Biochemical Genetics of Opossum
Aldehyde Dehydrogenase 3: Evidence for
Three ALDH3A-Like Genes and an ALDH3B-
Like Gene

Author
Holmes, Roger

Published
2010

Journal Title
Biochemical Genetics

DOI
https://doi.org/10.1007/s10528-009-9318-3

Copyright Statement
Copyright 2010 Springer Netherlands. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. The original publication is available at www.springerlink.com

Downloaded from
http://hdl.handle.net/10072/30852
BIOCHEMICAL GENETICS OF OPOSSUM ALDEHYDE DEHYDROGENASE 3. Evidence for three ALDH3A-like genes and an ALDH3B-like gene

Roger S Holmes

School of Biomolecular and Physical Sciences, Griffith University, Nathan 4111 Brisbane Queensland Australia

Email: r.holmes@griffith.edu.au

ABSTRACT

Mammalian ALDH3 isozymes participate in peroxidic and fatty aldehyde metabolism, and in anterior eye tissue UV-filtration. BLAT analyses were undertaken of the opossum genome using rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2 amino acid sequences. Two predicted opossum ALDH3A1-like genes and an ALDH3A2-like gene were observed on chromosome 2; as well as an ALDH3B-like gene, which showed similar intron-exon boundaries with other mammalian ALDH3-like genes. Opossum ALDH3 subunit sequences and structures were highly conserved, including residues previously shown to be involved in catalysis and coenzyme binding for rat ALDH3A1 by Liu and coworkers (1997). Eleven glycine residues were conserved for all of the opossum ALDH3-like sequences examined, including two glycine residues previously located within the stem of the rat ALDH3A1 active site funnel. Phylogeny studies of human, rat, opossum and chicken ALDH3-like sequences indicated that the common ancestor for ALDH3A- and ALDH3B-like genes predates the appearance of birds during vertebrate evolution.

INTRODUCTION

Mammalian aldehyde dehydrogenases (ALDH; EC 1.2.1.3) are encoded by a super-family of ALDH genes which oxidize a wide range of endogenous aldehydes in the body, generated in many metabolic pathways (eg. alcohols, lipids, amino acids), as well as exogenous aldehydes, derived from environmental substances and from drugs (http://aldh.org/superfamily.php) (Vasiliou & Nebert, 2005; Sophos et al.,
Mammalian ALDH3 isozymes have been investigated predominantly because of their roles in fatty, peroxidic and aromatic aldehyde metabolism (Algar & Holmes, 1989; Hempel et al., 1989), and their high expression levels in stomach, cornea, liver hepatoma cells and liver microsomes (Timms & Holmes, 1982; Miyauchi et al., 1991).

Four major ALDH3-like genes have been investigated in humans and rodents: ALDH3A1 encodes the major cytosolic stomach, corneal and tumor-associated isozyme (Jones et al. 1988; Holmes et al. 1988; Vasiliou et al. 1994), acts as a corneal ‘crystallin’ and UV-filter in mammalian anterior eye tissues (Abedinia et al. 1990; Lassen et al. 2007), and promotes survival of human corneal epithelial cells (Pappa et al. 2005); AL3A2 encodes the liver microsomal fatty aldehyde ALDH (Miyauchi et al. 1991; Rogers et al. 1997), for which gene mutations in human populations have been implicated in the Sjogren-Larsson syndrome (De Laurenzi et al. 1996); ALDH3B1 encodes a cytosolic ALDH with high expression levels in human lung, prostate, kidney and some tumors (Hsu et al. 1994a; 1994b); and ALDH3B2 encodes a high expression ALDH in human salivary gland, placenta and some tumors (Hsu et al. 1997).

This study describes the predicted sequences, structures and phylogeny of ALDH3 genes and enzymes in a South American marsupial, the gray short-tailed opossum (Monodelphis domestica), a marsupial animal model used to study the genetics of UV-induced eye cancers and lipid transport proteins (VandeBerg et al. 1994; Rainwater et al. 2001). In silico methods were used to predict the primary, secondary and tertiary structures for opossum ALDH3 isozymes and gene locations for opossum ALDH3-like genes, using data from the opossum genome sequence (Mikkelsen et al. 2007). This paper extends previous studies which examined the tissue and subcellular distribution and biochemical properties of opossum ALDHs (Holmes et al. 1990; 1991).
and the predicted structures and properties of opossum ALDH1-like genes and proteins (Holmes, 2009). Phylogenetic analyses also describe the relationships and potential evolutionary origins of multiple opossum ALDH3A-like genes and an ALDH3B-like gene in the opossum.

**MATERIALS AND METHODS**

**Opossum ALDH Gene and Enzyme Identification.**

BLAT (BLAST-Like Alignment Tool) in silico studies were undertaken using the UC Santa Cruz web browser [http://genome.ucsc.edu/cgi-bin/hgBlat](http://genome.ucsc.edu/cgi-bin/hgBlat) (Altschul et al. 1990; Kent et al. 2002) with the default settings. GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] and UniProtKB/Swiss-Prot Database [http://au.expasy.org] sequences for rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2 (Table 1) were used to interrogate the opossum genome sequence. Gene locations, predicted gene structures and ALDH protein subunit sequences were observed for each ALDH examined for those regions showing identity with the respective opossum ALDH gene products (Table 1).

