogenase of E. coli (EC 1.1.1.77), Adh2 of Zymomonas mobilis (EC 1.1.1.1); AdhIV of S. cerevisiae</td>
<td>Sridhara et al., 1989 Scopes, 1983 Conway et al., 1987 Conway and Ingram, 1989 Williamson and Paquin, 1987</td>
<td></td>
</tr>
</tbody>
</table>

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: GHEX₂GX₅(G,A)X₂(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)+-dependent oxidoreductase enzyme activities, though it exhibits a catalytic mechanism completely different
from the zinc-containing ADHs. 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 20β-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-β-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 α-helix/β-strand structure characteristic of nucleotide binding domains, while the C-terminal half does not appear to have this alternating α/β 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)+-independent ADHs. Members of this group are those enzymes that use neither NAD+ nor NADP+ 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 organophillum 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 crystalloographic and sequence studies of the horse liver enzyme
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.

![Ribbon diagram of horse liver alcohol dehydrogenase](image)

**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/alcdeyznrt.html](http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdeyznrt.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 \(\text{Cys}^{46}, \text{Cys}^{174}\) and \(\text{His}^{67}\) ([http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdeyznrt.html](http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdeyznrt.html)). An ionizable water molecule hydrogen bonds with the hydroxyl group of Thr\(^{48}\) 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.](image)

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 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$^+$ is bound in position by multiple residues of the Rossmann fold, a series of α-β-α folds (http://pps98.cryst.bbk.ac.uk/assignment/projects/debono/alcdehyznrt.html). Some of the NAD-binding residues include Gly$^{210}$, Asn$^{225}$, Pro$^{243}$, Asn$^{242}$, Val$^{268}$, Asp$^{223}$, Try$^{178}$, Arg$^{47}$, Gly$^{292}$, Val$^{203}$. The ADH active site is made up of three amino acid residues, Phe$^{93}$, Leu$^{57}$ and Leu$^{116}$, 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$^{48}$ and Phe$^{93}$) 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 a/β 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\(^{138}\), Tyr\(^{151}\), Lys\(^{155}\)) 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\(^+\), 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\(^{51}\) draws a proton from NAD, which in turn draws a proton from Thr\(^{48}\), 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/alcoholdehydroy/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).

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/alcddbase/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 Dueuter et al., (1999) has also been incorporated into the new system. Dueuter 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.

<table>
<thead>
<tr>
<th>Class</th>
<th>Old human symbol</th>
<th>New human symbol</th>
<th>Old mouse symbol</th>
<th>New mouse symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>ADH1</td>
<td>ADH1A</td>
<td>Adh1</td>
<td>Adh1</td>
</tr>
<tr>
<td></td>
<td>ADH2</td>
<td>ADH1B</td>
<td>Adh1</td>
<td>Adh1</td>
</tr>
<tr>
<td></td>
<td>ADH3</td>
<td>ADH1C</td>
<td>Adh5</td>
<td>Adh5</td>
</tr>
<tr>
<td>II</td>
<td>ADH4</td>
<td>ADH4</td>
<td>Adh5</td>
<td>Adh5</td>
</tr>
<tr>
<td>III</td>
<td>ADH5</td>
<td>ADH5</td>
<td>Adh5-ps</td>
<td>Adh5-ps</td>
</tr>
<tr>
<td>III</td>
<td>ADH5P1</td>
<td>ADH5P1</td>
<td>Adh5-ps</td>
<td>Adh5-ps</td>
</tr>
<tr>
<td>IV</td>
<td>ADH7</td>
<td>ADH7</td>
<td>Adh3</td>
<td>Adh7</td>
</tr>
<tr>
<td>V</td>
<td>ADH6</td>
<td>ADH6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 ADHα, ADHβ and ADHγ 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 ADHα, ADHβ and ADHγ, 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, ZEAma; 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 1.1.1236 and β-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). 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_2 – C_{14}, 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$_2$O$_2$ (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 alkanes 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 ω-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


Alcohol dehydrogenase: Chem 231 Lecture No. 16 Sp 1999 (9/10/2005)


**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


**Craterostigma plantagineum** and *Arabidopsis thaliana*. The Plant Journal, 28(5): 555-567.


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'lchenko 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'lchenko 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)\(^+\)-dependent FAOD (Kemp et al., 1988, 1994; I'lchenko et al., 1994; Mauersberger et al., 1992; Hommel et al., 1994). These authors argued that reduction of NAD(P)\(^+\) 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)\(^+\) reduction is due to the activity of NAD(P)\(^+\)-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°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\textsubscript{620nm} 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\textsubscript{620nm} 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°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$_{620\text{nm}}$ 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$_{22}$ 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® reagent (Invitrogen Life Technologies). The isolated RNA was stored in formamide at -80°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:

\[
\text{CH}_3(\text{CH}_2)_n\text{CH}_2\text{OH} + \text{NAD}^+ \xrightarrow{\text{FADH}} \text{CH}_3(\text{CH}_2)_n\text{CHO} + \text{NADH} + \text{H}^+
\]

\[
\text{CH}_3(\text{CH}_2)_n\text{CHO} + \text{NAD}^+ \xrightarrow{\text{FALDH}} \text{CH}_3(\text{CH}_2)_n\text{COOH} + \text{NADH} + \text{H}^+
\]

Assay Method

Principle: The reduction of NAD\(^+\) (i.e. NADH formation, noticeable by increase in absorbance) is measured at 30° 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 (ε) of 6.27M\(^{-1}\)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.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Assay mixture (volumes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM Tris-HCl buffer, pH8.5</td>
<td>37ul</td>
</tr>
<tr>
<td>20mM substrate (alcohols in DMSO, aldehydes in dioxane)</td>
<td>13ul</td>
</tr>
<tr>
<td>13mM NAD(^+)</td>
<td>50ul</td>
</tr>
<tr>
<td>13mM NaN(_3) (sodium azide)</td>
<td>50ul</td>
</tr>
<tr>
<td>Enzyme extract (10X diluted)</td>
<td>50ul</td>
</tr>
</tbody>
</table>

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\(^+\) and NADP\(^+\) (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:**

\[
\text{CH}_3(\text{CH}_2)_n\text{CH}_2\text{OH} \xrightarrow{\text{oxidase}} \text{C H}_3(\text{CH}_2)_n \text{CHO}
\]

**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\(_2\) \xrightarrow{\text{alcohol oxidase}} \text{product} + \text{H}_2\text{O}_2 \text{ followed by,}

(ii) \text{H}_2\text{O}_2 + \text{ABTS} \xrightarrow{\text{peroxidase}} \text{H}_2\text{O} + \text{ABTS}_{\text{ox}}

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

**Principle:** The spectrophotometric assay method measures the liberation of \text{ABTS}_{\text{ox}} 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 ($\varepsilon$) 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](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, MgCl₂-containing PCR buffer and one unit of Taq polymerase (Roche Diagnostics). PCR conditions for 25 cycles were as follows; denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 72°C for 1 min. For PCR amplification of *ALDH2*, however the annealing temperature was 48°C. PCR amplification of the β-actin gene was also performed under similar conditions with annealing temperature of 55°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°C for later use in Southern and northern blot analyses.

