ENZYMATIC STUDIES OF ALCOHOL DEHYDROGENASE BY A COMBINATION OF IN VITRO AND IN SILICO METHODS

Mikko Hellgren

Stockholm 2009
Abstract

The family of alcohol dehydrogenases (ADHs) catalyzes conversions of alcohols, ketons and aldehydes. The early discovery and isolation of ADH (1937) was followed by numerous investigations. It was also the first dimeric enzyme for which the three-dimensional structure was determined (1974). Recent findings have revealed new physiological functions for the ADH enzymes. One type is a key enzyme in hepatic retinol metabolism, another is a main formaldehyde scavenger and a regulator of S-nitrosothiols levels. ADH genes have been shown to be connected to diseases and syndromes, such as alcoholism and asthma. Hence, investigations of structure-function relationships of the ADH enzymes are of both physiological and medical interest. The aim of this thesis was to study structure-function relationships and to investigate and identify ligands for ADH, with in vitro and in silico methods.

The catalytic activities of human, mouse and rat ADH2 for retinoids, were determined. The Kₘ values for human ADH2 are the lowest among all known human dehydrogenases, which supports a key role for human ADH2 in the hepatic retinoid metabolism. ADH3 is an enzyme with a proposed role as an NO scavenger. Two new lines of ligands, bile acids and fatty acids, were investigated for their potential effects on NO homeostasis. The bronco dilatatory effect of NO suggests that ADH3 inhibition could potentially work as treatment of obstructive lung disorders. The stability of the quaternary structure of sorbitol dehydrogenase (SDH) was determined by in vitro experiments and in silico energy calculations. A hydrogen-bonding network crucial for the tetrameric stability in SDH was identified. This network is located at a region enclosing the structural zinc site in mammalian ADHs. The structural zinc site was studied in detail by a combination of molecular dynamics and quantum mechanics simulations. The simulations revealed that the interaction between the cysteine residues and the zinc atom is of an electrostatic and covalent nature.

With in silico and in vitro simulations, interactions between ligands and the active site were determined, revealing site specific interactions within both ADH2 and ADH3. Furthermore, studies of subunit interactions and the structural zinc site revealed properties of the quaternary stability.
List of original papers

This thesis is based on the following papers, printed in the appendix, and will be referred to in the text by their roman numerals. Publications are sorted by publication dates in chronological order:


II. Hellgren M, Kaiser C, de Haij S, Norberg A, Hoog JO (2007a) A hydrogen-bonding network in mammalian sorbitol dehydrogenase stabilizes the tetrameric state and is essential for the catalytic power. Cellular and Molecular Life Sciences 64(23): 3129-3138


Additional publications not included in the thesis


# Contents

1 Introduction to alcohol dehydrogenase (ADH) ........................................................ 11
   1.1 A brief historical background: From the Big Bang to the discovery of ADH .. 11
   1.2 Medium-chain dehydrogenase/reductase (MDR) proteins ................................ 13
   1.3 The MDR-ADH protein superfamily ............................................................... 14
   1.4 Human ADH1 to ADH5 .................................................................................. 15
      1.4.1 ADH1 ...................................................................................................... 16
      1.4.2 ADH2 ...................................................................................................... 16
      1.4.3 ADH3 ...................................................................................................... 17
      1.4.4 ADH4 ...................................................................................................... 18
      1.4.5 ADH5 ...................................................................................................... 18
   1.5 Tetrameric ADHs ............................................................................................ 18
   1.6 Sorbitol dehydrogenase ................................................................................... 19
   1.7 The central dogma of ADH ............................................................................. 20
      1.7.1 Genes and protein sequences ................................................................... 20
      1.7.2 Protein expression patterns ...................................................................... 21
      1.7.3 Tertiary structure ..................................................................................... 22
      1.7.4 Active site ................................................................................................ 23
      1.7.5 Structural zinc site ................................................................................... 25
      1.7.6 Quaternary structure ................................................................................ 25
   1.8 The metabolism of endogenous and exogenous substrates .............................. 27
      1.8.1 Catalytic activities ................................................................................... 28
      1.8.2 Ethanol .................................................................................................... 29
      1.8.3 Formaldehyde.......................................................................................... 30
      1.8.4 Nitric oxide.............................................................................................. 31
      1.8.5 Vitamin A................................................................................................ 32
      1.8.6 Sorbitol and the polyol pathway .............................................................. 33
      1.8.7 Coenzyme NAD ...................................................................................... 34
   1.9 Mutations, polymorphisms and disease ........................................................... 35
   2 Theories and methods for enzymatic studies ........................................................... 37
      2.1 Proteins as enzymes......................................................................................... 37
      2.2 In vitro methods for protein experiments ......................................................... 38
         2.2.1 Expression, purification and identification of proteins ............................ 38
         2.2.2 Gel filtration .......................................................................................... 38
         2.2.3 Enzyme kinetics .................................................................................... 39
      2.3 In silico methods for protein interaction .......................................................... 40
         2.3.1 Molecular dynamics simulations with GROMACS .................................. 41
         2.3.2 Monte Carlo simulations with ICM ......................................................... 42
         2.3.3 Quantum chemistry simulations with GAUSSIAN ................................. 43
   3 Results and discussions, Paper I to V ...................................................................... 44
      3.1 Paper I ............................................................................................................. 44
      3.2 Paper II ........................................................................................................... 46
      3.3 Paper III .......................................................................................................... 48
      3.4 Paper IV .......................................................................................................... 50
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-HDA</td>
<td>12-hydroxydodecanoic acid</td>
</tr>
<tr>
<td>ADH</td>
<td>Alcohol dehydrogenase (MDR-ADH)</td>
</tr>
<tr>
<td>ADH 1/2/3/4</td>
<td>Alcohol dehydrogenase Class I/II/III/IV</td>
</tr>
<tr>
<td>ALDH</td>
<td>Aldehyde dehydrogenase</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DEAE</td>
<td>Diethylaminoethyl</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>GROMACS</td>
<td>Groningen machine for chemical simulations (software)</td>
</tr>
<tr>
<td>GSH</td>
<td>Glutathione, reduced form</td>
</tr>
<tr>
<td>GSNO</td>
<td>S-nitrosoglutathione</td>
</tr>
<tr>
<td>GSSG</td>
<td>Glutathione, oxidized form</td>
</tr>
<tr>
<td>HMGSH</td>
<td>S-hydroxymethylglutathione</td>
</tr>
<tr>
<td>ICM</td>
<td>Internal coordinate mechanics (software)</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MDR</td>
<td>Medium-chain dehydrogenases/reductases</td>
</tr>
<tr>
<td>M-M</td>
<td>Michaelis-Menten</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NAD⁺</td>
<td>NAD, oxidized form</td>
</tr>
<tr>
<td>NADH</td>
<td>NAD, reduced form</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric oxide</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum mechanics</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SDH</td>
<td>Sorbitol dehydrogenase</td>
</tr>
</tbody>
</table>
VS Virtual screening
Wt Wild type
Å Ångström (10⁻¹⁰ m)

One and three letter codes for the 20 naturally occurring amino acids

A Ala Alanine  
C Cys Cysteine  
D Asp Aspartic acid  
E Glu Glutamic acid  
F Phe Phenylalanine  
G Gly Glycine  
H His Histidine  
I Ile Isoleucine  
K Lys Lysine  
L Leu Leucine  
M Met Methionine  
N Asn Asparagine  
P Pro Proline  
Q Gln Glutamine  
R Arg Arginine  
S Ser Serine  
T Thr Threonine  
V Val Valine  
W Trp Tryptophan  
Y Tyr Tyrosine
**List of definitions**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affinity</td>
<td>The affinity between two molecules is determined from the association and dissociation rates at equilibrium; this affinity is physically determined by non-covalent interactions.</td>
</tr>
<tr>
<td>Compound</td>
<td>A molecule with an unknown affinity to a protein.</td>
</tr>
<tr>
<td>Configuration</td>
<td>The three dimensional relationship between all atoms in a molecule that is fixed by chemical bonds.</td>
</tr>
<tr>
<td>Conformation</td>
<td>Structural differences from rotation around single bonds, which do not involve the breaking or formation of chemical bonds.</td>
</tr>
<tr>
<td>Chemical reaction</td>
<td>A reaction which include the breakage and/or formation of covalent bonds.</td>
</tr>
<tr>
<td>De Novo</td>
<td>Chemical reactions in an anabolic process where simple molecules are utilized to construct complex molecules.</td>
</tr>
<tr>
<td>In silico</td>
<td>Methods where simulations, models and calculations are carried out with a computer.</td>
</tr>
<tr>
<td>In vitro</td>
<td>Laboratory method where no living organism is involved.</td>
</tr>
<tr>
<td>In vivo</td>
<td>Laboratory method where living organisms constitute the basic medium for the experiment.</td>
</tr>
<tr>
<td>Ligand</td>
<td>A molecule with affinity for an enzyme.</td>
</tr>
<tr>
<td>Oxidation</td>
<td>The loss of electrons or gain of oxygen in a reaction.</td>
</tr>
<tr>
<td>Reduction</td>
<td>The gain of electrons or loss of oxygen in a reaction.</td>
</tr>
<tr>
<td>Redox reaction</td>
<td>A chemical reaction, where both oxidation and reduction are going on, side by side.</td>
</tr>
<tr>
<td>Substrate</td>
<td>This is a molecule that can be chemically converted by an enzyme.</td>
</tr>
<tr>
<td>Wt</td>
<td>This is the typical most common form of an organism, gene or protein that occurs in nature.</td>
</tr>
</tbody>
</table>
1 Introduction to alcohol dehydrogenase (ADH)

1.1 A brief historical background: From the Big Bang to the discovery of ADH

The history of the universe is believed to have started, with a "Big Bang", estimated to have happened 12.5 billion years ago (Cayrel et al, 2001). As long as 8 billion years later our solar system was created (Allegre et al, 1995; Patterson et al, 1955). In the beginning our earth was probably a rough place not suitable for life and evolution. However, gradually things cooled down and spontaneous organic reactions eventually created small functional units, today called RNA (Szostak et al, 2001). Within a lipid envelope the RNA molecules were protected from the surroundings in a primordial cell (Deamer, 1997). Here, RNA could go through autonomous replication (Kruger et al, 1982). The primordial cell, precursor of the modern cell, slowly evolved and transformed our earth into the biological world of DNA, RNA and proteins in different species. Still, it would take billion of years before the first human species evolved (Hedges, 2002).

With the prehistoric human life forms, the beginning of a new kind of evolution started - the technical evolution (Vitousek et al, 1997). The first *Homo sapiens* on earth probably evolved from *Homo erectus* 250 thousand years ago in Africa (Wood, 1992). At first, the humans gathered food from hunting, fishing and collecting naturally growing vegetables, nuts and fruits (McBrearty & Brooks, 2000). Early on, *Homo sapiens* were probably capable of controlling fire and with this ability they could spread out to cold places, cook food, preserve food and scare away predators (Balter, 1995). When we eventually started to grow the grains of wild grasses and other vegetables, larger and more stable settlements were founded. Eight thousand years ago, the invention of the plough resulted in increased crops which made it possible to feed a larger population (Janick, 2007). Through improvements in agricultural techniques less people were needed on the farms and they could move into cities, which lead to further specialisation and economic growth.