**Predicted Secondary and Tertiary Structures for Rat and Opossum AL3A and AL3B Subunits.**

Predicted secondary structures for rat ALDH3A2, ALDH3B1 and ALDH3B2 subunits, and for opossum ALDH3A1, ALDH3A2, ALDH3A3 and ALDH3B1 subunits were obtained using the PSIPRED v2.5 web site tools provided by Brunel University [http://bioinf.cs.ucl.ac.uk/psipred/psiform.html](http://bioinf.cs.ucl.ac.uk/psipred/psiform.html) (McGuffin et al. 2000). Predicted tertiary structures for opossum ALDH3A1, ALDH3A2 and ALDH3B1 subunits were obtained using the SWISS MODEL web tools [http://swissmodel.expasy.org/workspace](http://swissmodel.expasy.org/workspace) (Schwede et al. 2003). Tertiary structures for rat ALDH3A1 (Liu et al., 1997) served as a reference for obtaining the opossum ALDH tertiary structures: ALDH3A1 (residues 4-448)
and ALDH3A2 (residues 5-443) using the rat ALDH3A1 template at 2.6 Å resolution (PDB ID 1ad3B); and ALDH3B1 (residues 6-443) using the template for the rat ALDH3A1-NAD complex at 2.6 Å resolution (PDB ID 1ad3A).

Alignment of Mammalian ALDH Active Site Residues

Alignments of rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2, and of predicted opossum ALDH3A1, ALDH3A2, ALDH3A3 and ALDH3B1 sequences were undertaken using a ClustalW-technique (http://www.ebi.ac.uk/clustalw/) (Chenna et al. 2003) and previously reported sequences for the rat ALDH3A1 (Jones et al., 1988), ALDH3A2 (Miyauchi et al. 1991), ALDH3B1 and ALDH3B2 (Mammalian Gene Collection Team, 2004) ALDH sequences (Table 1).

Prediction of Transmembrane Regions for Rat and Opossum ALDHs

Predictions of transmembrane helices for rat and opossum ALDH3 sequences were conducted using the web resources of the Center for Biological Sequence Analysis, Technical University of Denmark TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

Phylogenetic Studies and Sequence Divergence

Phylograms were constructed using a ClustalW-derived amino acid alignment of ALDH protein sequences, obtained with default settings and corrected for multiple substitutions (http://www.ebi.ac.uk/clustalw/) (Chenna et al. 2003). An alignment score was calculated for each aligned sequence by first calculating a pairwise score for every pair of sequences aligned. The alignment ambiguous amino and carboxy terminus regions were excluded prior to phylogenetic analysis yielding alignments of 372 residues of human, rat, opossum and chicken ALDH sequences (Table 1). Pairwise scores were calculated using the number of identities in the best alignment divided by the number of residues compared. Scores were initially calculated as percent identity.
scores and were converted to distances by dividing by 100 and subtracting from 1.0 to give the number of differences per site. Percentage sequence identities for the ALDH3 family subunits examined were determined using the SIM-Alignment tool for Protein Sequences (http://au.expasy.org/tools/sim-prot.html) (Schwede et al. 2003).

RESULTS AND DISCUSSION

Alignments of predicted opossum ALDH3-like amino acid sequences with rat ALDH3-like sequences.

The deduced amino acid sequences of four predicted opossum class 3 ALDH subunits (designated as ALDH3A1, ALDH3A2, ALDH3A3 and ALDH3B1) are shown in Figure 1 together with previously reported sequences for rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2 (see Table 1). The alignments showed high levels of sequence identities for rat ALDH3A1 sequences with the opossum ALDH3A1 and ALDH3A3 sequences (73% and 74%, respectively); for the rat and opossum ALDH3A2 sequences (80%); and the opossum ALDH3B1 sequence with the rat ALDH3B1 and ALDH3B2 sequences (75% and 71%, respectively) (Table 2). This supports a proposal that the identified opossum ALDH sequences are members of the ALDH3 family and are products of the same ALDH3 gene sub-class in each case: opossum ALDH3A1 and ALDH3A3, being ALDH3A1-like; opossum ALDH3A2, ALDH3A2-like; and opossum ALDH3B1 being an ALDH3B-like gene. In contrast, the sequences for rat and opossum ALDH3-like ALDHs from different sub-classes showed lower levels of identity (51-55%) (Table 2). Comparisons of predicted opossum ALDH sequences with the rat ALDH3A1 sequence (for which the 3-dimensional structure has been previously reported) (Liu et al., 1997) enabled identification of key residues which may contribute to catalysis and function (Figure 1; Table 3). Active site residues (rat ALDH3A1 numbers used), which bind the substrate (Glu209; Cys243) or stabilize the
transition state for the catalyzed reaction (Asn116) (Hempel et al., 1996; Liu et al., 1997) have been strictly conserved for all of the predicted opossum ALDH3A and ALDH3B sequences. In addition, the rat ALDH3A1 novel dinucleotide-binding motif and NAD(P) binding domain near the N-terminal end of the αD helix (188Gly-Ser-Thr-Ala-Val-193Gly) (Liu et al., 1997) has been predominantly retained for the opossum ALDH3A1 and ALDH3A3 sequences (a 190Ala/Gly substitution was observed) but has undergone further substitutions for the rat and opossum ALDH3A2 and ALDH3B sequences. Eleven glycine residues were conserved for all of the rat and opossum ALDH3 sequences examined (Table 3) which is consistent with the roles proposed by Hempel and coworkers (1996) in contributing to the structure of rat ALDH3A1. Moreover, glycine residues 187 and 211 have been located in the 3D structure for this enzyme within the stem of the active site funnel, and these have also been conserved for all of the rat and opossum ALDH3-like sequences examined. Three proline residues were also retained (Pro103, 317 and 337) which have been previously shown to contribute to the three dimensional structure of rat ALDH3A1 (Liu et al., 1997).

