Molecular evolution and functional divergence of alcohol dehydrogenases in animals, fungi and plants

Claudia E. Thompson1,2, Loretta B. Freitas3 and Francisco M. Salzano3

1Departamento de Farmacociências, Unidade de Genômica e Bioinformática Clínica, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil.
2Unidade de Biologia Teórica e Computacional, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.
3Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Abstract

Alcohol dehydrogenases belong to the large superfamily of medium-chain dehydrogenases/reductases, which occur throughout the biological world and are involved with many important metabolic routes. We considered the phylogeny of 190 ADH sequences of animals, fungi, and plants. Non-class III Caenorhabditis elegans ADHs were seen closely related to tetrameric fungal ADHs. ADH3 forms a sister group to amphibian, reptilian, avian and mammalian non-class III ADHs. In fishes, two main forms are identified: ADH1 and ADH3, whereas in amphibians there is a new ADH form (ADH8). ADH2 is found in Mammalia and Aves, and they formed a monophyletic group. Additionally, mammalian ADH4 seems to result from an ADH1 duplication, while in Fungi, ADH formed clusters based on types and genera. The plant ADH isoforms constitute a basal clade in relation to ADHs from animals. We identified amino acid residues responsible for functional divergence between ADH types in fungi, mammals, and fishes. In mammals, these differences occur mainly between ADH1/ADH4 and ADH3/ADH5, whereas functional divergence occurred in fungi between ADH1/ADH5, ADH5/ADH4, and ADH5/ADH3. In fishes, the forms also seem to be functionally divergent. The ADH family expansion exemplifies a neofunctionalization process where reiterative duplication events are related to new activities.

Keywords: Glycolytic proteins, molecular evolution, alcohol dehydrogenase, functional diversification, positive selection.

Received: February 23, 2017; Accepted: June 19, 2017.

Introduction

The alcohol dehydrogenase (ADH, EC 1.1.1.1) enzyme belongs to the large superfamily of medium-chain dehydrogenases/reductases, which include different enzyme activities, such as alcohol, sorbitol, xylitol, threonine dehydrogenase and quinone reductase (Persson et al., 1993). Its activity appears to be universal in all life forms, derived from enzymes of separate family assignments and, frequently, involves multiple occurrences in a complex fashion (Norin et al., 1997).

ADH class III (ADH3), with little or almost no ethanol activity and similar to the glutathione-dependent formaldehyde dehydrogenase, seems to be an ancestral form. Moreover, it has been characterized in vertebrates (Jörnvall and Höög, 1995; Hjelmqvist et al., 1995b), invertebrates (Kaiser et al., 1993; Danielsson et al., 1994), plants (Martínez et al., 1996), fungi (Sasnaukas et al., 1992; Fernández et al., 1995), and prokaryotes (Guthiel et al., 1992; Ras et al., 1995). ADH3 acts as a glutathione-dependent dehydrogenase in the oxidative elimination of formaldehyde, but does not function in ethanol or retinol oxidation, a function that is realized by other ADH classes (Duester et al., 1999). Additionally, it is considered to be the most ancient form of vertebrate ADH, reflecting the fact that it is the only form also detected in invertebrates (Kaiser et al., 1993).

Vertebrate ADH is a cytosolic, dimeric, zinc-containing, NAD-dependent enzyme with a subunit molecular mass of 40 kDa. Based on sequence alignment, phylogenetic analysis, catalytic properties and gene expression patterns at least eight distinct classes have been identified in vertebrates. ADH classes share around 60% amino acid sequence identity, and multiple ADH isoenzymes within a single class share above 90% identity (Jörnvall, 2008). They metabolize a wide variety of substrates, including eth-
anol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products (Duester et al., 1999).

In humans, ADH classes I (with three isoforms: A, B, and C, earlier called α, β, and γ, respectively), II, III, IV and V have been identified, and in mouse, classes I, II, III and IV have been described (Boleda et al., 1993; Zheng et al., 1993; Höög and Brandt 1995; Höög et al., 2001). ADH class VI has been observed in rat and deer mouse (Zgombic-Knight et al., 1995), and ADH class VII has been found in chicken (Kedishvili et al., 1997), where it may act as a steroid/retinoid dehydrogenase. An amphibian ADH class VIII (class IV-like) has specificity towards NADP(H), with high catalytic efficiency specificity for retinoids and a high $K_m$ for ethanol (Rosell et al., 2003).