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence (5’→ 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1</td>
<td>FA 1-1F</td>
<td>ATCAACAAACGAGTGGTC</td>
</tr>
<tr>
<td></td>
<td>FA 1-1R</td>
<td>CAGCGAAATGGGAGATGG</td>
</tr>
<tr>
<td>ALDH2</td>
<td>FA 2-1F</td>
<td>GAGAACTACCTCCAGAC</td>
</tr>
<tr>
<td></td>
<td>FA 2-2R</td>
<td>GACCATCTCGACCCATC</td>
</tr>
<tr>
<td>FALDH3</td>
<td>FA 3-1F</td>
<td>CGATTCTCTGTGATGCAGG</td>
</tr>
<tr>
<td></td>
<td>FA 3-2R</td>
<td>TTGGCATCCACTCCGTC</td>
</tr>
<tr>
<td>FALDH4</td>
<td>FA 4-1F</td>
<td>AAAATGTCTACCTTTGATTGG</td>
</tr>
<tr>
<td></td>
<td>FA 4-1R</td>
<td>CATCGTACATGCTTGG</td>
</tr>
<tr>
<td>FADH</td>
<td>LADH1-1F</td>
<td>CGC CT A CAC TCT CAG TGG CTC TGA</td>
</tr>
<tr>
<td></td>
<td>LADH 1-1R</td>
<td>TGA AAG TCG CAA ACC TCC GTC GGG</td>
</tr>
<tr>
<td>*FAOD (C. tropicalis)</td>
<td>C-trop-FAOD-1F</td>
<td>ATGGCTAGTTTTTACCAGACAAAG</td>
</tr>
<tr>
<td></td>
<td>C-trop-FAOD-1R</td>
<td>CTACAATTTAGCCTTGGATTTCAAG</td>
</tr>
</tbody>
</table>

* 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 \(^*\) 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 (50\(\mu\)l) contained approximately 500ng/\(\mu\)l genomic DNA, 0.2mM dNTPs, 100\(\mu\)M of each of the two primers, MgCl\(_2\)-containing PCR buffer and one unit of Taq polymerase (Roche Diagnostics). PCR conditions for 25 cycles were as follows; denaturation at 94\(^0\)C for 1 min, annealing at 50\(^0\)C for 1 min, and extension at 72\(^0\)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\(^0\)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 (50\(\mu\)l) contained approximately 500ng/\(\mu\)l genomic DNA, 0.2mM dNTPs, 100\(\mu\)M of each of the two primers, MgCl\(_2\)-containing PCR buffer and 1\(\mu\)l Taq polymerase (Roche Diagnostics). Amplification was performed under the following conditions for 25 cycles; denaturation at 94\(^0\)C for 1 min, annealing at 55\(^0\)C for 1 min, and extension at 72\(^0\)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, SalI and XbaI, was electrophoresed on a 1% agarose gel and then blotted onto a nylon transfer membrane (0.22µ, 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 λ DNA was added to the hybridization mixture to allow visualization of the molecular weight marker. Low and high stringency hybridizations were performed at 52°C and 68°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 β-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 (0.22µ, magnacharge, Osmonics Inc.) by capillary blotting, using 10X SSC. The membranes were probed by overnight hybridization at 65°C with the radiolabelled (³²P-dATP) PCR amplified gene sequences of ALDH 1 and 2, FALDH 3 and 4, FADH, FAOD and β-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°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\textsubscript{12} 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 \textit{ALDH}s, \textit{FADH} and \textit{\(\beta\)-actin} genes were radiolabelled with \textsuperscript{32}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\textsubscript{10} aldehyde (1-decanal) and C\textsubscript{12} 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'lchenko *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_{16}) and glycerol was followed over time by taking OD_{620nm} 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.

![Graph of FAOD induction](image)

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_{16}). 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 YPER 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⁺. According to literature the enzyme uses both NAD⁺ and NADP⁺ (Cobessi *et al.*, 1999, Zhang *et al.*, 2001). The following experiment was aimed at determining exactly which cofactor, NAD⁺ or NADP⁺, 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\(^+\) and NADP\(^+\) as
cofactor. As was expected enzyme activity was higher when using NADP\(^+\) than NAD\(^+\).
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)\(^+\) binding pocket as well as structural shift that accompanies NADP\(^+\)
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\(^+\) and NADP\(^+\) dependence of FALDH activities of Y. lipolytica H222 cells grown in
hexadecane and glucose. FALDH activities were assayed using 1-dodecanal as a substrate.

<table>
<thead>
<tr>
<th>Growth media</th>
<th>FALDH activity (U/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-dodecanal+ NAD(^+)</td>
</tr>
<tr>
<td>Hexadecane</td>
<td>1.90 ± 0.71</td>
</tr>
<tr>
<td>Glucose</td>
<td>1.35 ± 0.36</td>
</tr>
</tbody>
</table>

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 (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 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](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.
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 $^{32}$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 λ (EcoRI/HindIII) marker (Lane 1) is shown along side for size comparison. The sizes of the MR marker fragments appear in appendix B.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-methanolica</td>
<td>--MSTAGKTITCKAAVAWEAGKDLSEIETIEVAPPAKEVRKIAYTGVCTHDTAFLTS6GK</td>
<td>58</td>
</tr>
<tr>
<td>C-boidinii</td>
<td>--MSTVGKTITCKAAVAWKPGEALSLSEIETIEVAPPAKEVRKIAYTGVCTHDTAFLTS6GAD</td>
<td>58</td>
</tr>
<tr>
<td>C-maltosa</td>
<td>MSESTVGKPTICTAWEAKAWEAKPLSLSEIETIEVAPPAKEVRKIAYTGVCTHDTAFLTS6GAD</td>
<td>60</td>
</tr>
<tr>
<td>Y-lipolytica</td>
<td>--MSTEGKTITCKAAVAWEAKPGLPLEDEVEAPPQAEVRKIAYTGVCTHDTAFLTS6GAD</td>
<td>58</td>
</tr>
<tr>
<td>S-cerevisiae</td>
<td>MSAATVGKPIKCAVDAKPLS6EITVDPAKEVRKIAYTGVCTHDTAFLTS6GAD</td>
<td>60</td>
</tr>
</tbody>
</table>

Zinc-ADH signature

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-methanolica</td>
<td>AEGEFPVI GHGEGAGESVGETVSVEKVGDHVCELTECECKFCKSKTNLCGIRRA</td>
<td>118</td>
</tr>
<tr>
<td>C-boidinii</td>
<td>AEGEFPVI GHGEGAGESVGETVSVEKVGDHVCELTECECKFCKSKTNLCGIRRA</td>
<td>118</td>
</tr>
<tr>
<td>C-maltosa</td>
<td>PEGIFPSVI GHGEGAGESVGETVSVEKVGDHVCELTECECKFCKSKTNLCGIRRA</td>
<td>120</td>
</tr>
<tr>
<td>Y-lipolytica</td>
<td>PEGIFPSVI GHGEGAGESVGETVSVEKVGDHVCELTECECKFCKSKTNLCGIRRA</td>
<td>118</td>
</tr>
<tr>
<td>S-cerevisiae</td>
<td>PEGIFPSVI GHGEGAGESVGETVSVEKVGDHVCELTECECKFCKSKTNLCGIRRA</td>
<td>120</td>
</tr>
</tbody>
</table>

NAD-binding site

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-methanolica</td>
<td>TGKGKLMDFGTSFRT-CGKLLHYMGCTSFQTVLADISVAVDPKFAPMDRTGCGGCG</td>
<td>177</td>
</tr>
<tr>
<td>C-boidinii</td>
<td>TGKGKLMDFGTSFRT-CGQPLLHYMGCTSFQTVLADISVAVDPKFAPMDRTGCGGCG</td>
<td>177</td>
</tr>
<tr>
<td>C-maltosa</td>
<td>TGKGKLMDFGTSFRT-CGQPLLHYMGCTSFQTVLADISVAVDPKFAPMDRTGCGGCG</td>
<td>177</td>
</tr>
<tr>
<td>Y-lipolytica</td>
<td>TGKGKLMDFGTSFRT-CGQPLLHYMGCTSFQTVLADISVAVDPKFAPMDRTGCGGCG</td>
<td>177</td>
</tr>
<tr>
<td>S-cerevisiae</td>
<td>TGKGKLMDFGTSFRT-CGQPLLHYMGCTSFQTVLADISVAVDPKFAPMDRTGCGGCG</td>
<td>177</td>
</tr>
</tbody>
</table>

---

97
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 \([GHEX_2GX_5(G,A)X_2(I,V,A,C,S)]\) and the NAD-binding site \((GxGxxG \text{ or } GxTxxG)\) (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\(_P\) of *S. cerevisiae* (Wehner et al., 1993) and human ADH3\(_P\) (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.