**Predicted secondary and tertiary structures for opossum ALDHs**

Predicted secondary structures for opossum ALDH3A1, ALDH3A3, ALDH3A2 and ALDH3B1 and for rat ALDH3A2, ALDH3B1 and ALDH3B2 subunits were compared in Figure 1, together with the previously reported secondary structure for rat ALDH3A1 (Liu et al., 1997). Similar α-helix and β-sheet structures were observed for all of the ALDH3 subunits examined. The most significant differences related to the predicted additional α-helix at the carboxy-terminus for the rat and opossum ALDH3A2 sequences, which exhibited a high probability as a transmembrane sequence (Figure 3). Given the reported location of human and mouse ALDH3A2 within the liver endoplasmic reticulum ((Timms & Holmes, 1982; Miyauchi et al. 1991), it is likely that this sequence
contributes to the membrane-bound location for this enzyme. In contrast, rat and opossum ALDH3A1- and ALDH3B1-like sequences, and the rat ALDH3B2 sequence, showed only low probability regions for potential transmembrane sequences, which supports their location within the cytoplasm, as previously reported for rat and opossum ALDH3A1-like sequences (Jones et al. 1988; Hempel et al. 1996). Moreover, ALDH3A1 serves as a major soluble or 'crystallin-like' corneal protein within several mammalian species, and acts as a UV-filtering agent, protecting anterior eye tissues from UV-induced damage (Abedinia et al. 1990; Lassen et al. 2007).

Predicted tertiary structures for opossum ALDH3A1, ALDH3A2 and ALDH3B1 were also compared with the previously reported structure for rat ALDH3A1 (Liu et al. 1997) (Figure 4). Opossum ALDH3A1 showed a similar structure to that reported for rat ALDH3A1 which contains 3 major domains, the catalytic domain and the NAD binding domain, which are separated by the active site cleft, and an oligomerisation domain, which is involved in forming dimers of the active enzyme (Liu et al. 1997). The three domain structure was also predicted for opossum ALDH3A2 and ALDH3B1 demonstrating that the major tertiary structural features for all three opossum class 3 ALDHs were similar to rat ALDH3A1.

**Predicted gene locations and exonic structures for opossum ALDH3-like genes.**

Table 1 summarizes the predicted locations for the four predicted opossum ALDH3-like genes examined which include two closely localized ALDH3A1-like genes (ALDH3A1 and ALDH3A3), an ALDH3A2-like gene, also on chromosome 2, and an ALDH3B1-like gene. These predictions are based upon BLAT interrogation of the opossum genome (Mikkelsen et al. 2007), using the reported sequences for rat AL3A1 (Jones et al. 1988), AL3A2 (Miyauchi et al. 1991) and AL3B1 (Mammalian Gene Collection, 2004) and the UC Santa Cruz Web Browser (Kent et al. 2002) ([http://genome.ucsc.edu/cgi-bin/hgBlat](http://genome.ucsc.edu/cgi-bin/hgBlat)). The opossum ALDH3A1-like genes (ALDH3A1 and ALDH3A3) were
located only 35 kilobases apart on chromosome 2, whereas the opossum ALDH3A2 gene was ~140 kilobases more distant from the ALDH3A1-like genes. In contrast, the predicted opossum ALDH3B1 gene was located on another as yet unidentified chromosome. The opossum ALDH3A1, ALDH3A2 and ALDH3B1 genes were transcribed on the negative DNA strand, which is comparable with the human ALDH3A1 and rat ALDH3A2 and ALDH3B1 genes, but different from the opossum ALDH3A3, rat ALDH3A1, rat ALDH3B2, human ALDH3A2 and human ALDH3B1 genes, which are transcribed on the positive DNA strand. Predicted exonic start sites for the opossum ALDH3 genes were examined with the opossum ALDH3A1 gene containing 9 exons, which are located in similar or identical positions to the 9 exons identified for the human ALDH3A1 and ALDH3A2 genes; for rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2 genes (Figure 1).

BLAT analysis of the opossum genome using the rat ALDH3A1 sequence, however, revealed evidence for a second ALDH3A1-like gene with 8 exons, which aligned with the amino acid sequences observed for 8 of the 9 opossum ALDH3A1 exons, but lacked the exon 1 sequence. Examination of the nucleotide sequences in this predicted region for exon 1 of the opossum ALDH3A3 gene revealed a gap in the known sequence, which has apparently prevented the identification of this exon. This gap in sequence is reflected in the alignment of amino acids reported for opossum ALDH3A3 (Figure 1), with 54 residues missing at the N-terminus, in comparison with the opossum ALDH3A1 sequence. The presence of this second opossum ALDH3A1-like gene (designated as ALDH3A3) is supported by a BLAT sequence analysis of two exons (exons 3 and 4 for ALDH3A1 and ALDH3A3 genes) and a segment of the intron located between these exons (Figure 4). Exons 3 and 4 show 99 percent sequence identity for both predicted genes whereas intron 3 from both genes were similar, but distinct in sequence with ~90% identity for this region as a result of a number of deletions (or insertions). This result supports a
proposal that ALDH3A1 and ALDH3A3 are distinct but nearly identical genes on chromosome 2 of the opossum, which may have arisen from a process of gene duplication following an unequal crossover event in the recent evolutionary history of the opossum, similar to that reported for the primate γ-globin genes (Fitch et al. 1991).