The development of steam engines for usage in transport and industries, sparked the industrial revolution in England 250 years ago (Berg & Hudson, 1992). This technical
advance led to an enormous growth in productivity but it also increased environmental pollution and caused devastation through modern warfare. The most recent leap in the technical evolution is the development of programmable computers. The IT revolution has and will continue to change our lives in many ways.

The field of medicine has developed in parallel with the technical evolution. Humans have most likely always been interested in finding explanations for life, disease and death. The earliest signs of this interest are displayed in cave paintings dated 12 000-14 000 BC (Valladas et al, 1992). The rapid technical development in the last century has also revolutionized our knowledge about, and possibilities to treat, diseases. The field of molecular biology sprung out of the molecular identifications of RNA, DNA and proteins during the last century. The central dogma of biology was first coined by Francis Crick in 1958 (Crick, 1970). This theory states the direction of information in biology as going from DNA to RNA by a transcription process and from RNA to protein by a translation process. For reproduction of the organism a replication process duplicates the DNA. The human genome contains three billions of nucleotides and complete mapping of the genome was finished only a few years ago.

In 1890, Emil Fischer, described an early mechanistic model for enzymatically catalysed reactions, and the pervasive concept of the “lock and key” model was born (Lemieux & Spohr, 1994). Another important model for enzymes was the Michaelis-Menten function, proposed in 1913 by Leonor Michaelis and Maud Menten, which describes the catalytic reaction rate relative to a substrate concentration.

The first ever isolated alcohol dehydrogenase (ADH) was purified in 1937 from Saccharomyces cerevisiae (baker’s yeast) (Negelein & Wulff, 1937). Many aspects of the catalytic mechanism for the horse liver ADH was investigated by Hugo Theorell and co-workers (Theorell & Mckinley-Mckee, 1961a; Theorell & Mckinley-Mckee, 1961b). ADH was also the first ever oligomeric enzyme for which an amino acid sequence and a three-dimensional structure were determined (Eklund et al, 1974; Jörnvall & Harris, 1970). Following the rapid advances in DNA techniques, isolation of proteins, recombinant protein expression and molecular characterisation, a large number of different ADHs have been isolated and characterised. Genetic evidence from comparisons of multiple organisms showed that a glutathione dependent formaldehyde
dehydrogenase, identical to an ADH, probably is the ancestral enzyme for the entire ADH family (Danielsson & Jörnvall, 1992; Persson et al, 2008). Early on in evolution, an effective method for eliminating endogenous and exogenous formaldehyde was important and this capacity has conserved the ancestral ADH through time. From genetic duplications of this ADH, followed by series of mutations, the other ADHs evolved (Danielsson & Jörnvall, 1992; Persson et al, 2008).

The ability to produce ethanol from sugar is believed to have evolved in yeast. This feature is not rational from an energetic point of view, but by making alcohol in such high concentrations so that they were toxic to other organisms, yeast cells could effectively eliminate their competition. Since rotting fruit can contain more than 4% of ethanol, animals eating fruit needed a system to metabolize exogenous ethanol. This can explain the presence of an ethanol active ADH in other species than yeast. This ethanol active ADH in humans has different allele types around the world, which could be a consequence of different drinking habits.

1.2 Medium-chain dehydrogenase/reductase (MDR) proteins

The medium-chain dehydrogenase/reductase (MDR) family of proteins includes ADH, sorbitol dehydrogenase (SDH), leukotriene B4 dehydrogenase, quinone reductase and many other enzymes (Persson et al, 1994). The MDR protein family can be further divided into eight large superfamilies (>100 members) in vertebrates, of which ADH is the most studied one (Brändén et al, 1973; Gonzalez-Duarte & Albalat, 2005; Jörnvall et al, 2000; Nordling et al, 2002).

The total amount of genomic and protein data has increased exponentially in the last years. As a consequence, the number of MDR sequences in the database UNIPROT was multiplied ten times, to a total of 11,000, in a period of five years (Persson et al, 2008). The total number of human MDR proteins is now determined to be 25, which corresponds to 0.1% of the total number of genes in the human genome (Persson et al, 2008). Transcribed and further translated MDR genes result in proteins of 350 to 400 amino acid residues per subunit. The wild type (wt) quaternary structure can vary between different proteins.
Although the general sequence identity is about 25% between the different MDR protein superfamilies, several important structural and functional properties are conserved within all the MDR protein superfamilies, e.g. the specificity for the coenzyme NAD(P), two separate domains, the existence of a catalytic zinc site and a similar topology for the subunit structure.

1.3 The MDR-ADH protein superfamily

In this thesis, the MDR-ADH protein superfamily is referred to as simply ADH. The ADH family can be traced back to an ancestral form of ADH3. The ancestral ADH3 gene was duplicated, and then a series of mutations followed, which resulted in the multiple forms of ADHs, which are divided into several classes, isoenzymes and isoforms (Jörnvall, 1985; Jörnvall & Höög, 1995; Vallee & Bazzone, 1983). ADHs have been identified in most life forms, including bacteria, archaea, plants, yeast and animals. This widespread distribution is due to ADHs general protective role against different toxic compounds attacking the cell from both endogenous and exogenous sources.

ADH can catalyze an oxidative reaction of a hydroxyl group or a reductive reaction of an aldehyde or keton. ADHs serve as important enzymes in detoxification of particularly primary and secondary alcohols or aldehydes. This is a reduction-oxidation (redox) reaction, where the substrate reduction/oxidation is synchronized with the reduction/oxidation of the coenzyme NAD+/NADH or to a low extent NADP+/NADPH (Peralba et al, 1999; von Wartburg et al, 1964). Important known substrates are ethanol, hydroxymethylglutathione and different retinoids (Danielsson et al, 1994b; Duester, 2000; von Wartburg et al, 1964).

The nomenclature of ADH is not consistent in the literature and under debate. In this thesis, the latest suggested nomenclature will be used, as in the table below (Duester et al, 1999; Jörnvall, 2008). Thus, in the reference list and in the original articles the names of the ADH enzymes, isoenzymes and genes do vary.
New and old nomenclature for human alcohol dehydrogenase:

<table>
<thead>
<tr>
<th>Class</th>
<th>New nomenclature</th>
<th>Old nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>ADH1A ADH1A</td>
<td>α ADH1</td>
</tr>
<tr>
<td></td>
<td>ADH1B ADH1B</td>
<td>β ADH2</td>
</tr>
<tr>
<td></td>
<td>ADH1C ADH1C</td>
<td>γ ADH3</td>
</tr>
<tr>
<td>II</td>
<td>ADH2 ADH2</td>
<td>π ADH4</td>
</tr>
<tr>
<td>III</td>
<td>ADH3 ADH3</td>
<td>χ ADH5</td>
</tr>
<tr>
<td>IV</td>
<td>ADH4 ADH4</td>
<td>σ or μ ADH7</td>
</tr>
<tr>
<td>V</td>
<td>ADH5</td>
<td>ADH6</td>
</tr>
</tbody>
</table>

- Based on nomenclature recommended in (Duester et al, 1999).
- Based on nomenclature recommended in (Jörnvall & Höög, 1995).
- Old nomenclature for a single protein subunit.

Another matter of some confusion is that conserved amino acid residues which directly bind prosthetic groups and ligands, can differ between different ADHs. Horse liver ADH was the first ADH with a determined protein sequence, thus many of the early paper used this sequence numbering system as a standard. However, since ADH3 is believed to be the ancestral type of ADH, this protein sequence could therefore be a better alternative to be used as the standard alignment of amino acid residues. In this thesis, unless stated explicitly, amino acid residues are given their sequence number according to the human ADH3 sequence (initiator Met not included).

1.4 Human ADH1 to ADH5

Five different classes of ADHs (ADH1 to ADH5) exist in humans. In mammals one more class (ADH6) and in vertebrates two additional classes (ADH7 and ADH8) have been identified, making the total number of eight classes (ADH1 to ADH8) in vertebrates.
1.4.1 ADH1
ADH1 is the major enzyme for conversion of ethanol and it is highly expressed in the liver. ADH1 is missing in invertebrates, but is ubiquitously expressed in vertebrates. This and other data suggests that the first duplication event within the ADH superfamily can be dated to approximately 500 million years ago (Canestro et al, 2000; Jörnvall et al, 2001). The three isoenzymes of ADH1 found in primates, evolved later. These isoenzymes correspond to three different gene loci, ADH1A, ADH1B and ADH1C. The different isoenzymes exist as homo or heterodimers in vitro. In addition, several alleles of ADH1B and ADH1C exist. The ADH1B*2 phenotype has a decreased risk of developing alcoholism, as much as eight times lower than for the ADH1B*1 phenotype (Chen et al, 1999). This can be explained by its more rapid conversion of alcohol to acetaldehyde, a molecule with several unpleasant side effects (Lee et al, 2004). In addition, a number of mutations have been found in the non-protein coding parts of the different ADH1 genes, in line with the other ADHs. ADH1 isoenzymes have a broad substrate spectrum and have therefore been suggested to have a physiological role for conversion of ethanol, retinoids (Yang et al, 1994), ω-hydroxylated fatty acids (Björkhem, 1972), norepinephrine and dopamine metabolites (Mårdh et al, 1985; Mårdh & Vallee, 1986), serotonin metabolites (Consalvi et al, 1986), bile acids (Marschall et al, 2000; Okuda & Okuda, 1983) and steroids (McEvily et al, 1988).

1.4.2 ADH2
ADH2 was discovered through electrophoresis analysis of human livers and subsequent measurement of its catalytic activity (Li & Magnes, 1975), which differed from ADH1. Early on, several differences were noted in stability and substrate specificity between species (Bosron et al, 1979; Ditlow et al, 1984; Höög et al, 1999). Large differences in substrate specificity and enzymatic activity have been observed between rodent and human ADH2, which actualises the question about the physiological role of ADH2. ADH2 is highly expressed in the liver and intestine (Estonius et al, 1993). ADH2 harbour a broad substrate specificity, i.e. different retinoids (Hellgren et al, 2007; Popescu & Napoli, 2000; Yang et al, 1994), medium to short chain alcohols/aldehydes, 4-hydroxyalkenals (Sellin et al, 1991) and ω-hydroxy fatty acids (Ditlow et al, 1984).
1.4.3 ADH3

ADH3 was also discovered through electrophoresis analysis of human livers and subsequent measurement of its catalytic activity (Pâres & Vallee, 1981). Genetic data, protein expression patterns and conserved substrate specificity suggest that ADH3 is the ancestral form of ADHs (Danielsson et al, 1994b; Persson et al, 2008). ADH3 is also known both as a glutathione-dependent formaldehyde dehydrogenase and a glutathione-nitricoxide reductase (Jensen et al, 1998; Koivusalo & Uotila, 1974). ADH3 shows a low catalytic activity for ethanol and retinol but has a high activity in oxidative detoxification of formaldehyde in the presence of GSH, a capacity that was detected early (Strittmatter & Ball, 1955). GSH and formaldehyde will spontaneously form S-hydroxymethylglutathione (HM GSH). This reaction has an equilibrium constant of 1.77 mM at pH 7.5 (Sanghani et al, 2000). ADH3 has been suggested to be the major enzyme in formaldehyde metabolism (Hedberg et al, 2001b; Koivusalo et al, 1989) due to its high enzymatic specificity towards HM GSH, corresponding to a $K_m$ over $k_{cat}$ of 1.3 million (mM$^{-1}$min$^{-1}$). In the reductive pathway, ADH3 has a high specificity towards GSNO (Jensen et al, 1998). GSNO affects the cellular equilibrium between protein S-nitrosothiols and GSNO, which will change the amount of protein S-nitrosylation in a cell and thereby regulate cell signalling (Haqqani et al, 2003; Jensen et al, 1998; Que et al, 2005; Staab et al, 2008).