<table>
<thead>
<tr>
<th></th>
<th>DNA sequence</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ALDH1</td>
<td>ALDH2</td>
</tr>
<tr>
<td>ALDH1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALDH2</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>FALDH3</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>FALDH4</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Human FALDH</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>S. cerevisiae ALDH1</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

* 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).

![Southern hybridization to probe for FAOD in Y. lipolytica.](image)

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, SalI and XbaI, respectively and the λ (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 \( ^{32} \)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 β-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, C₁₂ alkane, C₁₆ alkane, C₁₆ alkane + 2% pristane, C₂₂ alkane + 2% pristane and C₁₆ acid + 0.1% Tween 80. 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\textsubscript{12} 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 λ (HindIII/EcoRl) marker is in the middle.](image)

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


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°C in Luria-Bertani broth. Selection of Ampicillin resistance in E. coli was performed on media containing 100µg ml⁻¹ 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⁺ and Leu2⁺ 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⁺ selection and YNB broth containing 0.01% uracil and all amino acids except leucine (YNB leu-) for Leu2⁺ selection. The amino acid mixture and uracil were first dissolved in water, filter-sterilized and then added to autoclaved YNB broth cooled to 60°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

<table>
<thead>
<tr>
<th>Strain (host strain)</th>
<th>Plasmid, genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E coli strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 10</td>
<td>F mcrA ?(mrr-hsdRMS-mcrBC) F80lacZ?M15</td>
<td>Invitrogen</td>
</tr>
<tr>
<td></td>
<td>lacX74 deoR recA1 ara D139 ?(ara-leu)7697</td>
<td></td>
</tr>
<tr>
<td></td>
<td>galK rpsL(StuR) endA1 nupG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>?80dlacZΔm15, recA1, endA1, gyrA96, thi-1,</td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>hsdR17 (φλ-, mhs +), supE44, relA1, deoR,</td>
<td>Promega</td>
</tr>
<tr>
<td></td>
<td>Δ(lacZYA-argF)U169</td>
<td></td>
</tr>
<tr>
<td>JME459 (DH5α)</td>
<td>Bluescript KS+ (&lt;ColE1 LacZ bla)</td>
<td>Stratagene</td>
</tr>
<tr>
<td>JME130 (DH5α)</td>
<td>URA3 in Bluescript KS+</td>
<td>Nicaud et al. unpublished</td>
</tr>
<tr>
<td>JME512 (DH5α)</td>
<td>JMP112 (LPR synthetic oligonucleotide in Bluescript KS+)</td>
<td>Fickers et al 2003</td>
</tr>
<tr>
<td>JME507 (DH5α)</td>
<td>JMP113 (1.2 kb URA3 fragment in JMP112, MU cassette)</td>
<td>Fickers et al 2003</td>
</tr>
<tr>
<td>JME461 (DH5α)</td>
<td>pRRQ2 (&lt;Cre ARS68 LEU2 in Bluescript KS&gt;)</td>
<td>Richard et al 2001</td>
</tr>
<tr>
<td><strong>Y. lipolytica strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Po1d</td>
<td>MATa ura3-302 leu2-270 xpr2-322</td>
<td>CLIB139</td>
</tr>
<tr>
<td>H222</td>
<td>MATa (wild type)</td>
<td>Barth and Gaillardin, 1996</td>
</tr>
<tr>
<td>E150</td>
<td>MATb ura3-302 leu2-270 xpr2-322 lip2::hph</td>
<td>Barth and Gaillardin, 1996</td>
</tr>
</tbody>
</table>

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 50μl, 5μl 10X PCR buffer (Roche diagnostics), 10mM dNTP mix (containing 10mM each of dATP, dTTP, dGTP and dCTP), 20-50ng of template DNA, 100?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°C for 30s, annealing at 60°C for 30s and extension at 72°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) 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®-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®-T PT at corresponding I-Sce I site. Thus the correct plasmid FALDH- pGEM®-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®-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⁺ transformants, which appeared after approximately 36h, were selected on YNB ura⁻ 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

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

*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 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 \textit{FALDH} isogenes in \textit{Y. lipolytica} using \textit{FALDH4} gene deletion as an example. (A) The promoter and terminator regions of the \textit{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\textsuperscript{®}-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 \textit{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 *lox* P 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\textsubscript{10}), dodecane (C\textsubscript{12}), tetradecane (C\textsubscript{14}) and octadecane (C\textsubscript{18}). 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\textdegree 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 laurinaldehyde (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 FALDH 3 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- 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 \textit{FALDH2} gene, designated \textit{FALDH134?::loxp} initially seemed to exhibit arrested growth on the alkane plates. In later experiments this was however no longer the case. With four \textit{FALDH} genes deleted one would have expected to see a pronounced change of phenotype from the \textit{FALDH1234?::loxp} strain, but this was not the case. It thus appeared that the \textit{FALDH} deletions had not affected the growth on alkanes on solid media.

![Figure 3.9: Growth analysis of \textit{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 300C for 48h.](image)
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.

![Graph](image1)

**Figure 3.10**: A comparison of growth of the parent strain (Po1d) with the FALDH triple and quadruple deletion mutants in (a)YP 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\textsubscript{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_{10}), tetradecane (C_{14}), hexadecane (C_{16}) and octadecane (C_{18}) 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, 10 min) 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](image)

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 (Δ134) and the quadruple deletion mutant (Δ1234) failed, because FALDH activity for all strains was too low to detect significant differences (< 0.01U/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


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

PR Matatiele¹, MS Smit¹, J Albertyn¹ and J-M Nicaud².

¹ University of the Free State, Department of Microbial, Biochemical and Food Biotechnology, P.O. Box 339, Bloemfontein, 9300, Republic of South Africa
² 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$_6$ up to C$_{40}$) 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)$^+$-dependent aldehyde dehydrogenases (ALDHs) [aldehyde: NAD(P)$^+$ 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/aldcbase/aldhnomencl.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 Sjörgen-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, \textit{ALDH14} from \textit{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 \textit{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 \textit{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 \textit{et al.}, 2003) and SMART (http://smart.embl-heidelberg.de/) (Schultz \textit{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 \textit{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 \textit{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


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](http://www.ncbi.nlm.nih.gov/genomes/FUNGI/funtab.html)).