Several fungal and bacterial ADH enzymes are tetramers with two zinc atoms per monomer, while the animal and plant ADHs characterized to date are thought to be dimers also with two zinc atoms (Persson et al., 2008). Five distinct ADHs are found in Saccharomyces cerevisiae and Kluyveromyces. ADH classes I and II of S. cerevisiae are cytoplasmic enzymes expressed under fermentative and respiratory conditions. Class III corresponds to a mitochondrial protein. Class IV is distantly related to the other four classes and is probably originated from a bacterium (Williamson and Paquin, 1987). Finally, class V was discovered during sequencing of the S. cerevisiae genome. The function of fungi classes III, IV and V is not completely understood (Wills and Jörnvall, 1979; Young et al., 2000; Ladrière et al., 2000; Thomson et al., 2005).

In plants, the ADH gene family has been intensively studied in order to understand its genetics and molecular evolution. Generally, this family is characterized by a small number of copies and very diverse expression patterns. ADHs are involved in the energy production pathway, converting acetaldehyde into ethanol via fermentation during episodes of low oxygen concentrations or low temperatures (Dolferus et al., 1994). Despite a large number of studies, there does not exist a clear correlation among ADH molecular evolution, function, and structure. Thompson et al. (2007) proposed that functional diversification during evolution has been responsible for site-specific shifts after ADH gene duplication in plants, and they furnished the first three-dimensional model of a plant ADH. Subsequently, they evaluated the impact of functional divergence on Poaceae, Brassicaceae, Fabaceae, and Pinaceae enzymes (Thompson et al., 2010) and identified divergent amino acid residues in three important regions of plant ADH (the loop around the zinc ion, the region of monomer interactions and the active site).

In the present work we investigated the relationship among the different ADH classes of animals, fungi, and plants. Moreover, we identified the amino acid residues crucial for different types of functional divergence between duplicate genes using evolutionary and modeling tools in order to better understand the ADH diversification process.

Materials and Methods

Source of the data and sequence alignment

We obtained our protein data set from National Center of Biotechnology Information (NCBI). It consists of ADH amino acid sequences from the phyla Chordata (Classes Myxin, Actinopterygii, Elasmobranchii, Sarcopterygii, Amphibia, Reptilia, Aves, and Mammalia), Mollusca (Class Cephalopoda), Nematoda (Class Chromadorea), Platyhelminthes (Class Turbellaria), and Ascomycota (Classes Saccharomycetes, Sordariomycetes, and Eurotiomycetes). Plant amino acid sequences used in our previous studies (Thompson et al., 2007) were incorporated in the analysis. Thus, 190 protein sequences composed the complete protein dataset. Moreover, we also downloaded 46 nucleotide alcohol dehydrogenase sequences from the NCBI server to evaluate the occurrence of positive selection. Protein alignments were performed using the PRANK software (Whelan and Goldman, 2001; Löytynoja and Goldman, 2005) with default settings. After manual inspection using Aliview (Larsson, 2014) software, we excluded the positions 40-74, 76-97 and 521-572. Furthermore, we used the TranslatorX (Abascal et al., 2010) program to align DNA sequences based on their corresponding manually adjusted protein alignment. Alignments are available upon request.

Phylogenetic analysis

We performed the selection of the best-fit models of amino acid for the maximum likelihood (ML) and Bayesian Inference (BI) analyses with the ProtTest program version 3.4.2 (Darriba et al., 2011) using a fast strategy (optimization of model, branches, and topology of the tree) and without restricting the set of protein evolution candidate models. The program calculates a BIONJ tree, which is a distance based on a phylogeny reconstruction algorithm with better topological accuracy than Neighbor Joining (NJ) in all evolutionary conditions (Gascuel, 1997). The ProtTest program also uses the following criteria: Akaike Information Criterion (AIC, Akaike, 1974; Posada and Crandall, 2001), Corrected Akaike Information Criterion (AICc, Burnham and Anderson, 2003), Bayesian Information Criterion (BIC, Schwarz, 1978), and Decision Theory (DT). These criteria evaluate the relative importance and the model-averaged estimate of parameters. AICc and BIC include penalties for sample size. The jModelTest software (Posada, 2008) was used to evaluate the best evolutionary model for DNA sequences jointly with the use of the AIC, AICc, BIC, and DT criteria for the selection of the best model.