The evolutionary mutation rate of the coding regions in ADH3 has been estimated to be significantly lower than in the other ADH classes (Gonzalez-Duarte & Albalat, 2005), as was shown earlier at the protein level (Jörnvall et al, 1987). This is in line with the hypothesis of a conserved enzymatic function for ADH3 (Danielsson et al, 1994a). ADH3 seems to be ubiquitously expressed in most organs in vertebrates (Estonius et al, 1993), indicating that it has a role as a “housekeeping” enzyme (Funkenstein & Jakowlew, 1996). There is, however, some contradictory data that point out that the activity of ADH3 in different mammalian tissues, can vary up to a thirty fold (Gonzalez-Duarte & Albalat, 2005). ADH3 is the only ADH which has been identified in the nucleus (Iborra et al, 1992), indicating its protective role against formaldehyde toxicity, probably from interactions of formaldehyde with the chromatin (DNA and histones). In addition to its crucial role in the formaldehyde and NO metabolism, ADH3 has also been
suggested to be involved in the metabolism of retinoids and first pass metabolism of ethanol (Duester et al, 1991; Lee et al, 2006).

1.4.4 ADH4
ADH4 is an enzyme with a high $K_m$ and $k_{cat}$ for ethanol. It has been suggested to have a role in the first pass metabolism of ethanol because its $k_{cat}$ value for ethanol is the highest among all human ADHs. ADH4 is an extrahepatic enzyme, nicknamed “gastric ADH” since it is highly expressed in epithelial cells in the ventricle and gastro-intestinal tract (Moreno et al, 1996; Parés et al, 1992). Nonetheless, it has also been identified in epithelial cells in many other tissues. ADH4 is the most efficient retinol dehydrogenase among human ADHs, and it is suggested to play a role in retinoid metabolism, especially during fetal development (Duester et al, 1997).

1.4.5 ADH5
ADH5 has only been identified at the gene and mRNA levels, and no functional role has been established. The transcription pattern of ADH5 results in three different mRNAs. One mRNA codes for a truncated variant of ADH (missing eight amino acid residues at the C-terminus) and the two other forms code for variants similar to the other classes of ADH (Strömberg & Höög, 2000).

1.5 Tetrameric ADHs
Besides the dimeric form, several tetrameric ADHs have been identified, although not in vertebrates. The formation of a tetrameric structure in ADH has probably evolved after mutations that created a favourable subunit interface for the formation of a “dimers of dimers” complex (Banfield et al, 2001). Examples of tetrameric ADH structures are found in yeast and thermostable bacteria, such as Thermoanaerobacter brokii and Clostridium beijerinckii (Karlsson et al, 2003; Korkhin et al, 1999). The structural zinc site is conserved only in certain tetrameric ADHs. Thus, the site is not a definite determinant for the ADHs preference of folding into a dimeric or a tetrameric quaternary structure.
1.6 Sorbitol dehydrogenase

SDH is also a zinc-dependent enzyme within the protein family of MDRs, with a positional sequence identity of 24 % to the closely related human ADH1. The molecular mass per subunit is 38 000 Da and the human SDH consists of 356 amino acid residues (Johansson et al, 2001; Karlsson et al, 1991). SDH is expressed at the mRNA level in almost all tissues, except in the gastro-intestinal tract (Estonius et al, 1993). SDH converts secondary alcohols of polyols with NAD as coenzyme (Jeffery et al, 1981). SDH has no or a very low activity, with primary alcohols as substrates (Jeffery et al, 1981; Lindstad & McKinley-McKee, 1993; Maret & Auld, 1988).

Mammalian SDH has a catalytic zinc ion but lacks the structural zinc site that is found in mammalian ADHs. Silverleaf whitefly SDH and yeast ADH constitute an evolutionary bridge between ADH and SDH where both have a structural zinc site and a tetrameric quaternary structure (Banfield et al, 2001; Karlsson et al, 2003).

The interaction between the subunits in SDH was investigated early (1985) in a computer modelling study and later revealed in three-dimensional structure determinations of rat and human SDHs (Eklund et al 1985; Johansson et al 2001; Pauly et al 2003). These studies concluded that two dimers, as in mammalian ADH, form a tetramer in SDH. Still, several details on the interactions have not been determined and some of these were investigated in Paper II.

The illustration to the left shows the tetrameric quaternary structure of SDH. The subunits, corresponding to the dimers in ADH, are displayed in yellow and green colours. The zinc atoms are blue and the residue Tyr110, crucial for the tetrameric quaternary structure, has been displayed in red.
1.7 The central dogma of ADH

The central dogma in biology describes the transformation process from genetic information into a functional protein. Below, this transformation is described for ADH at structural and functional levels.

1.7.1 Genes and protein sequences
Throughout time, genes have evolved by duplication, modifications (mutations) and recombination. This process has given each protein a unique trait which can be studied by comparing gene families and their appearances and sequences in different species.

The human ADH genes are found at chromosome four in the region q21-25, aligned in a cluster of genes, as follows: ADH4-ADH1C-ADH1B-ADH1A-ADH5-ADH2-ADH3 (ADH4 is the closest to and ADH3 is the most distant from the centromere) (Gonzalez-Duarte & Albalat, 2005). The human ADH genes consist of 9 exons and 8 introns and in other vertebrates there are examples of up to 11 exons and 10 introns (Duester et al., 1986; Gonzalez-Duarte & Albalat, 2005). The protein sequence length of an ADH subunit varies, the shortest being 373 amino acid residues as in human ADH3, up to 383 amino acid residues as in C. elegans ADH3 (Gonzalez-Duarte & Albalat, 2005).

The ADHs have roughly 60% protein sequence identity between different classes and the isoenzymes have more than 90% sequence identity within each class (Duester et al., 1999; Jörnvall, 2008). Polymorphism has been identified in both promoter regions and in different exons for the isoenzymes ADH1A, ADH1B and ADH1C as well as in the diverse classes ADH2, ADH3 and ADH4. The two isoenzymes ADH1B and ADH1C have been shown to be correlated to drinking behaviour and risk for developing alcoholism (Chen et al., 1999). Polymorphism in the ADH4 promoter region has been linked to neurodegenerative disorders such as Parkinson’s disease (Buervenich et al., 2000).
1.7.2 Protein expression patterns

The protein subunit of human ADH has a molecular mass of approximately 40 000 Da. The protein concentration of a particular protein type in a cell is determined by the rate of translation of its mRNA and the rate of its degradation. The rate of transport in and out of a cell will also alter the protein concentration. Protein expression, transport and degradation are regulated in various ways.

In general, the concentration of most types of proteins varies between species, phenotypes, organs, cell types, every cell and in cellular compartments. The protein concentrations in one cell can also vary with time. These variations make an experimental determination of a specific protein concentration within a certain cell type, cell batch or organ, often somewhat imprecise. ADH is an intracellular enzyme which is found in several of the cellular compartments, e.g. in the nucleus and the cytosol. ADH enzymes are water soluble and thus probably not attached to the lipid membranes surrounding cellular compartments. By being soluble, ADHs can diffuse within different compartments.

Often mRNA expression analysis within cells are used to determine if a protein is expressed or not, even though the amount of mRNA is sometimes not linearly correlated to the protein concentration. This lack of direct translatability between concentrations of mRNA and of protein, is due to the many biological processes that are involved in protein translation and degradation. One direct method for determining protein levels, is to use antibodies, given that the antibody binds specifically to only the sought-after protein. From antibody and mRNA determinations, the proteins ADH1, ADH2, ADH3 and ADH4 were found to be expressed in the liver (Estonius et al, 1996). The ratio between ADH1B:ADH1C:ADH2 were determined to be 1:0:0:3:0:7 (Lee et al, 2004). ADH1 (ADH1A) is expressed in a majority of cell types as was derived from looking at immunohistochemistry images for about 70 cell types (Nilsson et al, 2005). It was found in a particularly high concentration in the glandular cells of the colon, duodenum, appendix, gallbladder, rectum, stomach and in hepatocytes and alveolar cells. ADH3 is ubiquitously expressed in all investigated tissues within mammals, but compared to the other ADHs, to a lesser degree in the liver (Estonius et al, 1996). ADH4 is expressed
in roughly 10% of the different cell types, primarily in epithelial cells throughout the gastro-intestinal tract, but also to a smaller extent in hepatocytes, myocytes, Leydig cells and glandular cells in the prostate (Nilsson et al, 2005). Northern blot analysis from mRNA expression studies showed that mRNA for ADH4 could only be observed in the stomach, indicating that the antibody does not bind mutually exclusively to ADH4 (Estonius et al, 1996).

1.7.3 Tertiary structure

In line with other soluble proteins, ADH has a hydrophilic surface mostly consisting of polar amino acid residues which form hydrogen-bonds with the surrounding water molecules. The interior of the tertiary structure has, compared to the surface, a relatively low amount of hydrophobic amino acid residues (Eklund et al, 1974; Jörnvall et al, 1978). The strict placement of hydrophobic and hydrophilic amino acid residues seems to be optimal for minimizing the solvation free energy for the system and thus it contributes to correct folding and stabilization of the tertiary structure (Jaenicke, 1996; Nicholls et al, 1991).

The image shows ADH3 solvated in water. It was generated from a snapshot during a MD simulation with GROMACS.

The topology of a single subunit in ADHs is divided into two domains which are denoted the catalytic and coenzyme binding domains. The catalytic pocket is centred in the middle of the subunit, between the catalytic and coenzyme binding domains. ADHs are zinc dependent enzymes. They have two zinc atoms per subunit, bound to amino acid residues in the catalytic domain. In 1973, the first structure of alcohol dehydrogenase
(Brändén et al, 1973) was resolved (liver ADH) and now as many as 82 deposited structures of ADHs, with the EC number 1.1.1.1 exist, in the RCSB database (June 2009). The protein structures of human: ADH1A, ADH1B, ADH1C, ADH2, ADH3 and ADH4, have all been determined with x-ray crystallographic method. The structure of human ADH5 is still missing.

The x-ray crystallography determined ADH1 and ADH3 structures, in both a holo and apo state, revealed that the binding of the coenzyme in ADH3 causes the catalytic domain to rotate (Ramaswamy et al, 1999). This closes the binding cleft between the two domains and thus narrows the active site pocket for the substrate, making it possible for the substrate to bind. These dynamical motions has also been studied with neutron spin-echo spectroscopy, which revealed a collective motion of domains due to the binding and release of the cofactor (Biehl et al, 2008).

It is important to understand that protein folding, binding of ligands and catalysis of substrates is dynamic processes which alters the protein structure. Thus, no single protein structure or conformation of ADH can reveal all of its properties. However, the large amount of determined ADH enzymes from different species and classes in different states can reveal many important links between the ADH enzymes and their catalytic function.