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Ascomycota (28)</th>
<th>Basidiomycota (4)</th>
<th>Microsporidia (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pezizomycotina (10)</td>
<td>Saccharomycotina (17)</td>
<td>Schizosaccharomyces (1)</td>
<td></td>
</tr>
<tr>
<td>Aspergillus fumigatus</td>
<td>Candida albicans – WGS</td>
<td>Schizosaccharomyces pombe - complete</td>
<td>Coprinopsis cinerea okavama7#130 - WGS</td>
</tr>
<tr>
<td>Aspergillus nidulans FGSC A4 - WGS</td>
<td>Candida glabrata CBS138 – complete</td>
<td>Cryptococcus neoformans var. grubii H99 - WGS</td>
<td></td>
</tr>
<tr>
<td>Aspergillus parasiticus</td>
<td>Candida tropicalis</td>
<td>Phanerochaete chrysosporium - WGS</td>
<td></td>
</tr>
<tr>
<td>Aspergillus terreus - WGS</td>
<td>Debaryomyces hansenii CBS767 - complete</td>
<td>Ustilago maydis 521 - WGS</td>
<td></td>
</tr>
<tr>
<td>Coccidioides immitis R5 - WGS</td>
<td>Eremothecium goessypi - complete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coccidioides posadastii C735</td>
<td>Kluyveromyces lactis NRRL Y-1140 – complete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusarium sporotrichioides</td>
<td>Kluyveromyces waltii NCYC 2644 - WGS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gibberella zeae PH-1 - WGS</td>
<td>Saccharomyces cerevisiae - complete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnaporthe grisea 70-15 - WGS</td>
<td>Saccharomyces bayanus 623-6C - WGS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neospora crassa strain OR74A - WGS</td>
<td>Saccharomyces bayanus MCYC 623 - WGS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naumovia castellii NRRL Y-12630 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces kluyvert NRRL Y-12651 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces kudriavzevii IFO 1802 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces mikatae IFO 1815 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces mikatae IFO 1815 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces paradoxus NRRL Y-17217 - WGS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yarrowia lipolytica CLIB990 - complete</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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>
<thead>
<tr>
<th>Gene sequence No.</th>
<th>Protein sequence No.</th>
<th>Organism</th>
<th>Description</th>
<th>TBLASTN</th>
<th>Amino acid Length</th>
<th>% Similarity to other FALDHs (ClustalW scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XM_404038</td>
<td>EAK87196.1</td>
<td><em>Ustilago maydis</em> 521</td>
<td></td>
<td>384</td>
<td>43</td>
<td>61</td>
</tr>
<tr>
<td>AACQ010000035.1</td>
<td>EAL00022.1</td>
<td><em>Candida albicans</em> SC5314</td>
<td></td>
<td>317</td>
<td>39</td>
<td>58</td>
</tr>
<tr>
<td>AACQ010000020.1</td>
<td>EAL02020.1</td>
<td><em>Candida albicans</em> SC5314</td>
<td></td>
<td>312</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>XM_323331.1</td>
<td>EAA49061.1</td>
<td><em>Magnaporthe grisea</em> 70-15</td>
<td></td>
<td>309</td>
<td>37</td>
<td>58</td>
</tr>
<tr>
<td>AACU01000968.1</td>
<td>EAA53613.1</td>
<td><em>Magnaporthe grisea</em> 70-15</td>
<td></td>
<td>303</td>
<td>37</td>
<td>57</td>
</tr>
<tr>
<td>CR382139</td>
<td>CAG90160</td>
<td><em>Debaryomyces hansenii</em> CBS767</td>
<td></td>
<td>300</td>
<td>35</td>
<td>57</td>
</tr>
<tr>
<td>AABX01000718.1</td>
<td>EAA28332.1</td>
<td><em>Neurospora crassa</em> OR74A</td>
<td></td>
<td>280</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*TIGR222929</td>
<td>contig:1140:c posadasii</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*TIGR5085</td>
<td>contig:5231:a fumigatus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACD0100098.1</td>
<td>EAA62737.1</td>
<td><em>Aspergillus nidulans</em> FGSC A4</td>
<td></td>
<td>211</td>
<td>41</td>
<td>59</td>
</tr>
<tr>
<td>AACD01000168.1</td>
<td>EAA64317.1</td>
<td><em>Aspergillus nidulans</em> FGSC A4</td>
<td></td>
<td>214</td>
<td>31</td>
<td>54</td>
</tr>
<tr>
<td>AACM01000412.1</td>
<td>EAA67898.1</td>
<td><em>Gibberella zeae</em> PH-T</td>
<td></td>
<td>230</td>
<td>33</td>
<td>56</td>
</tr>
<tr>
<td>AACU01000237.1</td>
<td>EAA6915.1</td>
<td><em>Magnaporthe grisea</em> 70-15</td>
<td></td>
<td>214</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>NC_001145.2</td>
<td>NP_013828.1</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td></td>
<td>248</td>
<td>32</td>
<td>53</td>
</tr>
<tr>
<td>AACD01000092.1</td>
<td>EAA28332.1</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td></td>
<td>280</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td>AACB01000665.1</td>
<td></td>
<td></td>
<td><em>Kluveromyces waltii</em> NCYC2644</td>
<td></td>
<td>267</td>
<td>35</td>
</tr>
<tr>
<td>*AACH01001154.1</td>
<td></td>
<td></td>
<td><em>Saccharomyces mikatae</em> IFO 1815</td>
<td></td>
<td>256</td>
<td>35</td>
</tr>
<tr>
<td>AACD01000400.1</td>
<td></td>
<td></td>
<td><em>Saccharomyces bayanus</em> MRCY 623</td>
<td></td>
<td>255</td>
<td>34</td>
</tr>
<tr>
<td>AACD01000215.1</td>
<td></td>
<td></td>
<td><em>Saccharomyces paradoxus</em> NRRL Y-17217</td>
<td></td>
<td>247</td>
<td>33</td>
</tr>
<tr>
<td>AACI01000788.1</td>
<td></td>
<td></td>
<td><em>Saccharomyces kudriavzevi</em> IFO 1802</td>
<td></td>
<td>236</td>
<td>32</td>
</tr>
<tr>
<td>AACF0100006.1</td>
<td></td>
<td></td>
<td><em>Saccharomyces castellii</em> NRRL Y-12630</td>
<td></td>
<td>219</td>
<td>30</td>
</tr>
<tr>
<td>CR382128.1/XM_500380</td>
<td>XP_500380 (FALDH3)</td>
<td><em>Yarrowia lipolytica</em> CLIB99</td>
<td></td>
<td>342</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>CR382132.1/XM_505802</td>
<td>XP_505802 (FALDH4)</td>
<td><em>Yarrowia lipolytica</em> CLIB99</td>
<td></td>
<td>340</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>CR382131.1/XM_503981</td>
<td>XP_503981 (FALDH2)</td>
<td><em>Yarrowia lipolytica</em> CLIB99</td>
<td></td>
<td>327</td>
<td>38</td>
<td>58</td>
</tr>
<tr>
<td>CR382127.1/XM_500179</td>
<td>XP_500179 (FALDH1)</td>
<td><em>Yarrowia lipolytica</em> CLIB99</td>
<td></td>
<td>307</td>
<td>38</td>
<td>57</td>
</tr>
<tr>
<td>AAЕY01000059</td>
<td>EAL17607.1</td>
<td><em>Cryptococcus neoformans</em> var. neoformans B-3901A</td>
<td></td>
<td>355</td>
<td>36</td>
<td>28-33</td>
</tr>
<tr>
<td>XP_452089.1</td>
<td>CR382122.1</td>
<td><em>Kluveromyces lactis</em> NRRL Y-1140</td>
<td></td>
<td>248</td>
<td>32</td>
<td>54</td>
</tr>
<tr>
<td>CR380957.1</td>
<td>CAG61338.1</td>
<td><em>Candida glabrata</em> strain CBS138</td>
<td></td>
<td>239</td>
<td>32</td>
<td>53</td>
</tr>
</tbody>
</table>
*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 DAH - 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 \( \geq 40\% \) similarity are described as belonging to two different families, whereas sequences that are \( \geq 60\% \) 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-alkalase1 1 ME-P ------------------------------P 4
C-alkalase1 2 MEKPSKIEKSAADIAARCASKERTKVKEPLQQETKEVLEQPVSKVRRNIKTSPKSEKTSQTPSTPAASYAKHNPNTAEF 91
s-bayanus 1 0
s-nakasato 1 0
s-cerevisiae 1 0
s-paradoxus 1 0
s-kaltaste 1 0
s-castelli 1 0
C-glabrata 1 0
K-lactis 1 0
k-waltii 1 0
y-lipolytica2 1 0
y-lipolytica3 1 0
y-lipolytica1 1 0
m-grisea1 1 0
m-grisea2 1 0
n-crassa 1 0
g-tea 1 0
a-nidulans1 1 0
hs-PALD 1 0
hs-PALD1 1 0
U-maydis 1 0
m-grisea1 1 0
m-grisea2 1 0
a-nidulans2 1 0
Sc-ALDH1 1 0
D-hansenl 1 0
At-ALDH3 1 0
C-neoforams 1 0
C-alkalase1 5 SKID ------------------SSD.------------------FPDID-MTSTSWY ------------------TEKIEQGVVMHELTDSSPT 41
C-alkalase1 92 AKPHEKSLTSPIPSQKQAGATTKEKSQVDSKTSSTTSYTVANNVSFLVNLY ------------------TELEISQIFVERTAFRPHS 162
s-bayanus 1 MNDIC -- KI ------------------------LYTVDEEIQDQA ------------------SGCQFD-KQLKSHSLE KS 39
s-nakasato 1 MNODA -- I ------------------------SRHDFD-KQLKSHSLE KS 39
s-cerevisiae 1 MENDS -- KI ------------------------LYNTPSVEKIDEEI ------------------SNRFKPE-KQLKSHSLE KS 39
s-paradoxus 1 MNOSG -- KI ------------------------LYNTPSVEKIDEII ------------------SNRFFFF-KQLKSHSLE KS 39
s-kaltaste 1 M -- HMLELTSELEDEQVII ------------------TALDIDSPCTEDKLMAS 67
s-castelli 1 M -- AT ------------------------LYTPEVACTDQIEK ------------------SNFKYLA-QQLQSHSHE 34
C-glabrata 1 MA -- EI ------------------------LYTELEIDAKAIEK ------------------SNFQQQ-NKRELAQK 35
K-lactis 1 M -- G --------------AKQFRS ------------------NAQALAAK 35
k-waltii 1 NE ------------------------ANYTPVEIADAIVAS ------------------CVNKSYK -QREKVEANS 33
y-lipolytica2 1 MEFWESLLLPF ------------------------TPDIEEQGRLQIKSFRAS 31
y-lipolytica4 1 MEFWESLPPF ------------------------TPDIEEQGRLQIKSFRAS 31
U-maydis 1 MTTAFTKTPNTVPTTSPLKETASKGTPVTSNVPTWEMSCXK ------------------TLPSEEISISKRKTFS 63
m-grisea1 1 MNIKSRPHLMLIVKIEILPVYKLNSGIPVPMRSKAGISSPFAEEA ------------------TALQDSIPCTDKVLKMAFS 67
n-crassa 1 MAKKEVIAFPEV ------------------------TPIDAIQAVCTWATATAS 32
n-crassa 1 MAKMTQVTTPGQF ------------------------TPDIEANAKOVLKMTKFS 24
a-nidulans1 1 MGAIDIPQFF------------------------TPQIEQGFSRKTQXTE 30
hs-PALD 1 0 MELEVRK ------------------------VRQAFSL 14
hs-PALD1 1 0 MUEIYAEK ------------------------AMAFAFS 17
U-maydis 46 TLAAHQPQLRALTAFSHALLLPFPSSPSPTLRLVCLVLVRNHSVHSVSSRLQSPQIMAATATAEALGWTPTIDDIPIVDRLGAALFT 136
C-neoforams 1 0 MTIITTIVHYSTPIRTY ------------------STPAQVDHATHTAFRT 37
C-alkalase1 42 KQRT8HD-QRIPMQGLMIAAIAVRQDAE -- ALSKDFHM ESETEEKVLTVQGIKVLRVTASILHENV ------------------KPEK 111
C-alkalase1 113 C-TEESWQFLRQLGARVTQGQKELCQDQKOAI -- ALQQRDFHM ESETEEKVLTVQGIKVLRVTASILHENV ------------------KPEK 231
s-bayanus 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
s-nakasato 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
s-cerevisiae 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
s-paradoxus 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
s-kaltaste 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
s-castelli 1 DPKREDILFQRLQGYKIAAEDKREDE ID -- AMYQDFHM TELEISLIRTLKMDLILHELIIKPRM ------------------KPEK 109
C-glabrata 1 0 MVRASRDKQKPQPQXWYKFRKYLE -- AKLFMRH QSREISLITVLKDLLLQIPPMN ------------------KPEK 105
K-lactis 1 0 SPRAT1MEIRIKQYFGQKDRHQAIA -- AASDRFDH ASLETITFLVPRHLYLIRNNRKQK ------------------EPKX 104
k-waltii 1 0 DPIKDEGELTEKAIKEVQFQDYKREDE ID -- AMQFDHM SQREYVLMTQGLVNLHIAELQPMI ------------------KPEK 103
y-lipolytica2 1 0 D-KTLDDNHRQDIQ1KRKMVAVKKK KDIKDV -- AIHDLDRPVFETEQLKFQSFUTEHNIQVYVLKMWKA ------------------ADMT 100
y-lipolytica4 1 0 D-KTLDDNHRQDIQ1KRKMVAVKKK KDIKDV -- AIHDLDRPVFETEQLKFQSFUTEHNIQVYVLKMWKA ------------------ADMT 100
y-lipolytica3 64 D-KTLDDLDRQDIQ1KRKMVAVKKK KDIKDV -- AKLKDPAFEDTEHAIYISQVWQFMFQVYNVHKLWKA ------------------KPEK 132
y-lipolytica1 29 D-KTLDDLAFQDIQ1KRKMVAVKKK KDIKDV -- AKLKDPAFEDTEHAIYISQVWQFMFQVYNVHKLWKA ------------------KPEK 97
s-castelli 1 68 H-RTEDVEKMVRQVXAVQHEKELQY -- AMQFDHM PASEEKLHRVFQSDPMGKMLHATPA ------------------KPEK 136
n-crassa 33 H-KTNQWQRLLQVVRQIYDDAFAA AKA -- ALQQRDFHM GYESIOTFQWYK废LMDLIELNLEPAF ------------------KPEK 101
m-grisea1 35 H-KTNQWQRLLQVVRQIYDDAFAA AKA -- ALQQRDFHM GYESIOTFQWYK废LMDLIELNLEPAF ------------------KPEK 101
a-nidulans1 31 H-KTNQWQRLLQVVRQIYDDAFAA AKA -- ALQQRDFHM GYESIOTFQWYK废LMDLIELNLEPAF ------------------KPEK 101
hs-PALD 15 G-RSHRFLRQGQLQHMERKELK -- AIAADLSAEFVNFQKLVTFUTGVSQDMLEMPPFF ------------------KPEK 83
hs-PALD1 18 G-RSHRFLRQGQLQHMERKELK -- AIAADLSAEFVNFQKLVTFUTGVSQDMLEMPPFF ------------------KPEK 83
U-maydis 137 G-KTVRESTQMKLQETYKMKEAY -- SLRTKDQHSFESIFAPMKTNGREIFAVTKIDMAK -- KFAPX 205
m-grisea1 38 G-LKSKLEKMKQKMKREMUKARMD -- AKLKDPAFEDTEHAEYISQVWQFMFQVYNVHKLWKA ------------------KPEK 106
m-grisea2 26 G-KTVREKVMQNLQWHPDEKAKDEEKE -- ALKMDMAHLEFVVEQWYQWYVLMLMNQFQ ------------------ATEP 94
a-nidulans2 49 G-STENVCEVSAETTEDCSSCAADQGAFDTKQPIGKLEKLEKQLGQDIQVSSALBNQGK--TALRKEFIPAIAAWCAAGAAA
D-hansenl 68 STEERHITDVYEALEEDIDTAVEAASAAFKS-SWSTGDPSVRANALLKLADLVDANADTLAHIEALDNGK-SLMCSRGDVALTAMYFRSCAG 45
At-ALDH3 77 MREKGGNLRAAFIIRIQSRPKNLSKFTVLDPLVHVRLKLPSS 44
C-neoforams 30 G-VTCSLAYRLKQKAGQHNEAAKA -- AVENQQQDPFDTGIA LALREIDAVLKVHVINMMK -- DESH 99