**Phylogeny and divergence of mammalian and chicken ALDH3A and ALDH3B sequences.**

A phylogram (Figure 5) was calculated by the progressive alignment of human, rat, opossum and chicken ALDH3 amino acid sequences which clustered into three main groups, corresponding to the ALDH3A1-like, ALDH3A2-like and ALDH3B-like proteins and genes. Two of the opossum ALDH3A1-like sequences (ALDH3A1 and ALDH3A3) were grouped together on one branch of the phylogram, with the human and rat ALDH3A1 sequences, indicating that these opossum ALDH genes are products of a recent gene duplication event of an ancestral marsupial ALDH3A1 gene. This is further supported by the tandem locations for these genes on chromosome 2, their high level (98%) of amino acid sequence identity (Table 2), and the near nucleotide sequence identity (~99%) observed for exons 3 and 4 of these genes (Figure 4). The average amino acid sequence divergence rates for mammalian and chicken class 3 ALDHs were also calculated using the average genetic distances observed for these ALDHs and the dates for the common ancestors of eutherian and marsupial mammals and birds (Table 4). The results showed lower amino acid substitution rates for ALDH3A- and ALDH3B-like genes and proteins prior to the appearance of mammals (0.05% substitution/million years), as compared with the apparent higher amino acid substitution rates (0.06-0.11%) observed during marsupial and eutherian evolution. In addition, given that the chick genome contains both ALDH3A- and ALDH3B-like genes, these results also support an hypothesis that the common ancestor for ALDH3 genes predates the evolutionary appearance of birds during vertebrate evolution, which has been
estimated at 300–320 million years ago (Kumar & Hedges, 1998). Moreover, the clustering and degree of sequence identities for the human, rat, opossum and chicken ALDH3A1-like, ALDH3A2-like and ALDH1A3 genes and protein sequences, together with similarities in gene structures and placements for exons, indicate that these genes have evolved from an ancestral ALDH3 gene earlier in vertebrate evolution.

Functions of Opossum ALDH3 Isozymes

The high levels of sequence identities for the two opossum ALDH3A1-like (ALDH3A1 and ALDH3A3) proteins with those for human and rat ALDH3A1-like sequences, (Figure 1; Table 2), together with the retention of key amino acid residues described earlier (Table 3), indicate that opossum ALDH3A1 and ALDH3A3 contribute to metabolic functions previously reported for eutherian ALDH3A1 enzymes. These include serving as a detoxifying agent for peroxidic aldehydes in the body (Algar & Holmes, 1989) and as a multifunctional protein in the eye, serving as a corneal and lens ‘crystallin’ in several mammalian species (including the opossum) (Holmes et al., 1990), as a protective agent against oxidative damage in the lens and cornea, as a UV-radiation filter in mammalian anterior eye tissues (Abedinia et al. 1990; Vasiloiu et al. 1994; King & Holmes, 1998; Pappa et al. 2003; Pappa et al. 2005; Lassen et al. 2007); in cyclophosphamide and related cancer drug metabolism (Sladek, 2003); as well as participating in the oxidative stress response and cell homeostasis (Moreb et al. 2008).

Opossum ALDH3A2 and ALDH3B1 genes and proteins were also observed in this study, which showed high levels of sequence identity and similarities in key functional residues with the corresponding human and rat genes and encoded proteins (Table 1; Figure 1). Human ALDH3A2 (also called fatty aldehyde
dehydrogenase) plays a major role in the body in the metabolism of long chain fatty aldehydes and several genetic mutations have been reported in human populations which lead to a recessive genetic disease called Sjogren-Larsson syndrome (De Laurenzi et al. 1996; Rogers et al. 1997). Mammalian ALDH3A2 is predominantly located within liver microsomes (Timms and Holmes, 1981), apparently via a hydrophobic C-terminal transmembrane helix (Miyauchi et al. 1991), a property which is also shared with opossum ALDH3A2 (Figure 3). It is proposed that opossum ALDH3A2 plays a similar role in fatty aldehyde metabolism, assisting in the degradation of long chain aldehydes in the body and probably located within liver microsomes of the opossum. The metabolic role(s) for the ALDH3B-like genes and proteins have not been described in mammalian organisms, however it is likely that they also perform detoxification roles for endogenous and exogenous aldehydes.

In summary, BLAT analyses of the recently published opossum genome (Mikkelsen et al., 2007) have been undertaken using the amino acid sequences reported for rat ALDH3A1, ALDH3A2 and ALDH3B1 protein subunits for interrogation of the genome. Evidence is reported for at least four opossum ALDH3-like genes, including two ALDH3A1-like genes closely localized on chromosome 2; an ALDH3A2 gene also on chromosome 2; and an ALDH3B1 gene which has not been mapped on the opossum genome. The predicted amino acid sequences and predicted secondary/tertiary structures for the opossum ALDH3A1-like subunits (designated as ALDH3A1 and ALDH3A3), and the opossum ALDH3A2 and ALDH3B1 subunits showed a high degree of similarity with the corresponding human and rat ALDHs. This report extends previous biochemical and genetic analyses of opossum ALDHs (Holmes et al., 1991; 1992). Phylogenetic analyses undertaken with
opossum, human, rat and chicken ALDH3 isozymes supported the proposed designation of the opossum enzymes as ALDH3A1-like (ALDH3A1 and ALDH3A3), ALDH3A2 and ALDH3B1 isozymes. It is likely that these opossum ALDH3 isozymes perform functions in the opossum, similar to those reported for human, mouse and rat ALDH3A1 in peroxidic aldehyde metabolism and in protecting the eye from UV-induced tissue damage; and for human ALDH3A2, which performs a major role in fatty aldehyde metabolism within liver microsomes.

ACKNOWLEDGEMENTS

I am grateful to Dr Laura Cox and Dr John VandeBerg of the Southwest Foundation for Biomedical Research in San Antonio Texas for helpful discussions and advice.