1.7.4 Active site
Between the catalytic and coenzyme binding domains there is a hydrophobic cleft that forms the active site (Eklund et al, 1976). The depth and width of the substrate pocket vary between different classes. Where ADH3 has the widest active site pocket and thus favours the binding of larger substrates and ADH1 with its smaller substrate pocket can better accommodate smaller substrates, such as ethanol. The catalytic domain seems much more disordered than the coenzyme domain, several windings back and forth and a secondary structure which contains only a few helical structures.

The inner part of the active site consists of the active site zinc, which is crucial for the catalysis reactions. The zinc binding areas of the catalytic domain is located toward
the N-terminal side. The active site zinc is bound by two cysteine residues and one histidine residue (Brändén et al, 1973). Zinc ions in proteins or enzymes are most often coordinated by four different atoms. The fourth ligand which binds to the active zinc ion is probably a water molecule. The water molecule is replaced by a substrate hydroxyl oxygen atom, prior to the catalyzing reaction. There has also been suggested that both the substrate and a water oxygen could coordinate the zinc ion, creating a penta coordinated active site zinc (Vononciul & Clark, 1993). This, however, is a coordination which seems unlikely as learned from a QM/MM combined study (Ryde, 1995).

Image of ADH3 with the substrate GSNO, zinc atoms and the coenzyme NAD bound to the enzyme. NAD (both sides) and GSNO (left side) has a yellow surface marking the active site.

The same image as above but with all protein atoms displayed explicitly. This shows the complexity and multiplicity of atomic interactions in macromolecules, such as ADH.
The coenzyme domain in ADHs has a similar topology as in other dehydrogenases. It consists of a Rossman fold with two separate binding halves for the two mononucleotide parts of the coenzyme (Eklund & Ramaswamy, 2008). This part is located on the C-terminal side.

1.7.5 Structural zinc site
The structural zinc ion is bound to a zinc binding motif consisting of four cysteine residues, Cys96, Cys99, Cys102 and Cys110, which form an almost tetrahedral coordination for the zinc ion (Brändén et al, 1973; Ryde, 1996). This zinc site is labelled a structural one because it has been shown to be of importance for the folding and stability of the enzyme. These are properties which also indirectly effect the catalytic activity of the enzyme (Jelokovà et al, 1994).

1.7.6 Quaternary structure
All mammalian ADHs have a dimeric structure. In bacteria, fungus and yeast a tetrameric structure in ADH is common. Recent findings also revealed a thermostable ADH which was a multimer, consisting of twelve subunits (Hess & Antranikian, 2008). Both the structural and catalytic zinc are probably important for the folding, but details on the folding process are difficult to determine (Karplus & Kuriyan, 2005). The main interaction area between the two subunits, is between amino acids 299-316 in ADH3, which has an antiparallel structure. This part is rich in hydrophobic amino acid residues and it forms hydrogen-bonds between the subunits, which, in turn, folds into a β sheet (Eklund & Ramaswamy, 2008).
This image shows the amino acids residues at the interface between the two subunits (A and B) of ADH3. The displayed amino acid residues are 299-316. Subunit A is coloured in pink and B is coloured in green. The top left picture is rotated 45 and 90 degrees in two different directions to illustrate the interaction from different angles.

The rate of association and dissociation between subunits determines the quaternary structure and stability. The following image shows the assembly of the four subunits (A, B, C and D) in tetrameric SDH and ADHs, Paper II.
1.8 The metabolism of endogenous and exogenous substrates

The chemically catalysed reaction with ADH does not directly produce reactive oxygen species (ROS), as is the case with cytochrome P450 (CYP) catalyzed reactions (Fleming et al., 2001; Wrighton & Stevens, 1992). The catalytic action of CYPs can potentially be very harmful since ROS can induce genetic damages and initiate apoptosis (Bondy & Naderi, 1994). However, the ADH oxidation reaction of an alcohol can also be harmful since the reduction produces an aldehyde which often is more toxic due to a higher spontaneous reactivity, than the alcohol. But at normal alcohol levels the aldehyde is continuously metabolized to the corresponding carboxylic acid. The metabolism of most alcohols by ADH is coupled to second step with ALDH, a general metabolic feature which is of importance for the breakdown of a large number of endogenous and exogenous compounds in humans.

The general picture of the coupled metabolism of ADH and ALDH enzymes is given in the following picture.

In the metabolism of alcohols, in the above scheme, additionally two coenzymes (NAD(P)) are reduced and two or three protons are released. The first proton is released in the reaction from alcohol to aldehyde and the second proton in the reaction from aldehyde to carboxylic acid. A third proton can also be released, however, this depends on the pH of the solution and the pKa for that specific carboxylic acid.

Different classes and isoenzymes of ADH and ALDH have varying specificities for different types of substrates. Nevertheless, for substrates such as ethanol or retinol, several classes and isoenzymes of ADH and ALDH have an overlapping complementary specificity.
1.8.1 Catalytic activities
The catalytic reaction rate by ADH depends on the rate of binding and release of substrates and products. Where the transition state free energy barrier for the chemical reaction needs to be overcome. Human ADH1, ADH2 and ADH4 have a kinetic mechanism that follows an ordered bi-bi mechanism for the binding and release of substrate and cofactor (Brändén et al, 1973), unlike human ADH3, which follows a random bi-bi mechanism (Sanghani et al, 2000). Both non-covalent interactions and formation or breakage of covalent bonds are important to consider for understanding the lowering of the transition state energy barrier in the catalysed reaction and for the binding specificities of the ligands. Since breakage or formation of covalent bonds is ruled by quantum mechanics (QM), methods for determination of these effects are important to use. Theoretical calculations with QM methods and kinetic experiments with deuterium isotopes showed that hydrogen tunnelling is important in ADH (Agarwal et al, 2000; Billeter et al, 2001; Ramaswamy et al, 1999). QM calculations also showed that the ring puckering of the coenzyme NAD⁺ are an important factor for the hydride transfer (Agarwal et al, 2000; Billeter et al, 2001). The mechanism for a complete circle of an ordered bi-bi catalyzed reaction of an alcohol by ADHs include: binding NAD⁺ to the enzyme, binding of the alcohol substrate, deprotonation creating an alcoxide ion, hydride transfer from the alcoxide ion to NAD⁺ leading to NADH and aldehyde, release of aldehyde, release of coenzyme NADH. However, in general, there are several possible intermediate states and catalytic transition states. These catalytic pathways and transition states are in general, difficult to determine and study.

The intracellular pH in most cells is around 7.2 (Srivastava et al, 2007). Small pH changes alter many cellular processes, including proliferation, migration, and transformation and a pH lowered by 0.3-0.4, is a sign of an apoptotic cell (Lagadic-Gossmann et al, 2004; Srivastava et al, 2007). This tight regulation of the pH is maintained by high concentrations of several types of molecules which can easily accept or donate hydrogen atoms at levels around physiological pH. Important physiological buffer molecules which can accept or donate protons, in accordance to their pKₐ values, are GSH, GSSG, albumin, hemoglobin and specifically histidine residues in proteins. However, during ethanol ingestion the pH can be lowered by protons released in the
reaction from ethanol to acetaldehyde and the following conversion of acetaldehyde to acetate. This can reduce the pH in hepatocytes (Carini et al, 2000). A lowered pH changes the chemical equilibrium for many reactions and protonation of amino acid residues. This is known to alter the function of many proteins and enzymes. In general, the catalytic reaction for ADHs is increased at higher pH (pH optimum around 10), and lowered by pH below the physiological pH. This lowered pH, due to ethanol ingestion, could serve as a protective mechanism, since it reduces the catalytic turn-over rates for ethanol by ADH, which in turn reduces the release of protons.

In the next section, the metabolism of the substrates ethanol, retinol, HM GSH and GSNO are described. These four substrates are important and specific for different human ADHs, with a certain described physiological function. With paper II in mind, the metabolism of the sugar alcohol sorbitol by SDH is included.

1.8.2 Ethanol
Ethanol is a small alcohol (46 Da) which is easily soluble in water because the two atoms in the hydroxyl group are connected by hydrogen-bonds with the water molecules.

Humans can endure an ingestion of several mol of ethanol each day, but if the blood concentration reaches to about 4 per mil (~90 mM) this leads, for most individuals, to a condition of unconsciousness and respiratory arrest. Ethanol rather easily passes the cell membrane and enters into all body fluids, intracellular compartments and the blood-brain barrier.

The enzymes involved in the metabolism of ethanol are catalysis by different ADHs in the first step reaction followed by catalysis of different ALDHs in a second reaction step (Edenberg, 2007).

$$\text{Ethanol} \xrightarrow{\text{Catalysis ADH}} \text{Acetaldehyde} \xrightarrow{\text{Catalysis ALDH}} \text{Acetic acid}$$

A part of the ingested ethanol is metabolized by a so called first pass metabolism, which is performed by the epithelial cells lining the stomach, and these cells have a high
concentration of ADH4 in their cytosol. However, the vast majority of the ingested ethanol is metabolized in the liver by different isoenzymes of ADH1. Other enzymes that are also believed to participate in the metabolism of ethanol in the liver, but to a smaller extent due to their higher $K_m$ values are; ADH2, ADH3 and CYP2E1. However, even if the contribution to metabolism of ethanol from CYP2E1 is small, the physiological effect from heavy drinking results in an induction of CYPs. This in turn alters the general metabolism of many other chemicals in primarily the liver. The increase in CYPs activity will lead to an increase of ROS in the cells (Bailey & Cunningham, 1998; Bondy & Naderi, 1994).

Ethanol is converted to acetaldehyde in the cytosol. The acetaldehyde is then converted to acetic acid by primarily ALDH2 in the inner mitochondria membrane matrix of the hepatocytes (Quintanilla & Tampier, 1992). Acetic acid is a substrate which can enter into the citric acid cycle. This metabolic pathway oxidizes acetic acid to carbon dioxide and water through several steps.

Ethanol drinking can affect the function and metabolism in many cells and organs and it has been linked to numerous syndromes and diseases. Prolonged heavy drinking induces organic damage to the heart, brain, liver, kidney and gastrointestinal organs. Nevertheless, the prognosis and recovery is very good for those who can stop abusing ethanol (Spanagel, 2009). In Sweden, somewhere between 300 000 to 500 000 people are believed to abuse ethanol, and worldwide, ethanol drinking is estimated to cause 3.2 % of all deaths (WHO, 2004).

1.8.3 Formaldehyde
Formaldehyde is the smallest of all carbonyl molecules (31 Da). Its reactivity is potentially harmful under normal cellular condition which has been demonstrated with in vitro experiments showing that formaldehyde induces multiple kinds of DNA damages. In addition, epidemiological studies has pointed out the carcinogenic effects of formaldehyde (Wilbourn et al, 1995). The catabolism of the amino acid residues serine and glycine, seem to be the largest endogenous contributors of formaldehyde in humans. Humans are exposed to exogenous sources of formaldehyde from the burning of carbon
fuels, in certain foods and in cosmetic preservatives (Wilbourn et al, 1995).

Formaldehyde reacts spontaneously with many molecules. GSH, found intracellularly in millimolar concentrations, reacts non-enzymatically with formaldehyde, to form HMGSH (Forman et al 2002; Staab et al 2008). HMGSH can then be converted by ADH3 to S-formylglutathione. This is a reaction scheme for the elimination of formaldehyde performed by ADH3.