ADH phylogenies were estimated using Neighbor Joining (NJ; Saitou and Nei, 1987), available in the MEGA program version 6 (Kumar et al., 2008, 2016), ML methods through the PhyML program (Phylogenetic Maximum
Likelihood; Guindon and Gascuel, 2003), and BI using MrBayes version 3.2.4 (Ronquist et al., 2012).

We applied p-distance, the Poisson-corrected amino acid distances, and the complete and pairwise deletion of gaps/missing data with 2,000 bootstrap repetitions to analyze the amino acid sequences using the NJ method. PhyML performed the analyses using the best models of protein and nucleotide sequence evolution that resulted from the ProtTest and jModelTest, respectively. This calculates an initial BIONJ tree and applies an approximate likelihood-ratio test (aLRT) for branch support. This approach is based on the conventional LRT principle. However, it is a faster test since the log-likelihood value $l_2$ is computed by optimizing over the branch of interest and the four adjacent branches, whereas other parameters are fixed at their optimal values corresponding to the best ML tree (Anisimova and Gascuel, 2006). We used four chains of 1,000,000 generations, a burn-in of 25% as criteria, and the best evolutionary models identified by ProtTest and jModelTest for Bayesian inference. An average standard deviation of split frequencies equal to or smaller than 0.01 was the convergence criterion. The consensus tree was constructed considering a 50% majority rule consensus. Finally, we used FigTree version 1.4.2 and MEGA to visualize and edit the resulting phylogenies.

Selection and functional diversification analysis

Branch lengths of the tree topologies were calculated using the M0 model available in the CODEML program of the PAML package (Yang, 2007) and, subsequently, the presence of positive selection was evaluated through the maximum likelihood models recommended by Yang (2007) using alcohol dehydrogenase DNA sequences. We carried out a series of LRTs to investigate whether $\omega$ was significantly different from 1 for each pairwise comparison: M1a vs. M2a, M0 vs. M3, and M7 vs. M8. LRT performs the comparison both with the constraint of $\omega=1$ and without such constraint: $LR=2(lnL_1-lnL_2)$ These LRT statistics approximately follow a chi-square distribution and the number of degrees of freedom is equal to the number of additional parameters in the more complex model (Anisimova et al., 2001, 2002). We applied the Naive Empirical Bayes (NEB) and Bayes Empirical Bayes (BEB) approaches available in the PAML package to calculate the posterior probability that each site belongs to the positively selected class.

It is important to note that a relationship between a statistically detectable positive selection ($\omega>1$) and functional divergence might not necessarily exist (Tennesen, 2008). Thus, to investigate further if any amino acid replacement could have led to adaptive functional diversification, we estimated the Type-I divergence by posterior analysis using DIVERGE version 3 (Gu and Vander Velden, 2002; Gu, 2006). The latter evaluates shifted evolutionary rates and altered amino acid properties after gene duplication (Gu, 2006). Type-I functional divergence (site-specific rate shift) refers to the evolutionary process resulting in site-specific rate shifts after gene duplication. It identifies amino acid residues highly conserved in one gene copy and highly variable in the other. The probability of a residue being under Type-I divergence is denoted $Q_i(k)$ is the site $(k)$-specific score corresponding to the posterior probability that site $k$ is related to type-I functional divergence (Zheng et al., 2007).

Three-dimensional structures of alcohol dehydrogenase were downloaded from the RCSB Protein Data Bank (Berman et al., 2000) to evaluate the impact of potential divergent amino acid residues. Moreover, PyMOL software version 1.8.4.2 was used to display and visualize Homo sapiens (ADH1, PDB ID 1HDX), Saccharomyces cerevisiae (ADH1, PDB ID 4W6Z), and Gadus morhua (ADH1, PDB ID 1CDO) structures.