149
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 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>
<thead>
<tr>
<th>Motif number</th>
<th>Length</th>
<th>Motif in all ALDHs</th>
<th>Motif in fungal FALDHs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td><img src="https://example.com/motif1.png" alt="Motif 1" /></td>
<td><img src="https://example.com/motif1-in-fungal-FALDHs.png" alt="Motif 1 in fungal FALDHs" /></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td><img src="https://example.com/motif2.png" alt="Motif 2" /></td>
<td><img src="https://example.com/motif2-in-fungal-FALDHs.png" alt="Motif 2 in fungal FALDHs" /></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td><img src="https://example.com/motif3.png" alt="Motif 3" /></td>
<td><img src="https://example.com/motif3-in-fungal-FALDHs.png" alt="Motif 3 in fungal FALDHs" /></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td><img src="https://example.com/motif4.png" alt="Motif 4" /></td>
<td><img src="https://example.com/motif4-in-fungal-FALDHs.png" alt="Motif 4 in fungal FALDHs" /></td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td><img src="https://example.com/motif5.png" alt="Motif 5" /></td>
<td><img src="https://example.com/motif5-in-fungal-FALDHs.png" alt="Motif 5 in fungal FALDHs" /></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td><img src="https://example.com/motif6.png" alt="Motif 6" /></td>
<td><img src="https://example.com/motif6-in-fungal-FALDHs.png" alt="Motif 6 in fungal FALDHs" /></td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td><img src="https://example.com/motif7.png" alt="Motif 7" /></td>
<td><img src="https://example.com/motif7-in-fungal-FALDHs.png" alt="Motif 7 in fungal FALDHs" /></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td><img src="https://example.com/motif8.png" alt="Motif 8" /></td>
<td><img src="https://example.com/motif8-in-fungal-FALDHs.png" alt="Motif 8 in fungal FALDHs" /></td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td><img src="https://example.com/motif9.png" alt="Motif 9" /></td>
<td><img src="https://example.com/motif9-in-fungal-FALDHs.png" alt="Motif 9 in fungal FALDHs" /></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td><img src="https://example.com/motif10.png" alt="Motif 10" /></td>
<td><img src="https://example.com/motif10-in-fungal-FALDHs.png" alt="Motif 10 in fungal FALDHs" /></td>
</tr>
</tbody>
</table>