Formaldehyde + GSH $\xrightarrow{\text{Spontaneous reaction}}$ HMGSH $\xrightarrow{\text{Catalysis ADH3}}$ S – formyl glutathione

In parallel to this route, formaldehyde can be converted by cytosolic ALDH1 and mitochondrial ALDH2 enzymes (Wang et al, 2002).

1.8.4 Nitric oxide
Nitric oxide (NO) is a small and reactive molecule (30 Da) formed as burning of combustion engines, cigarettes and lightening storms. The atmospheric concentration of NO varies between ten to five hundred parts per billion. Physiologically NO is formed from arginine by different NO synthases. NO can be toxic, but NO will also serve as an important cell signalling molecule at normal cellular concentrations (Droge, 2002). NO mediates regulation of vascular tone, cell death, cell signalling from membrane receptors and oxygen tension thorough control of the erythropoietin production (Droge, 2002; Ignarro et al, 1987; Palmer et al, 1987; Radomski et al, 1987).

NO has under normal cellular conditions different number of electrons, NO$^-$, NO$^-$ and NO$^+$, and these various electron configurations have different physiological effects (Bin-Nun & Schreiber, 2008). NO can also easily convert into reactive non radical species from both enzymatic and non enzymatic reactions, either directly or from subsequent reactions. NO reacts spontaneously with glutathione forming GSNO (Gow & Stamler, 1998). Depletion of GSNO can be catalyzed by ADH3 at a fast kinetic rate (110,000 mMol$^{-1}$ min$^{-1}$)(Staab et al, 2008). The reaction is in vitro inhibited by GSH at its physiological concentration (Staab et al 2009).

NO + GSH $\xrightarrow{\text{Spontaneous reaction}}$ GSNO $\xrightarrow{\text{Catalysis ADH3}}$ S – hydroxylaminoglutathione
The normally low intracellular NADH concentration, could question the depletion of GSNO by ADH3. However, a recent study showed that the parallel oxidation of HMGSH by ADH3 could serve as an NADH source for depleting of GSNO (Staab et al, 2008). Inhaled NO is a frequently used drug for treatment of respiratory failure and persistent pulmonary hypertension (Bin-Nun & Schreiber, 2008). This positive NO mediated effect on obstructive lung conditions could potentially be increased, if the NO depletion by ADH3 were inhibited.

The picture shows ADH3 with the coenzyme and substrate GSNO. To the right is a close-up view of the interaction.

1.8.5 Vitamin A

All vitamins are essential for survival. Vitamin A is commonly described as all the different active forms of retinoids in humans. Thus, vitamin A can be an alcohol, aldehyde or carboxylic acid in different configurations (cis or trans). Meat and certain vegetables (pro-vitamin A) are rich in vitamin A. The liver is the organ with the highest amount of retinoids in humans, and the major part is stored as retinyl esters (Biesalski & Nohr, 2004).

In a recent study, the concentrations of Vitamin A from chicken livers were determined to be between 0,23 \( \mu \text{M} \) to 0,29 \( \mu \text{M} \) (Santos et al, 2009). Thus, the alcohol and aldehyde retinoic concentration are in the low \( \mu \text{M} \) range.

Retinoids are light sensitive molecules, which mean that they can be excited by absorption of a photon; this electron excitation induces a configuration change of the molecule, between a trans and a cis configuration. In vertebrates this property is utilized for eye sight. Further, retinoic acids have an important physiological function as nuclear
transcription factors. The retinoic acids can bind to retinol binding factors within the cell nucleus, these complexes activates or enhances the transcription of different genes.

Since both a too high and too low retinoid level within the body is harmful, the metabolism needs to be regulated. The metabolism of ingested retinol to retinoic acid follows two steps: retinol is first oxidized to retinaldehyde, which in a second step is oxidized to retinoic acid. The enzymes participating in the first step, also the rate-limiting step, belongs to the ADH, short-chain dehydrogenases/reductases (SDR) families and aldo-keto reductases (AKR) (Crosas et al, 2003; Duester, 2000; Gallego et al, 2006; Jörnvall et al, 2000).

\[
\text{Retinol} \xrightarrow{\text{Catalysis ADH, SDR, AKR}} \text{Retinal} \xrightarrow{\text{Catalysis ALDH}} \text{Retinoic acid}
\]

In a comparative study with a sensitive method for determining of \(K_m\)-values different enzymes with a retinol catalytic activity were compared. The results from Paper I that utilized the same method are also included in the table (Gallego et al 2006).

<table>
<thead>
<tr>
<th>All-trans-Retinol</th>
<th>ADH1B1</th>
<th>ADH1B2</th>
<th>ADH2</th>
<th>ADH4</th>
<th>AKR1B10</th>
<th>RDH11</th>
<th>RoDH-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_{\text{cat}})</td>
<td>2.0</td>
<td>21.0</td>
<td>4.0</td>
<td>190.0</td>
<td>7.2</td>
<td>11.0</td>
<td>1.2</td>
</tr>
<tr>
<td>(K_m)</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>1.1</td>
</tr>
</tbody>
</table>

\(a\) RDH11 and RoDH-4 belongs to the SDR family

In the second and last step, the retinoic acid formation is catalyzed by cytosolic ALDHs, which oxidizes retinaldehyde to retinoic acid (Sirbu & Duester, 2006).

1.8.6 Sorbitol and the polyol pathway

Polyols contain multiple hydroxyl groups and the name is associated with the reduced product of a sugar or similar compounds. The polyol pathway is linked to the pentose phosphate pathway and glycolysis. This pathway is more active at high intracellular glucose concentrations. The first step in the polyol pathways is the conversion of D-glucose to sorbitol by the enzyme aldose reductase. The second step is the conversion of sorbitol to fructose by the enzyme SDH (Jeffery & Jörnvall, 1988). Aldose reductase and SDH have been shown to be coexpressed in several cell lines (Hedberg et al, 2001a), and
have been widely studied in relation to vascular dysfunction and complications due to diabetes (Oates, 2002).

1.8.7 Coenzyme NAD
The main pathway for nicotinamide adenine dinucleotide (NAD) synthesis in humans is from ingestion of nicotinic acid and nicotinamide (collectively called niacin) or from the recovery of breakdown, of NAD similar molecules. However, tryptophan can be utilized for de novo production of (NAD) in vertebrates. The production of NAD in a cell, is controlled by transcriptional regulation of involved enzymes (Belenky et al, 2007). The breakdown of NAD is also highly regulated and linked to gene regulation and signalling inside and outside a cell (Belenky et al, 2007).

NAD⁺ is the oxidized form of NAD and NADH is the reduced form. The phosphorylated form of NAD is NADP with its oxidized/reduced forms NADP⁺/NADPH. Reactions that change the oxidation state of NAD/NADH involve a transfer of a hydride ion, which is an unstable form of a hydrogen atom, which consists of a proton and two electrons (H⁻). This hydride ion is necessary for the catalytic action of several enzymes including ADHs. For the redox reaction with ADH, the following reaction with NAD can be drawn.

\[
\text{Alcohol} + \text{NAD}^+ \xleftrightarrow{\text{hydride transfer}} \text{Aldehyde} + \text{NADH}
\]

The relationships and concentrations of NAD⁺/NADH and the phosphorylated forms NADP⁺/NADPH are crucial for the metabolism, energy production and redox potential in cells. The human NAD pool is dominated by the oxidized form NAD⁺, thus often involved in oxidation of substrates. In order to maintain stable ratio, NADH is re-oxidized to NAD⁺, with e.g. NADH dehydrogenases. The NADP pool is dominated by the reduced form NADPH and thus often involved reductive pathways.

NAD is unable to passively diffuse from the cytosol to the inner mitochondria compartment. Without direct transport of NAD between the cytosol and the mitochondria, enzymatic reactions can indirectly change the ratio between NAD⁺ and NADH. The total concentration of NAD in hepatocytes is reported to be around 0.3 mM.
in the cytosol and 0.4 mM in the inner mitochondrial membrane matrix (Yamada et al 2006). This concentration is much higher than the $K_m$ values for NAD in human ADH1-ADH4.

1.9 Mutations, polymorphisms and disease

Mutations which affect the risk for disease is, for example, different allelic forms of ADH1 and ALDH2 (Crabb et al, 1993; Lee et al, 2006; Shibuya & Yoshida, 1988; Tanaka et al, 1996). The allelic variant ADH1B*2 protects against alcoholism because of a faster catalytic conversion of ethanol to acetaldehyde, which increases the blood acetaldehyde level, leading to unpleasant side effects from drinking (Lee et al, 2004). Homozygosity of ADH1B*2 variant has been shown to reduce the risk of alcoholism eight fold (Chen et al, 1999). In addition, homozygosity of ALDH2*2 almost completely protects against alcoholism through the lack of catalytic activity for by ALDH2*2, leading to severe unpleasant side effects, such as headache, increased heartbeat and a feeling of sickness from high acetaldehyde blood concentration. Other potential important polymorphisms that alter the catalytic activity are found in ALDH1 and ADH2 alleles. The general metabolic effect of all these mutations is that they increase the conversion to acetaldehyde or decrease the conversion of acetaldehyde. This leads to a higher intracellular and blood acetaldehyde concentration, which gives unpleasant side effects and reduces the incentive to drink alcohol. This metabolic effect is utilized by the drug Antabus, working as an inhibitor of ALDH.

Diabetes patients have difficulties to regulate the blood glucose concentration. This is simply explained, due to a decreased or delayed insulin secretion from pancreatic $\beta$-cells and decreased insulin sensitivity in cells. Moderate alcohol ingestion has been shown to be protective against type 2 diabetes (Carlsson et al, 2005). In a population study, the different isoforms of ADH1 and ALDH2 was recently shown to be linked to differences in risk to develop diabetes type 2, possibly due to differences in blood insulin levels (Dakeishi et al, 2008). SDH is important for the metabolism of sorbitol. Sorbitol has been suggested to cause damages to different tissues due to osmotic effects, thus studies of SDH and its function could possibly reveal new ways to modulate its
enzymatic function. This could potentially be used in treatment of tissue damages in diabetic patients.

Mutations in the ADH and ALDH genes have also been linked to cancer, where the link to acetaldehyde could explain an increased risk for several types of cancer (Brooks et al, 2009; Ding et al, 2008; Druesne-Pecollo et al, 2009; Homann et al, 2009). The flushing reaction due to ethanol drinking within the oriental population has been suggested as a clinical marker for an increased risk of esophageal cancer (Brooks et al, 2009; Yokoyama et al, 1996).
2 Theories and methods for enzymatic studies

2.1 Proteins as enzymes

A protein is an enzyme if it can catalyze a chemical reaction. Enzymes are macromolecules with a structure that can be defined at different levels, if sorted with an increasing complexity, defined as the primary, secondary, tertiary and quaternary structure. With nuclear magnetic resonance and x-ray crystallography the most detailed experimental structures are determined, where a protein structure can be given with a resolution down to roughly one Ångström (10^{-10} m). The relative position (in an x,y,z coordinate system) of all atoms are deposited as a Protein Data Bank (PDB) file into large databases, such as RCSB or Brookhaven. The structure of proteins can be freely downloaded from these databases and visualized by molecular graphics programs, at different structural levels.