Results

Phylogenetic analysis

In total, we performed a comparative phylogenetic analysis using 190 ADH amino acid sequences from animals, fungi and plants. The taxonomic classification, ADH types, accession numbers, and sequence sizes are shown in Table S1 (Supplementary Material). The best protein evolutionary model was LG (Le and Gascuel, 2008), with a proportion of invariable sites (+I) and rate variation among sites with a number of rate categories in the gamma distribution (+G), whereas GTR (Lanave et al., 1984; Tavaré, 1986) with a gamma distribution (+G) was the best evolutionary model for DNA sequences.

The tree topologies resulting from the BI (Figure 1A) and ML (Figure 1B) methods do not differ significantly, especially when major clades are considered. We identified three monophyletic groups, corresponding to fungi, plants, and a larger group formed by animals. Additionally, we identified a clade composed by ADH sequences from the phylum Nematoda, which includes two Caenorhabditis elegans sequences (ADH1 and ADH2) that are placed close to the tetrameric fungal ADHs (Figure 1). Within a large group of ADH3s from animals it is interesting to note that C. elegans ADH3 clustered with those of Octopus vulgaris (Phylum Mollusca) and Schmidtea mediterranea (a freshwater planarian from Phylum Platyhelminthes). The invertebrate ADH3s formed a highly supported monophyletic group in BI phylogeny (Figure 1A). Mammalian, avian, reptilian, amphibian and Elasmobranch ADH3s also formed a monophyletic cluster (Figure 1).

Most of the ADH1s were located in a large set that includes chordate ADH1, amphibian ADH8 and mammalian ADH4 and ADH5, with high bootstrap support for the individual clusters within the considered group (Figure 1). This form is the classical and highly variable liver enzyme responsible for ethanol metabolism. In fishes, we detected...
only two ADH groups: ADH3 and a second mixed class (here named ADH1, but also called ADH8 in the literature) that is separated from all other ADH1 forms (Figure 1 and Figure S1). Actinopterygian ADH1 seems to be basal to the highly supported clade formed by class III and non-class III ADHs (Figure 1). Mammalian ADH4s are highly similar to ADH1 in terms of primary sequence and are placed close to them in the phylogenetic tree (Figures 1 and 2). ADH2 is found in mammalian and avian/reptilian lineages, forming a sister group to tetrapod non-class III proteins (Figure 1). There was a distinct cluster of amphibian ADH8 close to Amphibian ADH1 (Figure 1 and Supplementary Material Figure S2) in the phylogenetic tree.

Phylogenetic relationships among mammalian ADH sequences are displayed in more detail in Figure 2, where monophyletic groups were formed according to ADH type. ADH1 showed sub-clusters (ADH1A, ADH1B, ADH1C), corresponding to different isoenzymes. Both in Figures 1 and 2, ADH4 was placed close to ADH1, suggesting that it originated from an ADH1 duplication. Mammalian ADH5 was placed close to avian ADH1 and, together with amphibian ADH1, formed a sister group in relation to a cluster that includes amphibian ADH1 and ADH8, mammalian ADH1 and ADH4, and ADH1 from Aves and Reptilia (Figure 1).

A new form (ADH8) appeared in amphibians, and it formed a separated cluster from ADH1 and ADH3 (Figure S2). Reptile ADH3 sequences formed a distinguishable group from ADH1 (Figure 1 and Figure S3). In addition to ADH1 and ADH3, ADH2 appears basal in relation to ADH1 in both mammals and birds (Figure 1), and ADH3 was basal to all sequences in these two animal groups.

A more complex pattern of sequence duplication was seen in fungi (Figure 3), where the ADH sequences clustered according to ADH type and fungi genera. A larger cluster composed by Saccharomycetes sequences is distinguishable. Additionally, Sordariomycetes and Eurotiomycetes ADHs formed distinct monophyletic groups. Our

Figure 1 - Evolutionary history of alcohol dehydrogenase proteins from plants, fungi and animals. A. Bayesian Inference; B. Maximum Likelihood. Numbers represent posterior probability and aLRT non-parametric branch support, respectively. Only values higher than 0.7 are shown. Scale bar indicates levels of sequence divergence.
phylogenetic analysis grouped sequences from *Saccharomyces* by ADH type, with ADH2 closer to ADH1. *Saccharomyces* ADH1, ADH2 and ADH5 are probably derived from a common ancestor. ADH1 and ADH2 forms from *Lachancea* grouped together; ADH4 from *Kluyveromyces* formed a different group, as well as *Saccharomyces* ADH3 and ADH5. The *Yarrowia* and *Candida* sequences were also separated according to ADH type. *Pichia* ADHs did not form a monophyletic cluster, whereas ADH3 from *Kluyveromyces* and *Lachancea* clustered together.