REFERENCES


Table 1: Aldehyde dehydrogenase genes and enzymes examined

GenBank mRNA (or cDNA) IDs identify previously reported sequences or predicted sequences (see http://www.ncbi.nlm.nih.gov/Genbank/); UNIPROT refer to UniprotKB/Swiss-Prot IDs for individual ALDHs (see http://kr.expasy.org). ALDH sequences are provided by the above sources. Opossum ALDH (ALDH3A1, ALDH3A3, ALDH3A2 and ALDH3B1) protein sequences were obtained from a blast of the opossum genome using web tools of the National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and rat ALDH3A1, ALDH3A2, ALDH3B1 and ALDH3B2 sequences. Predicted exon/intron locations and gene sizes were obtained by BLAT interrogations of the opossum genome using the predicted ALDH sequences and UC Santa Cruz web tools (http://genome.ucsc.edu); *refers to the ALDH3A3 sequence for which exon 1 is apparently located in a gap in the current opossum genome sequence.

<table>
<thead>
<tr>
<th>Animal</th>
<th>ALDH Gene</th>
<th>RefSeq mRNA</th>
<th>No of Amino Acids</th>
<th>Chromosome Location</th>
<th>Strand</th>
<th>Gene Size</th>
<th>Exons</th>
<th>Uniprot ID</th>
<th>Alternate Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>AL3A1</td>
<td>NP_000682.3</td>
<td>453</td>
<td>17: 19,582,228-19,589,034</td>
<td>Negative</td>
<td>6,807</td>
<td>9</td>
<td>P30838</td>
<td>ALDH3A1</td>
</tr>
<tr>
<td></td>
<td>AL3A2</td>
<td>NP_000373.1</td>
<td>485</td>
<td>17: 19,492,877-19,515,861</td>
<td>Positive</td>
<td>22,985</td>
<td>9</td>
<td>P51648</td>
<td>ALDH3A2</td>
</tr>
<tr>
<td></td>
<td>AL3B1</td>
<td>NP_000685.1</td>
<td>468</td>
<td>11: 67,539,344-67,551,956</td>
<td>Positive</td>
<td>12,613</td>
<td>10</td>
<td>P43353</td>
<td>ALDH3B1</td>
</tr>
<tr>
<td>Rat</td>
<td>AL3A1</td>
<td>NP_114178.1</td>
<td>453</td>
<td>10: 47,368,564-47,374,585</td>
<td>Positive</td>
<td>6,022</td>
<td>9</td>
<td>P11883</td>
<td>ALDH3A1</td>
</tr>
<tr>
<td></td>
<td>AL3A2</td>
<td>NP_113119.2</td>
<td>484</td>
<td>10: 47,404,346-47,420,947</td>
<td>Negative</td>
<td>16,602</td>
<td>9</td>
<td>P30839</td>
<td>ALDH3A2</td>
</tr>
<tr>
<td></td>
<td>AL3B1</td>
<td>NP_001006999.1</td>
<td>468</td>
<td>11: 67,492,352-67,515,956</td>
<td>Negative</td>
<td>10,899</td>
<td>9</td>
<td>Q5XI42</td>
<td>ALDH3B1</td>
</tr>
<tr>
<td>Opossum</td>
<td>AL3A1</td>
<td>XP_001372749.1</td>
<td>462</td>
<td>12: 519,102,420-519,137,640</td>
<td>Negative</td>
<td>18,809</td>
<td>9</td>
<td>na</td>
<td>ALDH3A1</td>
</tr>
<tr>
<td></td>
<td>AL3A2</td>
<td>XP_001365258.1</td>
<td>507</td>
<td>11: 519,892,342-519,907,548</td>
<td>Positive</td>
<td>15,207</td>
<td>8</td>
<td>na</td>
<td>ALDH3A3</td>
</tr>
<tr>
<td></td>
<td>AL3B1</td>
<td>XP_001362651.1</td>
<td>465</td>
<td>11: 519,102,420-519,137,640</td>
<td>Negative</td>
<td>35,221</td>
<td>10</td>
<td>na</td>
<td>ALDH3A2</td>
</tr>
<tr>
<td></td>
<td>AL3A2</td>
<td>NP_001006223.1</td>
<td>490</td>
<td>19: 6,791,454-6,794,935</td>
<td>Positive</td>
<td>3,482</td>
<td>9</td>
<td>na</td>
<td>ALDH3A2</td>
</tr>
<tr>
<td></td>
<td>AL3B1</td>
<td>XP_426371.2</td>
<td>496</td>
<td>5: 79,353-85,060</td>
<td>Positive</td>
<td>5,708</td>
<td>10</td>
<td>na</td>
<td>ALDH3B1</td>
</tr>
</tbody>
</table>
Table 2: Percentage identities for human, rat, opossum and chicken aldehyde dehydrogenase (ALDH) amino acid sequences.