The catalyzing capacity of enzymes, is accomplished upon the binding of a substrate (single substrate reactions) followed by a chemical reaction, where the transition state free energy barrier is lowered (Garcia-Viloca et al, 2004). One complete circle of a chemical reaction is finalized upon the release of the product to the solution. Different types of enzymatically catalysed reactions are classified according to the Enzyme Commission number (EC number), where each enzyme is placed in a main group and subcategories depending on which type of chemical reaction that is catalysed. There are six main classes; oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5) and ligases (EC 6). MDR belongs to the oxidoreductases (EC 1), and ADH has been assigned EC number of 1.1.1.1.

There are still several obstacles left to fully understand enzymes at atomic and biological levels. Firstly, enzymes are large molecules of up to millions of atoms with complex three dimensional structures. Secondly, enzymes are very dynamic, because
non-covalent interactions under physiological conditions are generally unstable. Thirdly, the mechanisms behind the lowering of a transition state free energy barrier is complex. What further complicate our physiological understanding of enzymes is that in vivo, the enzymatic function is altered by “crowding effects”, compartmentalisation, membrane interactions, activators, inhibitors, multiple kinds of substrates and products, pH changes, new synthesis and degradation of enzymes, molecular transport in and out of the cell, post transcriptional modifications and post translational modifications. Today numerous methods exist which try to determine the in vitro activities and physiological function of enzymes. Below, methods for enzymatic studies which are related to this thesis are presented.

2.2 In vitro methods for protein experiments

2.2.1 Expression, purification and identification of proteins
The ADH and SDH proteins were expressed in penicillin resistant E. coli cells. The cells were lysed and the cytosolic fractions were separated from membrane lipids and other organelles, by centrifugation. The recombinantly expressed proteins were purified by an ion exchange chromatography DEAE-cellulose column, as a first step. Often affinity chromatography was used as a second purification step, either AMP-sepharose or blue-sepharose. As a final third step, gelfiltration chromatography was used. To determine the purity of the samples, gel electrophoresis was used. Finally, massspectrometry was used to verify the proteins.

2.2.2 Gelfiltration
Gelfiltration (size exclusion chromatography) separates molecules based on their hydrodynamic volume (Meredith, 1984), and for globular proteins the hydrodynamic volume is approximately proportional to the molecular mass (M). The resolution of the column is based on both column length and separation media as well as packing conditions. Therefore it is necessary to calibrate the column with proteins of known molecular masses. The following graph shows the calibration of the gelfiltration column and the elution volume for the different quaternary states of SDH used in Paper II.
The elution volumes ($V_e$) of the proteins used in the calibration are divided by the void volume ($V_o$), which here is the elution volume for blue dextran ($M = 2,000,000$ Da), and plotted against the logarithm of the protein standard molecular masses. This technique was used in paper II to determine the quaternary states and size of proteins (McEvily & Harrison, 1986; Mullen & Jennings, 1998).

2.2.3 Enzyme kinetics

Enzyme kinetics models are designed to describe enzymatic catalytic turnover rates for chemical reactions at different conditions. Due to the complexity of catalytic reactions with enzymes, most models are only valid under strict steady states and in vitro conditions. The initial reaction velocity as a function of the substrate concentration has the following expression:

$$V_o = \frac{V_{\text{max}} \times S}{K_m + S}$$

The substrate concentration ($S$) is known and the initial reaction velocity ($V_o$) is
measured. The two “unknown” kinetic constants are $V_{\text{max}}$ and $K_m$, where $V_{\text{max}}$ is the maximum reaction velocity and $K_m$ is the substrate concentration at half the maximum reaction velocity. The initial reaction rate has to be determined when the concentration of product can be approximated to zero, which in reality means that the concentration of products is much less than the concentration of substrates. Steady-state for the enzyme-substrate (ES) complex must be valid, i.e. the rate of association equals the rate dissociation for the ES complex. Furthermore, the concentration of enzyme has to be much less than the substrate concentration. The general applicability for many catalysed reactions by enzymes makes the M-M determinations useful for comparisons between enzymes, substrate specificities and metabolic pathways.

Inhibition of proteins is commonly the mechanism behind the physiological action of many drugs. The reaction rate for enzymes can be decreased by denaturation, covalent modifications or from inhibition by other molecules. Inhibition of enzymes is studied by measuring the reaction rate of an isolated enzyme at different concentrations of a molecule with an inhibitory effect on the enzyme. Four main types of mechanisms used to describe inhibition mechanisms are: competitive, non-competitive, un-competitive and mixed inhibition.

With systems biology approaches, larger metabolic networks which are open and not in equilibrium can be simulated. These models often use the experimental determined enzyme kinetic constants (such as $k_{\text{cat}}$ and $K_m$) for predictions of metabolic reaction schemes under different conditions.

### 2.3 In silico methods for protein interaction

Bioinformatics, computational chemistry and molecular modelling were used for calculations and simulations of the biomolecular systems in this thesis. In addition, programs for visualization of proteins or for analyzing of large quantities of biological data were used. The different in silico methods have their pros and cons and no single method is outstanding.

Today, molecular modelling is a field that explores the dynamics and interactions
between molecules in biological systems, down to an atomic level. This field ranges from QM calculations of small systems with a few atoms to the dynamics of large biomolecular systems consisting of millions of atoms, such as proteins in solution. Macroscopic properties which are measured in lab experiments are, in most cases, determined from an ensemble of a large number (>10^6) of molecules, in various different states. This has an important consequence, namely that a single structure can not be used to determine macroscopic properties from the total system. To determine macroscopic properties with in silico methods, such as free energies in molecular dynamics (MD) simulations, extensive sampling and long simulations are needed. However, the general complexity of even a small biological system makes macroscopic properties difficult to determine with computational methods. Still, experimentally determined structures in a single conformation are useful, since it, in most cases, represents a conformation which is stable and often represents a global energy minimum. This experimentally resolved structure is often used as a starting point in simulations. From x-ray structure determinations several important structural and functional properties and relations can be identified: substrate and cofactor binding sites, quaternary structure, subunit interfaces, amino acid residues relative position, secondary structure, protein domains and hydrogen-bonds between amino acids. Protein structure modelling is used to deepening our understanding about the connection between protein sequences and protein functions.

2.3.1 Molecular dynamics simulations with GROMACS
MD is a simulation technique which uses Newton’s classical mechanics theory. QM are not a direct part of the MD simulation algorithm but can be used for determinations of force-field parameters or in mixed MD-QM simulations. From MD simulations studies both static equilibrium properties, such as Gibbs free energies, or dynamic properties such as binding kinetics can be determined. GROMACS is one out of several programs for MD simulations and was originally designed for simulations of biomolecular systems (Berendsen et al, 1995).

The programming of the MD simulation is started by figuring out an appropriate level of detail for the system, e.g. system size, complexity and time scale of the problem.
MD simulations of solvated enzymes are in general between ns (10⁻⁹) to µs (10⁻⁶), which limits, the biological events that can be modelled. In order to determine wt properties of enzymes, a reasonable starting position and conformation for the protein-ligand complex is important for as a correct results as possible. Important approximations and limitations which are associated with MD simulations are: system size is typically much smaller than in vitro and in vivo experiments, relativistic effects are often not included, all electrons stay in the same electronic state during simulations, the force field is pair addittative and system boundaries is artificial which can produce artefacts. Still, with MD simulations many physical properties of proteins in solutions have been accurately predicted.

2.3.2 Monte Carlo simulations with ICM

Molecular simulations that are based on Monte Carlo (MC) algorithms, often has an iterative search through the potential energy landscape. These MC algorithms are commonly written so that the probability for a structure to be kept or rejected for the next iteration step, is coupled to the Boltzmann probability function. There is always a chance to keep a state with a higher potential energy which avoids being trapped in a local potential energy minimum.

MC simulations of molecules have similar limitations as described for the MD method. In line with MD simulations, MC methods are limited by: the complexity of large molecular structures which makes it impossible to search through the whole energy landscape, and most force-field parameters are based on empirical results and averages. In addition, MC simulations treat water molecules implicitly (not always) and lack time dependence in simulations.

ICM (Internal Coordinate Mechanics) was originally built to predict potential energies for biomolecular systems (Abagyan & Totrov, 1994; Abagyan et al, 1994). The ICM is programmed after a method which uses a biased probability, based on experimentally determined differences in torsions angels for different amino acid residues. This method is called a biased probability MC procedure (BPMC)(Abagyan & Totrov, 1994).
Most predictions, homology modelling, of three dimensional structures for proteins are based on template structures from proteins with a similar sequence. There is a strong correlation between the accuracy of results and how similar the sequences of the template and the model are.

Molecular docking in ICM can be performed with the BPMC method. It is possible to reduce the degrees of freedom and thus the computational cost, by a reduction of movable torsion angels. If e.g. the active site residues are allowed to rotate, only amino acid residues there are allowed to adjust to the binding of ligands. Thus, an induced-fit between different ligands and the protein can be simulated.

Virtual screening (VS) methods are used to identify new ligands within a large set of compounds, by docking them to an enzyme or protein. With ICM one can determine a static grid energy potential for a selected part of the enzyme, often corresponding to that enzymes active site. This use of a grid energy potential, allows for large screenings of compounds. To avoid unnatural short term repulsions due to the static nature of the grid potential, a soft van der Waals potential can be used.

2.3.3 Quantum chemistry simulations with GAUSSIAN
With QM methods a system of atoms can be simulated by numerical algorithms which can solve the time dependent Schrödinger equation. However, even with numerical solutions, only small systems of atoms can be simulated and different approximations is necessary. GAUSSIAN is one program for QM simulations that can be used for calculations of e.g. chemical reactions free energies, chemical reaction rates, bond enthalpies, fractional charge distributions, chemical shifts and molecules electronic orbitals. Density functional theory (DFT), used in Paper III, are a method which has been shown to be very accurate for a comparatively small computational cost (Becke, 1993; Godbout et al, 1992; Johnson et al, 1993).
3 Results and discussions, Paper I to V

The overall aim of this thesis was to determine structure-function relationships to investigate and identify new ligands for ADH. In order to accomplish this, different in vitro and in silico methods were used.

3.1 Paper I

The aim of this paper was to determine and explain differences in kinetic rate constants for the conversion of retinoids (Vitamin A) by rodent and human ADH2. In addition, the physiological role by ADH2 in hepatic liver metabolism of vitamin A, was evaluated.

Compared to human ADH2, the rodents ADH2 contain a proline at position 47 instead of a histidine. By a Pro47His exchange in mouse ADH2, it was possible to evaluate the effect of this mutation for the retinoid catalytic rates in rodents as compared to human.

Since the solubility for retinoids is poor in water, due to their hydrophobic properties, a mixture with albumin was used, in a 1:1 ratio to the retinoid molecules, in a mixture with phosphate and water. The enzyme kinetics was performed at 37°C, pH 7.5, to mimic physiological conditions.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Human</th>
<th>Rat</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADH2</td>
<td>ADH2</td>
<td>ADH2</td>
</tr>
<tr>
<td>all-trans-retinol</td>
<td>(5.2 μM)</td>
<td>48,600</td>
<td>6,100</td>
</tr>
<tr>
<td>all-trans-retinaldehyde</td>
<td>(6.2 μM)</td>
<td>27,500</td>
<td>500</td>
</tr>
<tr>
<td>9-cis-retinol</td>
<td>(7.6 μM)</td>
<td>220,000</td>
<td>11,400</td>
</tr>
<tr>
<td>9-cis-retinaldehyde</td>
<td>(4.8 μM)</td>
<td>38,000</td>
<td>6,400</td>
</tr>
<tr>
<td>all-trans-retinol</td>
<td>(30 μM)</td>
<td>150,500</td>
<td>105</td>
</tr>
</tbody>
</table>
The reaction rates in the table above are given as the amount of converted substrate in pmol per minute per mg of protein at a single substrate concentration. The enzymes correspond to ADH2 in human, rat, mouse and the mouse Pro47His compared to ADH1, ADH3 and ADH4 from mouse.