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 cystein 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\textsuperscript{243}, Glu\textsuperscript{333}, Phe\textsuperscript{335}, Gly\textsuperscript{187} and Gly\textsuperscript{240} (Perozich \textit{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\textsuperscript{243}, which serves as the catalytic thiol, is found within motif 6. while Glu\textsuperscript{333} and Phe\textsuperscript{335} reside in motif 8 (Hempel \textit{et al.}, 2001). Glu\textsuperscript{333} serves as the general base, which activates the catalytic thiol through a water molecule. It may also be involved in cofactor binding (Hempel \textit{et al.}, 2001). ALDHs lacking enzymatic activity such as the Ω-crystallins and allergens have been found to have other residues in place of Cys\textsuperscript{243}, Glu\textsuperscript{333} or Phe\textsuperscript{335}. Gly\textsuperscript{187} together with Phe\textsuperscript{335} form an integral part of the coenzyme binding Rossmann fold by interacting with the nicotinamide portion of NAD(P). Gly\textsuperscript{240} 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\textsuperscript{25}, Gly\textsuperscript{105}, Asn\textsuperscript{114}, Pro\textsuperscript{116}, Gly\textsuperscript{131}, Lys\textsuperscript{137}, Gly\textsuperscript{211}, Pro\textsuperscript{337}, Gly\textsuperscript{383}, Asn\textsuperscript{388}, and Gly\textsuperscript{403}) were found conserved in more than 95% of 145 ALDH sequences examined by Perozich \textit{et al} (1999). Some of these residues lie at critical turns and loops in the ALDH structure. For example, Gly\textsuperscript{211} is part of the dipeptide, which marks the boundary between the coenzyme binding and catalytic domains. Gly\textsuperscript{403} is involved in one of the U-turn regions whereas Asn\textsuperscript{114} 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\textsuperscript{388} is replaced by glycine in \textit{U. maydis} and the \textit{Saccharomyces} sequences. Asn\textsuperscript{114} is missing in the \textit{A. nidulans} isozyme1 sequence while Pro\textsuperscript{337} is replaced by threonine in the \textit{A. nidulans} isozyme2 sequence.

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

Another group of residues has been proposed to be specifically important in the ALDH3 family (Hempel \textit{et al.}, 2003). The group consists of Asp\textsuperscript{247}, Tyr\textsuperscript{248}, Lys\textsuperscript{359}, Pro\textsuperscript{360}, Glu\textsuperscript{69}, Gln\textsuperscript{122}, Pro\textsuperscript{123}, Lys\textsuperscript{235}, and Thr\textsuperscript{242}. The first four residues are important in subunit/subunit interaction recognition at the dimer interface. All these residues except Pro\textsuperscript{360} and Glu\textsuperscript{69} are generally very variable among these fungal FALDHs. Even Glu\textsuperscript{69} is replaced by aspartic acid in the \textit{M. grisea} sequences. Glu\textsuperscript{69} and Gln\textsuperscript{122} lie in the substrate-binding pocket of the ALDH structure. Gln\textsuperscript{122} is very variable in the fungal FALDHs, while Pro\textsuperscript{123} is relatively conserved (exceptions are \textit{C. albicans} isozyme1, \textit{D. hansenii} and \textit{G. zeae}). Mutation of Pro\textsuperscript{123} is implicated in SLS (Rizzo \textit{et al.}, 2001). Asp\textsuperscript{247} 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 \textit{et al.}, 2001). Histidine is found in place of Tyr\textsuperscript{248} in the \textit{Magnaporthe} and \textit{Aspergillus} sequences. Lys\textsuperscript{235} and Thr\textsuperscript{242} 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 \textit{et al.} (2004) strongly suggested that Val\textsuperscript{363} is important for structural folding of the catalytic domain and is therefore essential for the normal functioning of the FALDH protein. However, the Val\textsuperscript{363} 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^{363} 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 *Y. lipolytica*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 (Miyauchi 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.