Numbers show the percentage of amino acid sequence identities; hu3A1-human ALDH3A1; ra3A1-rat ALDH3A1; op3A1A-opossum ALDH3A1; op3A3-opossum ALDH3A3; hu3A2-human ALDH3A2; ra3A2-rat ALDH3A2; op3A2-opossum ALDH3A2; ch3A2-chick ALDH3A2; hu3B1-human ALDH3B1; hu3B2-human ALDH3B2; ra3B1-rat ALDH3B1; ra3B2-rat ALDH3B2; op3B1-opossum ALDH3B1; ch3B1-chick ALDH3B1. Numbers in **bold** show higher sequence identities for ALDHs from within the same sub-family eg 3A1, 3A2 and 3B.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>hu3A1</td>
<td>100</td>
<td>85</td>
<td>78</td>
<td>79</td>
<td>68</td>
<td>68</td>
<td>71</td>
<td>68</td>
<td>55</td>
<td>51</td>
<td>55</td>
<td>54</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>ra3A1</td>
<td>85</td>
<td>100</td>
<td>73</td>
<td>74</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>67</td>
<td>56</td>
<td>51</td>
<td>55</td>
<td>55</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>op3A1</td>
<td>78</td>
<td>73</td>
<td>100</td>
<td>98</td>
<td>67</td>
<td>66</td>
<td>69</td>
<td>65</td>
<td>54</td>
<td>51</td>
<td>54</td>
<td>53</td>
<td>50</td>
<td>48</td>
</tr>
<tr>
<td>op3A3</td>
<td>79</td>
<td>74</td>
<td>98</td>
<td>100</td>
<td>68</td>
<td>66</td>
<td>70</td>
<td>66</td>
<td>55</td>
<td>52</td>
<td>55</td>
<td>54</td>
<td>51</td>
<td>49</td>
</tr>
<tr>
<td>hu3A2</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>68</td>
<td>100</td>
<td>87</td>
<td>83</td>
<td>76</td>
<td>57</td>
<td>52</td>
<td>58</td>
<td>56</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>ra3A2</td>
<td>68</td>
<td>68</td>
<td>66</td>
<td>66</td>
<td>87</td>
<td>100</td>
<td>80</td>
<td>76</td>
<td>56</td>
<td>51</td>
<td>57</td>
<td>54</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>op3A2</td>
<td>71</td>
<td>69</td>
<td>69</td>
<td>79</td>
<td>83</td>
<td>80</td>
<td>100</td>
<td>77</td>
<td>55</td>
<td>50</td>
<td>54</td>
<td>54</td>
<td>51</td>
<td>54</td>
</tr>
<tr>
<td>ch3A2</td>
<td>68</td>
<td>67</td>
<td>65</td>
<td>66</td>
<td>76</td>
<td>76</td>
<td>77</td>
<td>100</td>
<td>56</td>
<td>51</td>
<td>55</td>
<td>52</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>hu3B1</td>
<td>55</td>
<td>56</td>
<td>54</td>
<td>55</td>
<td>57</td>
<td>56</td>
<td>55</td>
<td>56</td>
<td>100</td>
<td>84</td>
<td>87</td>
<td>81</td>
<td>74</td>
<td>68</td>
</tr>
<tr>
<td>hu3B2</td>
<td>51</td>
<td>51</td>
<td>51</td>
<td>52</td>
<td>52</td>
<td>51</td>
<td>50</td>
<td>51</td>
<td>84</td>
<td>100</td>
<td>79</td>
<td>80</td>
<td>70</td>
<td>63</td>
</tr>
<tr>
<td>ra3B1</td>
<td>55</td>
<td>55</td>
<td>54</td>
<td>55</td>
<td>58</td>
<td>57</td>
<td>54</td>
<td>55</td>
<td>87</td>
<td>79</td>
<td>100</td>
<td>85</td>
<td>75</td>
<td>68</td>
</tr>
<tr>
<td>ra3B2</td>
<td>54</td>
<td>55</td>
<td>53</td>
<td>54</td>
<td>56</td>
<td>54</td>
<td>54</td>
<td>52</td>
<td>81</td>
<td>80</td>
<td>85</td>
<td>100</td>
<td>71</td>
<td>64</td>
</tr>
<tr>
<td>op3B1</td>
<td>50</td>
<td>51</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>50</td>
<td>51</td>
<td>51</td>
<td>74</td>
<td>70</td>
<td>75</td>
<td>71</td>
<td>100</td>
<td>68</td>
</tr>
<tr>
<td>ch3B1</td>
<td>51</td>
<td>49</td>
<td>48</td>
<td>49</td>
<td>56</td>
<td>52</td>
<td>54</td>
<td>51</td>
<td>68</td>
<td>63</td>
<td>68</td>
<td>64</td>
<td>68</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3: Key aldehyde dehydrogenase (ALDH) amino acid residues for mammalian ALDH3 sequences.