![Graph comparing enzyme activity](image)

The graph above compares the relative activity for RoDH-4, RDH11, ADH2 and ADH1B1 in the conversion of all-trans retinol at different concentrations. This graph shows that the relative importance of ADH2 increases with decreasing substrate concentration.

In addition to the kinetic experiments, molecular docking of all-trans-retinol and 9-cis-retinol with the x-ray determined structures of mouse ADH2 and homology models of rat and human ADH2, were performed. To evaluate catalytic features and structural differences. ICM was used for construction of homology models of human and rat ADH2, and as a template structure, the x-ray determined structure of mouse ADH2, was used. All-trans retinol and 9-cis-retinol were docked into the ADH2 structures with restraints to the active site zinc, coenzyme NAD$^+$ and to the side chain of Thr48. These restraints were programmed to mimic the catalytic binding mode for the retinoids. In the molecular docking, the docked retinoid molecule and the amino acids residues surrounding the active site, were able to rotate freely. The results showed a very similar binding mode between species and no strong interaction between neither His47 in human ADH2 and Pro47 in rodent ADH2s were observed. These results indicated that the
difference in activities between species is not directly related to neither His47 in human nor Pro 47 in mouse.

Finally, the in vitro results clearly indicate a role for ADH2 in hepatic retinol oxidation in humans at physiological concentrations of retinoids. Furthermore, the mutation of Pro47His increased the reaction rate and can thus explain a part of the difference in catalytic activity between mouse and human ADH2.

3.2 Paper II

The aim of this paper was to explore the molecular interaction between subunits that form the homo tetrameric structure of SDH. Alignments, structural comparisons, energy calculations, gelfiltration experiments and enzyme kinetics were used.

The x-ray determined structure of human SDH, pdb entry 1PL7, was used to determine the interactions between the subunits (A,B,C,D) as judged from the calculated scoring energies (ICM).

<table>
<thead>
<tr>
<th>Subunit</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.06</td>
<td>0.18</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.06</td>
<td>0.50</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.18</td>
<td>0.50</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.50</td>
<td>0.18</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

The calculations showed that the scoring energies between the subunits A with D and B with C are strong. This interaction corresponds to the interaction between two subunits in the formation of a dimeric structure in mammalian ADHs. The interactions between subunits; A with B, A with C, B with D and C with D, is then probably crucial for the formation of a tetrameric structure. Of these interactions, the highest scoring energies were calculated to be between subunits A with C and B with D. From these calculations of the interaction, we identified an important region, of which the Tyr110 is a part of.
A hydrogen-bonding network was identified, which seemed to be crucial for the tetrameric stability in SDH: (Tyr110(subunit A)-Glu94(subunit A)-Tyr140(subunit A)-Lys106(subunit C)-Asn301(subunit A)-Tyr110(subunit A)).
<table>
<thead>
<tr>
<th>Protein Species</th>
<th>Species</th>
<th>Protein Sequence</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDH</td>
<td>Human</td>
<td>PRENDEFCKMGRYMNLSPSIFFCAT</td>
<td>T/D</td>
</tr>
<tr>
<td>SDH</td>
<td>Rat</td>
<td>PREIHEFCGIRYNLYTSPFFCAT</td>
<td>T</td>
</tr>
<tr>
<td>SDH</td>
<td>Whitefly</td>
<td>PCRCQFCKEKGYNLCPDTFCAT</td>
<td>T</td>
</tr>
<tr>
<td>SDH</td>
<td>Yeast</td>
<td>PDRFSPEMKERYNLDPKNFAAT</td>
<td>T</td>
</tr>
<tr>
<td>SDH</td>
<td>S. haemolyticus</td>
<td>PCRECYESGQYNLCPHMFMAT</td>
<td>T</td>
</tr>
<tr>
<td>ADH</td>
<td>T. brokii</td>
<td>PDWRTSEVORYHSHGGLAGWK</td>
<td>(T) (Z)</td>
</tr>
<tr>
<td>ADH</td>
<td>C. beijerinckii</td>
<td>PDWSLEVGQPFHSGNMLAGWK</td>
<td>(T) (Z)</td>
</tr>
<tr>
<td>ADH1</td>
<td>Yeast</td>
<td>SCMACYECLEGNESNCPHA-DLSG</td>
<td>T</td>
</tr>
<tr>
<td>ADH1</td>
<td>Rat</td>
<td>QCXKWRCI CKNPESNLCCQTKNLPPQKAGALDTGSRFSCRGKPI HHFIST</td>
<td>D</td>
</tr>
<tr>
<td>ADH1B1</td>
<td>Human</td>
<td>QCXKWRCI CKNPESNLCCQTKNLPPQKAGALDTGSRFSCRGKPI HHFGLGT</td>
<td>D</td>
</tr>
</tbody>
</table>

This above alignment shows differences between protein sequences from SDH and ADH in different species. The alignment clearly shows that the structural zinc is not a determinant for the quaternary difference in either ADH or SDH, however the insertion of about 20 amino acids residues after the loop region 100-110 in SDH, seems to play a role, since all dimeric ADHs contain this insertion.

This study showed how a network of hydrogen-bonds in mammalian SDH upholds a protein interface, which is favourable for a tetrameric quaternary state, and thus crucial for the catalytic power. Moreover, the free energy between wt and Tyr110Phe SDH showed a large difference in protein stability.

### 3.3 Paper III

In this paper the binding between zinc and two peptides were determined with *in silico* methods. One of the peptides (Peptide(4Cys)) corresponds to the binding motif of the structural zinc site in horse liver ADH, and the other peptide (Peptide(3Cys)) corresponds to the same binding motif with one of the cysteine replaced with an alanine. These results were compared to *in vitro* determined binding affinities.

Zinc binding to proteins or peptides is not simple to simulate and two different approaches can be used, a bonded or a non-bonded interaction. A bonded interaction ensures a stable binding but can not be used in dynamic descriptions of binding or releasing of the zinc ion.
In ADH both the catalytic and structural zinc ions are vital for the catalytic function and the stability of the enzyme. The structural zinc ion in ADHs is bound to four sulphur atoms from the side chains of cysteine residues (Cys97, Cys100, Cys103 and Cys111 in horse liver ADH).

This image shows the zinc binding peptide which corresponded to the structural zinc site in horse liver ADH. The image is generated from a MD simulation.

The classical binding free energies were calculated with three different techniques; thermodynamic integration, umbrella sampling and the linear interaction energy method. The results showed that the absolute classical binding free energies were consistently overestimated with all three methods, but that the relative binding free energies between the two peptides were determined within the statistical error. In order to understand the reasons behind the suboptimal prediction of absolute binding free energies, an additional QM treatment was conducted.

The QM simulations were based on DFT and performed on \( \text{Zn-(S-CH}_3\text{)_4}^2^- \) and \( \text{Zn-(H}_2\text{O)}_6^{2+} \) clusters. From the QM calculations it was observed that the binding of water molecules was underestimated in the classical treatment by 178 kJ/mol. This error was caused by the polarisation of water molecules by the zinc ion, an effect that strengthens the electrostatic interaction. The binding between the zinc ion and the sulphurs, was also underestimated, which was calculated to be 93 kJ/mol. Also, the charge of the zinc ion was reduced from 2e to 1,4e. This implies that the negatively charged sulphur atoms donate or share a part of their electrons with the zinc ion. This effect would decrease the classical electrostatic interaction, an effect which here only can
be understood if part of the binding energy was due to a covalent interaction. With the addition of the QM corrections both the hydration free energy for zinc and the absolute binding free energy for the Peptide(4Cys) were determined within the statistical error of the method.

In this paper, the solvation free energy for zinc ions in solutions and the absolute binding free energy between a zinc ion and a peptide were determined in line with experiments. In order to succeed both MD and QM simulations were needed. The study revealed details on the complex nature of zinc interactions with proteins. These interactions are crucial to determine in order to fully understand the role of zinc in biology.

3.4 Paper IV

In this paper the inhibition of ADH3-mediated GSNO reduction by substrate analogues of medium-chain fatty acids and glutathione derivatives were investigated. ADH3 has been suggested to be an efficient GSNO reductase which thereby will regulate protein nitrosylation, a protein modification which is known to be important for many cellular processes. Depletion of GSNO has been suggested to be related to several diseases, e.g. asthma, making ADH3 a potential drug target, where inhibition of the enzyme could be helpful.

The catalytic activity of ω-hydroxy fatty acid oxidation was lower for 8-hydroxyoctanoic acid compared to 10-hydroxydecanoic acid and 12-hydroxydodecanoic acid, primarily due to a higher Km value. The kcat values for 10-hydroxydecanoic acid and 12-hydroxydodecanoic acid were decreased ten-fold if the pH was changed from 10 to 7.5. Fatty acids of chain lengths varying between 8 to 13 carbon atoms were tested as inhibitors of GSNO reduction. The K_i values showed an optimal inhibitory effect for eleven and twelve carbons and the obtained K_i values for inhibition of GSNO by the corresponding inhibitors octanoic acid, decanoic acid and dodecanoic acid, were close to the K_m values for hydroxyoctanoic acid, 10-hydroxydecanoic acid and 12-hydroxydodecanoic acid.

The low K_m values for GSNO and HMGSH (11.1 µM and 0.12 µM, respectively)
suggested that glutathione derivatives could work as inhibitors. The $K_i$ for GSH was shown to be in the millimolar range. For improved inhibition a zinc binding moiety was searched for. S-acetoamidoglutathione was tested, which has an acetoamido group attached to the glutathione scaffold and the carbonyl oxygen could theoretically be in a zinc coordination distance. This improved the inhibition threefold in comparison to GSH, but the $K_i$ was still in the millimolar range. Another glutathione derivative, S-methylglutathione, had a $K_i$ close to GSH. GSSG showed a very weak inhibition and glutathionesulfonic acid lacked inhibition capacity (10mM).

Overall, neither the fatty acid nor the glutathione derivatives yielded very strong inhibitors to ADH3. In general the inhibition was best fitted to a non-competitive model, but quantitatively the $K_i$ values did not change much if a competitive model for inhibition was used.

To further understand the differences in inhibitor mechanisms and specificity a docking study of GSNO compared to MGSH was performed. These docking simulations showed that MGSH binds poorly to the catalytic zinc ion due to sterical hindrance from the more bulky methyl group. These findings imply that for glutathione-derived ligands a nucleophilic group in coordination distance to the active site zinc, is important for strong binding. The docking simulations support that this interaction is fruitful for stabilization and for correct orientation of the $\gamma$-glutamyl and glycyl branches towards hydrogen bonding partners from Asp55, Glu57 and Arg114 in an otherwise mostly hydrophobic active site.