### Selection and functional diversification analyses

Generally there was no indication of positive selection acting on *Adh* genes (Table 1), because LRTs comparing M1 (neutral) and M2 (selection), as well as M7 (beta) and M8 (beta & \(\omega\)) were not statistically significant considering 0.01 as a cutoff. Additionally, the NEB and BEB approaches did not identify any site with posterior probability equal or higher than 0.95. However, the LRT comparing M0 (one-ratio) against M3 (discrete) was highly significant, indicating that selective pressure is highly variable among sites.

Coefficients of functional divergence (\(\theta\)) of pairwise comparisons between mammalian, fishes, and fungal alcohol dehydrogenases are reported in Table 2. They showed statistically significant site-specific shifts of evolutionary rates, with \(\theta\) varying markedly from 0.35 to 0.85. We used a site-specific profile based on the posterior probability (\(Q_k\)) to identify amino acid residues responsible for functional divergences after gene duplication or speciation. To reduce false positives, a conservative cut-off value was empirically used: \(Q_k \geq 0.90\). Functionally important amino acid residue positions between the mammalian ADH forms and their respective \(Q_k\) values are shown in Table 3, whereas those important for the differentiation between fungi and fish forms are listed in Tables 4 and 5, respectively.

For mammals (Table 3), one site (253) seemed to be especially important for the differences between ADH3/ADH2 and ADH3/ADH5. The sites 44, 228 and 246 were also identified as divergent for ADH5/ADH3, whereas site number 54 was so for ADH1/ADH2. A number of differences in functionally important sites occurred mainly between ADH4 and ADH1 (33 sites) and ADH3/ADH5 (4 sites). Site number 122 showed a \(Q_k = 0.95\) for the ADH1/ADH4 comparison. We located sites 44, 54, 122, 228, 246 and 253 in the three-dimensional structure of human ADH1 (PDB ID: 1HDX, corresponding to sequence AAA51884; Table S1). They were located in a \(\alpha\)-strand, an \(\alpha\)-helix near NAD, a coil close to a zinc ion, and in a coil, \(\alpha\)-helix, and \(\beta\)-strand in the molecular surface, respectively (Figure 4A).

For fungi (Table 4), several sites accounted for differences between ADH1 and ADH5 from *Saccharomyces*. In fact, there are 18 sites, considering a \(Q_k \geq 0.95\). Additionally, ADH5 from this fungus was identified as functional divergent from ADH4 from *Kluyveromyces* (ADH4\(^S\)) and ADH3 from *Kluyveromyces* and *Lachancea* (ADH3\(^KL\)). These sites, 271, 272, 279, and 280 (\(Q_k \geq 0.95\)), were likely responsible for the divergence between ADH3\(^KL\) and ADH5\(^S\), whereas sites number 126 and 320

---

**Figure 2** - Phylogenetic tree of alcohol dehydrogenase proteins from mammals obtained by the neighbor-joining algorithm. Numbers represent bootstrap values; values higher than 80% are shown. Scale bar indicates levels of sequence divergence. Clusters distinguishable by ADH type are highlighted.
(Qk \geq 0.95) are associated with that between ADH4S/ADH5S. We identified these amino acids in the three-dimensional structure of Saccharomyces cerevisiae ADH1 (PDB ID 4W6Z, chain A). The sites 126, 271, 272, 279, and 280 corresponding to coil, coil, coil and an \alpha-helix, respectively, are all located in the protein molecular surface, whereas site 320 corresponds to a \beta-strand important to interaction with chain B of this tetrameric ADH (Figure 4B).
ADH1 and ADH3 from fishes are also functionally divergent, as indicated by the $Q_k$ values for specific amino acids. The sites 302, 328 and 355 all showed a $Q_k/\omega_1 > 0.90$ (Table 5). They were identified in the 3D structure of *Gadus morhua* ADH3 (PDB ID 1CDO, chain A). The first two are close to the NAD (nicotinamide-adenine-dinucleotide) coenzyme, while the site number 355 is in a coil in the molecular surface (Figure 4C).