<table>
<thead>
<tr>
<th>FALDH (organism)</th>
<th>No of TMSs</th>
<th>Position of TMD residues</th>
<th>Length</th>
<th>Cell localization (% Reliability)</th>
<th>Other features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human ALDH10</td>
<td>1</td>
<td>463-480</td>
<td>18</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 463)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible peroxisomal targeting signal: AKL at 148</td>
</tr>
<tr>
<td>A-nidulans1</td>
<td>1</td>
<td>119-135</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 119)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No special features found</td>
</tr>
<tr>
<td>A-nidulans2</td>
<td>3</td>
<td>136-142, 381-397, 494-510</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 3a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leucine zipper pattern: LLYLVVLLGLADVGLDDRLGVL at 495</td>
</tr>
<tr>
<td>C-albicans1</td>
<td>0</td>
<td></td>
<td></td>
<td>Cytoplasmic (94.1%)</td>
<td>ER membrane retention signals: TNTK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vacuolar targeting motif: ILPI at 490</td>
</tr>
<tr>
<td>C-albicans2</td>
<td>0</td>
<td></td>
<td></td>
<td>Nuclear (55.5%)</td>
<td>No special features found</td>
</tr>
<tr>
<td>C-glabrata</td>
<td>0</td>
<td>136-152, 506-522</td>
<td>17</td>
<td>Cytoplasmic (89%)</td>
<td>ER membrane retention signals: KLF4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2nd peroxisomal targeting signal: KLLGDILHL at 85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Membrane topology: type 3a</td>
</tr>
<tr>
<td>C-neodermatidis</td>
<td>0</td>
<td></td>
<td></td>
<td>Cytoplasmic (94.1%)</td>
<td>ER membrane retention signals: YKAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vacuolar targeting motif: ILPI at 34</td>
</tr>
<tr>
<td>L-hansenii</td>
<td>0</td>
<td></td>
<td></td>
<td>Cytoplasmic (94.1%)</td>
<td>No special features found</td>
</tr>
<tr>
<td>C-zeae</td>
<td>2</td>
<td>138-154, 505-521</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 3a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No special features found</td>
</tr>
<tr>
<td>K-lactis</td>
<td>2</td>
<td>122-138, 503-519</td>
<td>17</td>
<td>Cytoplasmic (89%)</td>
<td>Membrane topology: type 3a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vacuolar targeting motif: ILPI at 364</td>
</tr>
<tr>
<td>K-waltii</td>
<td>0</td>
<td></td>
<td></td>
<td>Cytoplasmic (89%)</td>
<td>No special features found</td>
</tr>
<tr>
<td>M-grisea1</td>
<td>2</td>
<td>127-143, 499-521</td>
<td>23</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 141)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vacuolar targeting motif: ILPI at 100</td>
</tr>
<tr>
<td>M-grisea2</td>
<td>1</td>
<td>127-144</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 128)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No special features found</td>
</tr>
<tr>
<td>M-grisea3</td>
<td>1</td>
<td>173-189</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 173)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No special features found</td>
</tr>
<tr>
<td>N-crasa</td>
<td>2</td>
<td>137-153, 510-527</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 137)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ER membrane retention signals: KARR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leucine zipper pattern: LQWRVQLRKLWALDDFKASL at 38</td>
</tr>
<tr>
<td>S-bayanus</td>
<td>2</td>
<td>130-152, 513-529</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 3a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2nd peroxisomal targeting signal: KLMNDILHL at 89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vacuolar targeting motif: ILPI at 100</td>
</tr>
<tr>
<td>S-castellii</td>
<td>1</td>
<td>127-143</td>
<td>17</td>
<td>Cytoplasmic (94.1%)</td>
<td>Membrane topology: type 2 (cytoplasmic tail 1 to 127)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ER membrane retention signals: YKAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Di-leucine motif in the tail: LL at 41 and LL at 85</td>
</tr>
</tbody>
</table>
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


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


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


Studies intended to elucidate the biochemistry and genetics of the alkane-degradation pathway have mainly focused on cytochrome P450 monooxygenases and the β-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'lichenko 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