This study identified and determined the effect of several inhibitors to ADH3, in the reduction of GSNO. The in vitro and in silico results, identified several structure-function properties of interest for finding specific inhibitors for ADH3.
3.5 Paper V

In this paper a VS approach was used. The ICM software was used to perform molecular docking of a large number of (40,962) compounds into the active site pocket of human ADH3. The compounds ($C_i$) were ranked based on the distance ($r$) between the closest oxygen atom in each compound to the catalytic zinc ion in ADH3. The main objective with this VS approach was to find new ligands to ADH3. To decrease the number of compounds to be tested in vitro, a scoring method was used to only test plausible candidates in vitro. In this study a set of known inhibitors ($I_i$) and substrates ($S_i$) to ADH3 were used to evaluate the choice of scoring method and to determine what compounds to test in vitro. For this purpose, a general function for evaluation of the VS approach were constructed, called Enrichment:

$$
\text{Enrichment}(r) = \frac{\sum S_i + \sum I_i}{\sum S_i + \sum I_i} \times \frac{\sum C_i}{\sum C_i}
$$

For a VS scoring method which can not perform better than a random selection of compounds, would on average produce an enrichment value of one. For scoring methods better than random selection the Enrichment value is above one.

Above is a plot of the Enrichment at different cut-off distances (1 Å to 5.5 Å) to the catalytic zinc. The Enrichment($r$) at different distances from the catalytic zinc was
highest at 2.5 Å.

The cut-off distance for compounds after VS step 1 was based on a high Enrichment(r) and a suitable amount of compounds (~2700). The second docking was more computationally demanding per compound and included flexibility of the amino acid side chains lining the active site pocket. This flexibility opened for specific adjustments to compounds within the active site. The compounds to test in vitro were selected based on either a short zinc binding distance (<2.6 Å) or a total short distances to NAD⁺ and Arg114 (<6.0 Å). These top candidates from VS step 2 were matched against their commercial availability. Six top ranked compounds that were not known to interact with ADH3, were tested in vitro. Two showed substrate activity (dodecyltetraglycol and 9-decen-1-ol), two showed inhibition capacity (7-deoxycholic acid and doxorubicin) and two compounds did not show neither substrate nor inhibitor properties.

For increased knowledge about site specific interactions, a number of known ligands were docked with an extended docking algorithm and the binding scoring energies were calculated with the REBEL method. These scoring energies correlated well with the logarithm of the substrates kcat over Km values.

In this study new substrates and inhibitors to ADH3 were identified. The VS approach and in vitro experiments revealed a novel group type of inhibitors, bile acids. As being naturally produced this type of inhibitors are potentially useful, with less risk of toxic reactions.
4 Conclusions

Within this thesis, the role of ADH in both the retinoid and the NO metabolism were assessed. New substrates were identified and active site interactions were studied in relation to alterations in structure and the composition of amino acids residues within the active site.

The role of ADH2 in the retinoid metabolism was investigated. For all-trans retinol and 9-cis retinol with corresponding aldehydes, the determined $K_m$ value were lower than in other retinoic active enzymes in humans. Furthermore, our docking of all-trans retinol and 9-cis retinol into the active site of mouse, rat and human ADH2 revealed site specific interactions which were in line with the results from the in vitro experiments. Clearly, ADH2 play an important role in hepatic retinoid metabolism.

The stability of the quaternary structure of SDH was determined by in vitro experiments and in silico energy calculations. This revealed a hydrogen-bonding network crucial for the tetrameric stability in SDH. This network is located at a region enclosing the structural zinc site in mammalian ADHs. In turn, the structural zinc site is essential for correct folding and stability of the human ADH. The determined free energies for the bonds between peptides and a zinc ion revealed that the interaction is predominantly ruled by electrostatic interaction, however, with a covalent contribution. The study revealed details on the complex nature of zinc interactions.

With both in vitro and in silico methods, new substrates and inhibitors to ADH3 were identified. A new ligand, deoxycholic acid, was identified through a VS approach and confirmed by in vitro experiments. Experiments with other bile acid derivatives confirmed that this compound group constitutes a novel type of inhibitor of ADH3. The inhibition of the catalysis of GSNO by ADH3 was studied with glutathione derivatives and fatty acids.

ADH, with its recognised important role as an eliminator of toxic compounds and as a historic model system, will continue to draw attention from researchers and new discoveries are yet to be presented.
5 Reflections

Medicine is an interdisciplinary research field that includes the disciplines of physics, chemistry, biology, psychology and social sciences. Inspiring goals for medical researchers are to develop both tools and practices for treating diseases. In this thesis, my work has its basis in both a practical and a theoretical approach. I used experiments and simulations to explain structure-function relationships in a specific protein family. The goal was to determine substrate and inhibitor specificities for different ADHs and to place the results on the human metabolic map.

The ADH enzymes are only one family, out of many, that form intracellular metabolic networks. In multicellular organisms, proteins also mediate the cell to cell interaction between billions of cells, interactions which can be organized into separable units. Each organ has a specialized function for the survival or the reproduction of the organism. Today, the rapid advances within cell and molecular biology open up for new interpretations and an improved understanding of life. Remaining, however, is the obstacle of how to integrate the ongoing and sometimes exponential increase of data and the knowledge about protein structure, protein function, protein networks and cellular networks. Since each cell and indeed every protein molecule, is in a unique state from at least a physicist’s point of view, no end of discoveries can be anticipated and no "Grand Unified Theory" will be recognised in the field of medicine. Nonetheless, the borders between the different fields of medical research will gradually become vaguer as new methods and studies are presented. This ongoing process will slowly lead to a more integrated understanding of life for the benefit of mankind.

"Practice without theory is blind, theory without practice, empty"
6 Acknowledgments

My time as a PhD student has been filled with challenges and rewards. Much time was spent analysing and solving problems, sometimes of a trivial nature and occasionally rather complex ones. With some hard work, intuition, fine supervision and a little bit of luck, a few of my research projects ended with a paper in print. To all of my colleagues, friends and family who have helped and inspired me, I would hereby like to express my gratitude.

Plenty of hugs goes to:

Formal and informal senior scientific supervisors, 
Jan-Olov Höög, you both guided and trusted me, always being professional, and added a realistic handle to my problems. Y ou are a true diplomat who is dedicated to both science and teaching. To me, you have been an excellent supervisor. Bengt Persson, who always showed up with a smile and a solution in hand. This has been very valuable to me. Hans Jörnvall, you are always a source of inspiration and you helped me with your profound scientific knowledge, intuition and critical questions. Olle Edholm, for your passionate answers and your help related to physics conundrums and scientific collaborations.

Tomas Bergman, you kindly introduced me to the world of zinc and zinc interactions. Jan Sjövall, who are always kind and supportive. Birgitta Agerberth, who, very generously, lent out equipment and chemicals to me.

Colleagues in the lab and collaborators
Margaretha Brandt, my main lab supervisor who gave me support, time and protocols I could follow. Andreas Cederlund, also my good friend, who have helped me and the students numerous times in the lab. Y ou have very much contributed to a nicer atmosphere and helpful attitude within the lab. Åke Norberg, my friend and also a valuable teacher of the HPLC machine. Always keen to help out and to have discussions about both science and life. Claudia Staab, whom I shared both experiences and experiments with. Torbjörn Nordling, we share both research and Tambourrelli interests. Jonas Carlsson, with whom I had a nice collaboration in the last paper.

Past and present members of the lab:
Andreas Almén, Gunvor Alvelius, Gudmundur Bergsson, Peter Bergman, Karl Bodin, Ella Cederlund, Michael Fitzen, Mikael Henriksson, Daniel Hirschberg, Moin Israt, Theres Jägerbrink, Monica Lindh, Ylva Kai-Larsen, Emma Lindahl, Ermias Melles, Ingegerd Nylander, Carina Palmberg, Essam Refai, Protim Sarker, Jawed Shafqat, Rannar Sillard, Johnny Söderlund, Sam Tryggvason, Marie Stålberg and Jan Wiberg.

Universitat Autònoma:
During three weeks in 2005 I got an opportunity to visit Xavier Parés research group in Barcelona. Oriol Gallego kindly showed me around and helped me with experiments in the lab. I had a pleasant stay with lots of work and some time for evenings out in Barcelona.

MBB staff members who make the wheels turn:
Angelica Arribada, Svante Backlund, Susie Björkholm, Tomas Cronholm, Jan Hallensjö, Sören Lundmark, Anders Lundsjö, Agneta Nordberg, Annika Pettersson,
Susanne Rothstein and Åke Rökaeus.
Additional Co-authors:
Patrik Strömberg, Silvia Martras, Jaume Farrè, Christina Kaiser, Simone de Haid, Roland Graafström

Students
Erik Brandt, even from the very start of your master thesis work, you were quite independent and worked hard, which gave results. Haneen Hamada, you showed very much of a fighting spirit and stubbornness, which in the end gave active and stable enzymes with which you could generate experimental results. Linus Östberg, you voluntarily spent each of your summer holidays in the lab, performing kinetic experiments, fixing computers and asking a lot of questions. I hope that you continue on the path of research.

Friends
Seth Malmersjö, Finn Lennartsson and Markus Kruusmägi, are my good friends with whom I created the “KI Community” site. Siwert Nilsson, it was a long journey and no words are really necessary. Anders Edquist, Dan Sjöstrand and Patrik Brusheim, we had lots of fun during our KTH studies. Camilla Appelgren, Frida Haglund, Anders Lindström, Karin Wahlberg, Jenny Davidsson for all the help with, and love you show for Tim.

Södertälje friends
Mats “Sune” Ericsson, not many people nowadays knows my nick name “Vante”. Per Karlsson, my best friend for several years, and Johan Ritzler was also an unforgettable part of the gang. Mats Pettersson for being a great grappler and friend. Kristina Hulth and Pernilla Stark, two stars who both have a special place in my heart.

MBB friends
Sofi Eriksson, Masatoshi Nukui, Juha Ojala, Jaakko Patrakka, Bernhard Lohkamp, Edvard Wigren, Magnus Claesson, Robert Schnell, Yilmaz Mahshid, Ásmundur Oddsson

Family
Claes and Birgitta Hellgren, for your support, love and care. Anna-Karin Hansen, Maria Hellgren De Vylder and Matti Hellgren are my loved sisters and brother and Niklas Hansen, Andreas Maldonado and Therese Hellgren are their loved partners in life. Line, Ville, Einar, Angnes, Stina and Frida, I wish you all the luck in life. Eddy Jonsson and Elisabeth Hellgren you are such great bonus parents. Klas and Birgit de Vylder, both of you are an inspiration and source of joy to me. Stina De Vylder, thank you for your love and the joyful times we had, which I will always remember. Elisabet van Nostrand, Jim van Nostrand, Jennifer Banta and Tracy Rhodes, you are my dear American family. Stefan, Martin and Sara De Vylder, I hope to see you in Kårköle next summer! Rolv Brenna, Johan, Ingeborg and Kristina Arntyr, you are my new family.... Tim, I am sure you know this, but for the record, you are perfect and I love you.

Paulina Arntyr, with your heart, support and care you have showed me that I am worth to be loved. You are sincere, honest, and intelligent and much more, I love you!
7 References


65

Lemieux RU, Spohr U (1994) How Fischer, Emil was led to the lock and key concept for enzyme specificity. Advances in Carbohydrate Chemistry and Biochemistry 50: 1-20


69


Vallee BL, Bazzone TJ (1983) Isozymes of human-liver alcohol-dehydrogenase. Isozymes-Current Topics in Biological and Medical Research 8: 219-244