## Discussion

Gene duplication is an important precursor of evolutionary diversification. The majority of new genes originate through duplication, chromosomal rearrangement, and the subsequent divergence of pre-existing genes (Lawton-Rauth, 2003). The existence of several multigene families is an indication of the importance of gene duplication in the
 Phylogenetic analysis has been a powerful approach to investigate the role of gene duplications in evolution. The alcohol dehydrogenase enzymes form a large and diverse family that has contributed to the understanding of protein evolution, enzymatic mechanisms, metabolic functions, and regulatory roles. They show chemically modified sub-forms, isoenzymes, classes, and separate enzymes, presenting a wide range of distinct functions, as well as redundancy with overlaps in activity (Jörnvall, 2008). We have theoretically demonstrated that different plant ADH forms may be submitted to an evolutionary diversification pro-

<table>
<thead>
<tr>
<th>Amino acid residues</th>
<th>ADH1/ADH4</th>
<th>ADH1/ADH2</th>
<th>ADH5/ADH3</th>
<th>ADH3/ADH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>44 (Val41)</td>
<td></td>
<td></td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>54 (His51)</td>
<td></td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>122 (Leu112)</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>142</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>228 (Ala213)</td>
<td></td>
<td></td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>246 (Lys231)</td>
<td></td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>248</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>253 (Thr238)</td>
<td></td>
<td>0.96</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>257</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>261</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>262</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>271</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>281</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*In bold are amino acid residues with $Q(k) \geq 0.95$. The correspondent amino acid residues in the three-dimensional structure of human ADH1 (PDB ID 1HDX, Figure 4A) are indicated.
cess that occurred after gene duplication (Thompson et al., 2007, 2010). The next step was to evaluate the importance of this process in ADHs of other organisms, to obtain a comprehensive panorama for ADH molecular evolution.

We identified three monophyletic groups composed by fungi, plants and animals. Glasner et al. (1995) analyzed a smaller number of sequences (22) and found a similar pattern of evolution for these proteins. The identification of two Caenorhabditis elegans sequences (ADH1 and ADH2) close to the tetrameric fungal ADHs (Figure 1) agrees with that obtained by Glasner et al. (1995) who described for the first time fungal-like ADH sequences among metazoans. Both C. elegans ADH forms show ethanol degradation activity, preferentially for longer alcohols. It may be possible that additional fungal-like sequences will be discovered in other animals or plants, which could be explained by one or multiple deletions in lineages generating the modern plants and animals, or it may be the result of convergent evolution (Glasner et al., 1995).

We also identified a close evolutionary relationship among ADH3s from C. elegans, Octopus vulgaris, and Schmidtea mediterranea, a freshwater planarian, within the large group of ADH3s from all animals. Godoy et al. (2007) also found a close relationship between the S. mediterranea and C. elegans ADH3s. Kaiser et al. (1993) described the O. vulgaris ADH3, which was the first-detected group of animals that lack ethanol dehydrogenase activity. No other ADH classes are present in planarians also, as suggested by in silico analysis that indicated that only one contig was sufficient to account for the cDNA and 40 trace sequences from the planarian databases (Godoy et al., 2007).

We observed a monophyletic cluster of ADH3 (Figure 1) in this work. This enzyme is widely known as a glutathione-dependent formaldehyde dehydrogenase that can oxidize ethanol at high concentrations (Dasmahapatra et al., 2001) but preferentially metabolizes longer aliphatic and aromatic alcohols (Reimers et al., 2004). ADH3 has been described as a ubiquitous enzyme in vertebrates (Funkenstein and Jakowlew, 1996), with a spatio-temporal regulation in zebrafish development (Dasmahapatra et al., 2001; Cañestro et al., 2003). Additionally, ADH3 is found in the cell nucleus, where it may have a probable DNA protection function (Iborra et al., 1992; Fernández et al., 2003), differently from the other ADHs, which commonly have a cytosolic location (González-Duarte and Albalat,
In invertebrates, its expression is mainly found in digestive tissues (Godoy et al., 2007). We demonstrated that sites 44, 228, 246, and 253 seem to be fundamental for the divergence of ADH3/ADH5 and ADH3/ADH2 in mammals. ADH1 is the classical liver enzyme responsible for ethanol metabolism. In fishes we identified ADH3 and a second mixed class (as previously remarked, it is here named ADH1, but also called ADH8 in the literature) that is structurally similar to class III but functionally similar to ADH1 (the classical alcohol-metabolizing enzyme; Dasmahapatra et al., 2005). This hybrid characteristic may explain why the Actinopterygii ADH1 cluster is separated from all other ADH1s (Figure 1). Fishes constitute the first vertebrate class with documented expression of more than one ADH class (Dasmahapatra et al., 2005). In this report we identified some amino acid residues important for functional differentiation between ADH1 and ADH3. They are located in regions of functional importance, such as those close to the NAD coenzyme and the zinc ion.