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.
<table>
<thead>
<tr>
<th>ID</th>
<th>Accession No.</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>At_ALDH3F1</td>
<td>CAE48163.1</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>At_ALDH3H1</td>
<td>CAE51203.1</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>At_ALDH3I1</td>
<td>CAC84903.1</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>BENZ_ACI</td>
<td>Q6FCB6</td>
<td>Acinetobacter sp.(strain ADP1)</td>
</tr>
<tr>
<td>BENZ_CAU</td>
<td>Q9A5Q0</td>
<td>Caulobacter crescentus</td>
</tr>
<tr>
<td>BENZ_HAL</td>
<td>Q9HMJ6</td>
<td>Halobacterium sp. (NRC-1)</td>
</tr>
<tr>
<td>BENZ_PSE</td>
<td>P43503</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>BENZ_RAL</td>
<td>Q8XT86</td>
<td>Ralstonia solanacearum</td>
</tr>
<tr>
<td>BENZ_STR</td>
<td>Q9L124</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>BENZ_XAN</td>
<td>Q8PDI4</td>
<td>Xanthomonas campestris (pv. campestris)</td>
</tr>
<tr>
<td>Ce_ALDH3C1_1</td>
<td>Q60WY8</td>
<td>Caenorhabditis briggsae</td>
</tr>
<tr>
<td>Ce_ALDH3C1_2</td>
<td>O16518</td>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>Ce_ALDH3C1_3</td>
<td>Q86S57</td>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>Ce_ALDH3C2</td>
<td>Q8MXJ7</td>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>DHA1_BOVIN</td>
<td>P48644</td>
<td>Bos taurus</td>
</tr>
<tr>
<td>DHA1_CHICK</td>
<td>P27463</td>
<td>Gallus gallus</td>
</tr>
<tr>
<td>DHA1_HUMAN</td>
<td>P00352</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>DHA1_MACFA</td>
<td>Q8HYE4</td>
<td>Macaca fascicularis</td>
</tr>
<tr>
<td>DHA1_MOUSE</td>
<td>P24549</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>DHA1_RABIT</td>
<td>Q8MI17</td>
<td>Oryctolagus cuniculus</td>
</tr>
<tr>
<td>DHA1_SHEEP</td>
<td>P51977</td>
<td>Ovis aries</td>
</tr>
<tr>
<td>DHA2_HUMAN</td>
<td>O94788</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>DHA2_MOUSE</td>
<td>Q62148</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>DHA2_RAT</td>
<td>Q63639</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>DHA5_HUMAN</td>
<td>P30837</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>DHAM_BOVIN</td>
<td>P20000</td>
<td>Bos taurus</td>
</tr>
<tr>
<td>DHAM_HORSE</td>
<td>P12762</td>
<td>Equus caballus</td>
</tr>
<tr>
<td>ID</td>
<td>Accession No.</td>
<td>Organism</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>DHAM_HUMAN</td>
<td>P05091</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>DHAM_MOUSE</td>
<td>P47738</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>DHAM_RAT</td>
<td>P11884</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>DHAS_CHICK</td>
<td>O93344</td>
<td>Gallus gallus</td>
</tr>
<tr>
<td>Dr_ALDH3D1_1</td>
<td>NP_775328.2</td>
<td>Danio rerio</td>
</tr>
<tr>
<td>Eh_ALDH13</td>
<td>AAA19741.1</td>
<td>Entamoeba histolytica</td>
</tr>
<tr>
<td>FUNG_ASH</td>
<td>Q758W1</td>
<td>Ashbya gossypii</td>
</tr>
<tr>
<td>FUNG_CAN</td>
<td>Q6FVP8</td>
<td>Candida glabrata CBS138</td>
</tr>
<tr>
<td>FUNG_CAN2</td>
<td>Q6FPK0</td>
<td>Candida glabrata CBS138</td>
</tr>
<tr>
<td>FUNG_DEB</td>
<td>Q6BJB3</td>
<td>Debaryomyces hansenii CBS767</td>
</tr>
<tr>
<td>FUNG_KLU</td>
<td>Q6CLU0</td>
<td>Kluyveromyces lactis NRRL Y-1140</td>
</tr>
<tr>
<td>FUNG_NEU</td>
<td>Q8X0L4</td>
<td>Neurospora crassa</td>
</tr>
<tr>
<td>FUNG_PIC</td>
<td>Q12648</td>
<td>Pichia angusta</td>
</tr>
<tr>
<td>FUNG_SAC</td>
<td>P46367</td>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>FUNG_YAR</td>
<td>Q6CD79</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
<tr>
<td>GAPDH_ARA1</td>
<td>P25857</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>GAPDH_CAP</td>
<td>Q8VWP2</td>
<td>Capsicum annuum</td>
</tr>
<tr>
<td>GAPDH_NIC</td>
<td>P09044</td>
<td>Nicotiana tabacum</td>
</tr>
<tr>
<td>GAPDH ORY</td>
<td>Q7X8A1</td>
<td>Oryza sativa (japonica cultivar-group)</td>
</tr>
<tr>
<td>GAPDH_PIN</td>
<td>P12859</td>
<td>Pisum sativum</td>
</tr>
<tr>
<td>GAPDH_PIS</td>
<td>Q41019</td>
<td>Pinus sylvestris</td>
</tr>
<tr>
<td>GAPDH_SPI1</td>
<td>P12860</td>
<td>Spinacia oleracea</td>
</tr>
<tr>
<td>GAPDH_ZEA</td>
<td>Q6LBU9</td>
<td>Zea mays</td>
</tr>
<tr>
<td>HOM_SAPL</td>
<td>P51648</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Hs_ALDH3A1</td>
<td>Q6PKA6</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Hs_ALDH3A2</td>
<td>P51648</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Hs_ALDH3B1</td>
<td>Q8N515</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Hs_ALDH3B2</td>
<td>P43353</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Hs_ALDH3B2</td>
<td>P48448</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>ID</td>
<td>Accession No.</td>
<td>Organism</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Mm_ALDH3A1</td>
<td>P47739</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>Mm_ALDH3A2</td>
<td>P47740</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>Mm_ALDH3B1</td>
<td>Q8VHW0</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>MmALDH3A2</td>
<td>Q99L64</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>MUS_MUSL</td>
<td>AAK01551.1</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>Po_alkH</td>
<td>P12693</td>
<td>Pseudomonas oleovorans</td>
</tr>
<tr>
<td>Pp_alkH</td>
<td>CAB51050.1</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Rn_ALDH3A1</td>
<td>P11883</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>Rn_ALDH3A2</td>
<td>P30839</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>Rn_ALDH3A2</td>
<td>P30839</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>SSDH_ACI</td>
<td>Q6F9G0</td>
<td>Acinetobacter sp. (strain ADP1)</td>
</tr>
<tr>
<td>SSDH_BUR2</td>
<td>Q62B48</td>
<td>Burkholderia mallei ATCC 23344</td>
</tr>
<tr>
<td>SSDH_BURK1</td>
<td>Q63NL9</td>
<td>Burkholderia pseudomallei</td>
</tr>
<tr>
<td>SSDH_ESC</td>
<td>Q8X950</td>
<td>Escherichia coli O157:H7</td>
</tr>
<tr>
<td>SSDH_HOM</td>
<td>Q8N3W6</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>SSDH_NEU</td>
<td>Q7SFB1</td>
<td>Neurospora crassa</td>
</tr>
<tr>
<td>SSDH_PSE1</td>
<td>Q9I6M5</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>SSDH_PSE2</td>
<td>Q88AT6</td>
<td>Pseudomonas syringae (pv. tomato)</td>
</tr>
<tr>
<td>SSDH_RAT</td>
<td>P51650</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>SSDH_YAR</td>
<td>Q6C0B4</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
<tr>
<td>ASP_NID1</td>
<td>EAA62737.1</td>
<td>Aspergillus nidulans FGSC A4</td>
</tr>
<tr>
<td>ASP_NID2</td>
<td>EAA64317.1</td>
<td>Aspergillus nidulans FGSC A4</td>
</tr>
<tr>
<td>CAN_ALB1</td>
<td>EAL0022.1</td>
<td>Candida albicans SC5314</td>
</tr>
<tr>
<td>CAN_ALB2</td>
<td>EAL02020.1</td>
<td>Candida albicans SC5314</td>
</tr>
<tr>
<td>CAN_GLA</td>
<td>CAG61338.1</td>
<td>Candida glabrata strain CBS138</td>
</tr>
<tr>
<td>CRY.neo</td>
<td>EAL17607.1</td>
<td>Cryptococcus neoformans var. neoformans B-3501A</td>
</tr>
<tr>
<td>DEB_HAN</td>
<td>XP_461954.1</td>
<td>Debaryomyces hansenii CBS767</td>
</tr>
<tr>
<td>GIB_ZEA</td>
<td>EAA67698.1</td>
<td>Gibberella zeae PH-1</td>
</tr>
<tr>
<td>ID</td>
<td>Accession No.</td>
<td>Organism</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>KLU_LAC</td>
<td>CR382122.1</td>
<td>Kluyveromyces lactis NRRLY-1140</td>
</tr>
<tr>
<td>KLU_WAL</td>
<td>AADM01000092.1</td>
<td>Kluyveromyces waltii NCYC2644</td>
</tr>
<tr>
<td>MAG_GRI1</td>
<td>EAA49061.1</td>
<td>Magnaporthe grisea 70-15</td>
</tr>
<tr>
<td>MAG_GRI2</td>
<td>EAA53613.1</td>
<td>Magnaporthe grisea 70-15</td>
</tr>
<tr>
<td>MAG_GRI3</td>
<td>CAG90160</td>
<td>Debaryomyces hansenii CBS767</td>
</tr>
<tr>
<td>NEU_CRA</td>
<td>EAA28392.1</td>
<td>Neurospora crassa OR74A</td>
</tr>
<tr>
<td>SAC_BAY</td>
<td>AACA01000400.1</td>
<td>Saccharomyces bayanus MCYC 623</td>
</tr>
<tr>
<td>SAC_CAS</td>
<td>AACF01000006.1</td>
<td>Saccharomyces castellii NRRL Y-12630</td>
</tr>
<tr>
<td>SAC_CER</td>
<td>NP_013828.1</td>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>SAC_KUD</td>
<td>AACI01000788.1</td>
<td>Saccharomyces kudriavzevii IFO 1802</td>
</tr>
<tr>
<td>SAC_MIK1</td>
<td>AABZ01000665.1</td>
<td>Saccharomyces mikatae IFO 1815</td>
</tr>
<tr>
<td>SAC_PAR</td>
<td>AABY01000215.1</td>
<td>Saccharomyces paradoxus NRRL Y-17217</td>
</tr>
<tr>
<td>UST_MAY</td>
<td>EAK87196.1</td>
<td>Ustilago maydis 521</td>
</tr>
<tr>
<td>YAR_LIP1</td>
<td>Q6CG32</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
<tr>
<td>YAR_LIP2</td>
<td>Q6C0L0</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
<tr>
<td>YAR_LIP3</td>
<td>Q6CGN3</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
<tr>
<td>YAR_LIP4</td>
<td>Q6C5T1</td>
<td>Yarrowia lipolytica CLIB99</td>
</tr>
</tbody>
</table>
APPENDIX B

The α-DNA/ HindIII+EcoRI Marker

APPENDIX C

Copies of pages from website references used are attached.