Sequences examined include: rat ALDH3A1 (rat 3A1); opossum ALDH3A1 (opossum 3A1); opossum ALDH3A3 (opossum 3A3); rat ALDH3A2 (rat 3A2); opossum ALDH3A2 (opossum 3A2); rat ALDH3B1 (rat 3B1); rat ALDH3B2 (rat 3B2); and opossum ALDH3B1 (opossum 3B1). Identification of predicted key catalytic and structural amino acid residues is based on 3D structural studies for rat ALDH3A1 (Liu et al. 1997). S refers to substrate. See Figure 1 for the complete amino acid sequences for rat and opossum ALDHs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>s binding glycine residues</td>
<td>Glu209</td>
<td>Glu209</td>
<td>Glu155</td>
<td>Glu206</td>
<td>Glu206</td>
<td>Glu209</td>
<td>Glu209</td>
<td>Glu212</td>
</tr>
<tr>
<td>Cys243</td>
<td>Cys243</td>
<td>Cys189</td>
<td>Cys240</td>
<td>Cys240</td>
<td>Cys243</td>
<td>Cys246</td>
<td>Cys243</td>
<td></td>
</tr>
<tr>
<td>catalysis glycine residues</td>
<td>Asn116</td>
<td>Asn116</td>
<td>Asn62</td>
<td>Asn113</td>
<td>Asn113</td>
<td>Asn116</td>
<td>Asn119</td>
<td>Asn116</td>
</tr>
<tr>
<td>Gly131</td>
<td>Gly131</td>
<td>Gly77</td>
<td>Gly128</td>
<td>Gly128</td>
<td>Gly131</td>
<td>Gly134</td>
<td>Gly131</td>
<td></td>
</tr>
<tr>
<td>Gly167</td>
<td>Gly167</td>
<td>Gly113</td>
<td>Gly164</td>
<td>Gly164</td>
<td>Gly167</td>
<td>Gly170</td>
<td>Gly166</td>
<td></td>
</tr>
<tr>
<td>Gly305</td>
<td>Gly305</td>
<td>Gly251</td>
<td>Gly302</td>
<td>Gly302</td>
<td>Gly305</td>
<td>Gly308</td>
<td>Gly301</td>
<td></td>
</tr>
<tr>
<td>Gly403</td>
<td>Gly403</td>
<td>Gly349</td>
<td>Gly400</td>
<td>Gly400</td>
<td>Gly403</td>
<td>Gly406</td>
<td>Gly399</td>
<td></td>
</tr>
<tr>
<td>conserved proline residues</td>
<td>Pro103</td>
<td>Pro103</td>
<td>Pro49</td>
<td>Pro100</td>
<td>Pro100</td>
<td>Pro103</td>
<td>Pro106</td>
<td>Pro103</td>
</tr>
<tr>
<td>Pro317</td>
<td>Pro317</td>
<td>Pro263</td>
<td>Pro314</td>
<td>Pro314</td>
<td>Pro317</td>
<td>Pro320</td>
<td>Pro313</td>
<td></td>
</tr>
<tr>
<td>Pro337</td>
<td>Pro337</td>
<td>Pro284</td>
<td>Pro334</td>
<td>Pro334</td>
<td>Pro337</td>
<td>Pro340</td>
<td>Pro333</td>
<td></td>
</tr>
<tr>
<td>Gly211</td>
<td>Gly211</td>
<td>Gly157</td>
<td>Gly208</td>
<td>Gly208</td>
<td>Gly211</td>
<td>Gly214</td>
<td>Gly207</td>
<td></td>
</tr>
<tr>
<td>Glu333</td>
<td>Glu333</td>
<td>Glu279</td>
<td>Glu330</td>
<td>Glu330</td>
<td>Glu333</td>
<td>Glu336</td>
<td>Glu329</td>
<td></td>
</tr>
<tr>
<td>funnel</td>
<td>Phe335</td>
<td>Phe335</td>
<td>Phe281</td>
<td>Phe332</td>
<td>Phe332</td>
<td>Phe335</td>
<td>Phe338</td>
<td>Phe331</td>
</tr>
</tbody>
</table>
Table 4: Genetic distance and amino acid substitution rate predictions for human, rat, opossum and chick aldehyde dehydrogenases (ALDH)

Common ancestors are identified on Figure 4 for human and rat (hu-ra), human and opossum (hu-op) and human and chicken (hu-ch) ALDH3A1, ALDH3A2 and ALDH3B sequences. Substitution rate is presented as a percentage of amino acid substitutions per million years. MY—million years ago. Dates for common ancestors were obtained from Kumar and Hedges, 1998; Woodburne et al, 2003; and Nilsson et al, 2004.

<table>
<thead>
<tr>
<th>Ancestral Gene</th>
<th>Genetic Distance¹ MY ago</th>
<th>Common Ancestor</th>
<th>% Substitution Rate/MY</th>
<th>Relative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA1: hu-ra AL3A1</td>
<td>0.073±0.005</td>
<td>84-99²</td>
<td>0.07-0.09</td>
<td>1.0⁵</td>
</tr>
<tr>
<td>CA2: hu-ra AL3A2</td>
<td>0.064±0.01</td>
<td>84-99²</td>
<td>0.06-0.07</td>
<td>0.8</td>
</tr>
<tr>
<td>CA3: hu-op AL3A1</td>
<td>0.116±0.004</td>
<td>173-193³</td>
<td>0.07-0.08</td>
<td>0.9</td>
</tr>
<tr>
<td>CA4: hu-op AL3A2</td>
<td>0.089±0.005</td>
<td>173-193³</td>
<td>0.05</td>
<td>0.6</td>
</tr>
<tr>
<td>CA5: hu-ch AL3A2</td>
<td>0.112±0.008</td>
<td>300-320⁴</td>
<td>0.035</td>
<td>0.4</td>
</tr>
<tr>
<td>CA6: hu-ra AL3B1/2</td>
<td>0.089±0.02</td>
<td>84-99²</td>
<td>0.09-0.11</td>
<td>1.25</td>
</tr>
<tr>
<td>CA7: hu-op AL3B1/2</td>
<td>0.133±0.02</td>
<td>173-193³</td>
<td>0.07-0.08</td>
<td>0.9</td>
</tr>
<tr>
<td>CA8: hu-ch AL3B1</td>
<td>0.155±0.02</td>
<td>300-320⁴</td>
<td>0.05</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Figure 1: Amino acid sequence alignments for rat and opossum aldehyde dehydrogenases (ALDH).