ADH1 has tissue-specific expression and is involved in different metabolic pathways, such as ethanol oxidation, norepinephrine, dopamine, serotonin and bile acid metabolism (Höög et al., 2001), oxidation of retinol in vitro (Boleda et al., 1993) and in vivo (Deltour et al., 1999). It is highly expressed in the liver and also significantly expressed in the uterus, adrenal, small and large intestine, kidney, testis, and epididymis (González-Duarte and Albalat, 2005). The ADH1 structure has three conserved positions, His67, Glu68, and Phe140, which have been proposed as a signature for class assignment (Norin et al., 1997), and three variable segments near the substrate-binding pocket and the subunit interaction region. In contrast, these regions are among the most conserved in ADH3 (Cañestro et al., 2003). It is important to note that preservation of those previously cited conserved amino acids does not necessarily imply ethanol-oxidizing activity (Reimers et al., 2004). Additionally, there are two main domain conformations of ADH1 described as ‘open’ in the apoenzyme and ‘closed’ in the binary and ternary complexes. Different substrate specificity and kinetic mechanisms of ADH1 and ADH3 may be due to these ‘open’ and ‘closed’ conformations (Sanghani et al., 2002).

Mammalian ADH4s were placed close to ADH1 in the phylogenetic tree (Figures 1 and 2), which suggests that it originated from ADH1 duplication. Our results corroborated the hypothesis proposed by González-Duarte and Albalat (2005) that ADH4 may be the result of a mammalian-specific Adh1 duplication, since this class has not been detected in birds or reptilians (Figure 1). Estonius et al. (1994), Parés et al. (1994) and Strömberg and Höög (2000) obtained similar results. In mammals, ADH4 is specifically expressed in epithelial tissues, such as stomach mucosa (Parés et al., 1994). ADH4 functions in retinoid oxidation in vitro (Boleda et al., 1993). However, ADH4-null mutant mice showed weak phenotypic effects, which may indicate a contribution in specific routes, not involved in systemic retinol metabolism (Deltour et al., 1999). In this work we identify a significant functional divergence of mammalian ADH4 and ADH1, with some amino acid residues of these differences located in functional important regions, such as site no. 122 close to the zinc ion.

Figure 4 - Three-dimensional structures of alcohol dehydrogenase from: A. Homo sapiens (ADH1, PDB ID 1HDX, chain A), B. Saccharomyces cerevisiae (ADH1, PDB ID 4W6Z, chain A), and C. Gadus morhua (ADH1, PDB ID 1CDO chain A). Blue spheres = zinc ions. Blue bars = nicotinamide-adenine dinucleotide (NAD) in A and C, and nicotinamide-8-ido-adenine dinucleotide (8ID) in B. Amino acids responsible for functional divergence and their respective position in the PDBs are indicated in the figures.
ADH2 was found in mammalian and avian/reptilian lineages forming a sister group to tetrapod non-class III proteins, reinforcing the results of Hjelmqvist et al. (1995b). Based on the phylogenetic analysis, as well as biochemical and structural characteristics (Höög et al., 2001, González-Duarte and Albalat, 2005), it is reasonable to suggest that ADH2 is derived from a tetrapod ADH3. ADH2 proteins have higher $K_m$ values toward ethanol and preferentially metabolize larger aliphatic and aromatic alcohols/alddehydes (Reimers et al., 2004). Moreover, they are structurally more divergent than the ADH1 forms, for which variation is classically known (Hjelmqvist et al., 1995a). A functionally important site (54, close to the zinc ion in the ADH three-dimensional structure) seems to be important for ADH1/ADH2 divergence in mammals.

Amphibian ADH8 formed a distinct cluster, which confirms the distinct characteristics of ADH8, such as a large active site pocket, very different proton-relay pathway, very specific rearrangements in the phosphate-binding site cofactor, and weak interactions of the adenine moiety (Rosell et al., 2003). This form has a unique NADP(H) specificity and was first described as ADH4-like. However, these characteristics led to its classification in a new class (Rosell et al., 2003).

In relation to the alcohol dehydrogenases from fungi, the ADH1-ADH2 duplication seems to have occurred before the divergence of the Saccharomyces species and after the divergence between Saccharomyces and Kluyveromyces, which has been estimated to have occurred 80 ± 15 million years ago (Thomson et al., 2005). Indeed Saccharomyces ADH1, ADH2 and ADH5 probably derived from a common ancestor, as suggested by Ladrière et al. (2000). Moreover, ADH5 has the highest rate of ADH sequence divergence. In this report, ADH5 was shown to be functionally divergent from ADH1. Saccharomyces ADH1 and ADH2 are cytoplasmatic enzymes acting in the fermentation and gluconeogenesis processes, respectively, while ADH3 is located in the mitochondria (de Smidt et al., 2008). Kluyveromyces ADH has two cytoplasmatic (ADH1 and ADH2) and two mitochondrial (ADH3 and ADH4) enzymes. In the present work, we have shown that ADH4$^K$ and ADH3$^KL$ are functionally divergent from Saccharomyces ADH5. We recall that Lertwattanasakul et al. (2007) have proposed that Kluyveromyces marxianus ADHs have distinct roles in cells, because the different Adh genes are differentially expressed depending on growth phase and carbon source. Since the Saccharomyces and Kluyveromyces genomes are similar, while their ADH sequences have been submitted to different rates of divergence (Ladrière et al., 2000), they may have a lower structural constraint or submission to a functionally divergence process, and this could lead to new enzyme functions. Indeed, this seems to occur in animals (Höög et al., 2001) and was theoretically demonstrated in plants (Thompson et al., 2007).

Natural selection has been described as responsible for the evolution of many genes (Hey, 1999). A widely used method to detect positive selection is through the ratio of nonsynonymous to synonymous rates ($\omega = d_\text{ns}/d_\text{s}$). It is assumed that synonymous substitutions are neutral, whereas the nonsynonymous are subject to selection. Consequently, a $\omega$ statistically higher than 1 would indicate the action of positive selection or a relaxed selective constraint, whereas low $d_\text{ns}/d_\text{s}$ values would mean conservation of the gene product due to purifying selection (Tennesen, 2008). Although we did not directly identify positive selection acting on the alcohol dehydrogenase genes, there appears to be variable selective pressure acting among sites, as indicated by LRT when the M0 (one-ratio) and M3 (discrete) models are compared. Therefore, we tested if any amino acid replacement could have led to adaptive functional diversification and the results indicated that there are some sites in different species that exhibit different evolutionary rates and altered amino acid properties after gene duplication, but experimental structural-functional studies are mainly restricted to the ADH1 and ADH3 enzymes. Future theoretical and experimental studies are needed to establish the impact of these amino acid replacements in the ADH structure and function. For instance, docking and molecular dynamics simulations could add valuable information about the functional divergence of these proteins.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and the Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM) of Universidade Federal do Rio Grande do Sul.

References


Boleda MD, Saubi N, Farrés J and Parés X (1993) Physiological substrates for rat alcohol dehydrogenase classes: Aldehydes...


Zgombic-Knight M, Ang HL, Foglio MH and Duester G (1995) Cloning of the mouse class IV alcohol dehydrogenase (retinol dehydrogenase) cDNA and tissue-specific expression


Internet Resources


Supplementary material

The following online material is available for this article:

Table S1: Alcohol dehydrogenase proteins used in this study.

Figure S1: Evolutionary relationships of fish ADH proteins.

Figure S2: Evolutionary relationships of amphibian ADH proteins.

Figure S3: Evolutionary relationships of reptilian ADH proteins.

Figure S4: Evolutionary relationships of avian ADH proteins.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.