See Table 1 for sources of ALDH sequences; 3A1 Rat-rat ALDH3A1; 3A1Opo-opossum ALDH3A1; 3A3 Opo-opossum ALDH3A3; 3A2 Rat-rat ALDH3A2; 3A2 Opo-opossum ALDH3A2; 3B1 Rat-rat ALDH3B1; 3B1 Opo-opossum ALDH3B1; and 3B2 Rat-rat ALDH3B2. * shows identical residues; bold font shows known or predicted exon junctions; predicted β-sheet (grey shading) and α-helix (yellow shading) secondary structures are shown; key residue identification is based on previous 3D studies of rat ALDH3A1 and likely predicted roles for amino acid residues [Liu et al, 1997]: AS-active site residues; N,Asn116; E, Glu209; and C, Cys243 (rat ALDH3A1 numbers); Initiation methionine is shown; α-helices and β-sheets are identified as αA, αB or α10 etc or β0, β1 etc, respectively, as identified by Liu et al, 1997; NP refers to coenzyme binding site identified in bold; note the C-terminal helix predicted for rat and opossum ALDH3A2.
C-terminus helix AL3A2 membrane bound sequences

3B1 Rat
IEFINRERKPLALYAFSRSQVIKQLARTSGGFGCNDGPWHMTLSSLPGGVTSGMGRHKFSFDTDQACLLRSFGMEKINDLRYPFY 446

3B1 Opossum
VAIIFKRERKPLALYAFSNNSVVTQMLECTSGGGFGCNDGPWHMTLSSLPGGVGSGMGMYHGFTEPDQAHRSCQCLRSSGLKVNGLRLYPFY 442

3B2 Rat
NVFINRERKPLALYAFSNNQVTQMLECTSGGGFGCNDGPWFLTLPALPGGVGSGMGMYHGFSDFTFSQACLLRSFGMEKINDLRYPFY 449

3A1 Rat
SPAKMPRH-----------------------------------------------452

3A1 Opossum
PFAKVSWIPGRSKLGST-----------------------------------461

3A3 Opossum
PFAKVSWIPGRSKLGST-----------------------------------407

3A2 Rat
SEKYSWKFLKLQFNGRGLLVLVAVAAVIWQQL-------------------483

3A2 Opossum
SQSKIDWTFIKRPNFGRGLFLMQLFLVGLVATISIKVLKRRKALLVFLVHRIPNSNK Q 506

3B1 Rat
TRNLKVLVAMEKKECTLL-----------------------------------467

3B1 Opossum
AQSLGGLSLVEISAKSQCTLL-----------------------------------464

3B2 Rat
GTWDGDLISWANGQ--SCTLL-----------------------------------468
Figure 2: Predicted transmembrane helices for rat and opossum aldehyde dehydrogenases (ALDH).

See Table 1 for sources of ALDH sequences. The TMHMM web tools of the Center for Biological Sequence Analysis, Technical University of Denmark TMHMM Server plots the probability of the ALDH sequence forming a transmembrane helix (0-1) (shown in red for the relevant amino acid sequences) (http://www.cbs.dtu.dk/services/TMHMM/). The predicted C-terminal transmembrane helices observed for the rat and opossum AL3A2 are identified. Regions of the ALDH predicted to be located inside or outside the membrane are shown in blue and pink, respectively.
Figure 3: Tertiary structure for rat ALDH3A1 and predicted tertiary structures for opossum ALDH3A1, ALDH3A2 and ALDH3B1

The structure for rat ALDH3A1 and the designation of the three domains is taken from Liu et al (1997)
**Figure 4: Nucleotide sequence alignments for opossum ALDH3A1 and ALDH3A3 genes: predicted exons 3 and 4 and intron 3**

Exon 3 and 4 sequences are in red and intron 3 sequences (incomplete sequence shown) are in black; *identical sequence; 3A1-opossum ALDH3A1 sequence; 3A3-opossum ALDH3A3 sequence. The numbers refer to nucleotides from the commencement of exon 3 in each case.

**Exon 3 99% identical**

3A1  GGCAAFCTGCTAGTCATCAAGCCCTCTCGAGTTGAGCGAGCAGACAGCCATCATGTTGGCCACACTGATCCCTAAGTACCTGGACAAGGAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2005
3A3  GACTTGTATCCAGTTATCAATGGTGGTATTCCTGAGACCACAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2124

---

**Intron 3 90% identical**

3A1  GTGAGTCACAGCCCCCAAGGAGTTCTCGTGGAGGGAAAGAATGAGCAGAGCTGGAGGTGGGCAGGCCAAGGAGCTCCTTCTGGGCTGAGGGGCCC 418
3A3  AGCAGCAGCCAAGCAGCTGCTGCTCTTTGGGGTTCTGGTGCCCCTCCTAGGGGATCTTAGGAAAGTCCCCTAACTCCTTCTGGGCTCCAGTTTGCTGATGAC 2005

---

**Exon 4 >99% identical**

3A1  GACTTGTATCCAGTTATCAATGGTGGTATTCCTGAGACCACAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2005
3A3  GACTTGTATCCAGTTATCAATGGTGGTATTCCTGAGACCACAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2030

---

3A1  AGCAGCAGCCAAGCAGCTGCTGCTCTTTGGGGTTCTGGTGCCCCTCCTAGGGGATCTTAGGAAAGTCCCCTAACTCCTTCTGGGCTCCAGTTTGCTGATGAC 2005
3A3  AGCAGCAGCCAAGCAGCTGCTGCTCTTTGGGGTTCTGGTGCCCCTCCTAGGGGATCTTAGGAAAGTCCCCTAACTCCTTCTGGGCTCCAGTTTGCTGATGAC 2030

---
Figure 5: Phylogram of human, rat, opossum and chick class 3 aldehyde dehydrogenase (ALDH) sequences.

Each branch of the phylogram is labeled with an abbreviation for the gene name: 3A1-ALDH3A1; 3A2-ALDH3A2; 3A3-ALDH3A3; 3B1-ALDH3B1; and 3B2-ALDH3B2, followed by the species name. Note the clustering into three ALDH groups (ALDH3A1; ALDH3A2; and ALDH3B). Common ancestors for ALDH genes are identified: CA1-human and rat ALDH3A1; CA2-human and rat ALDH3A2; CA3-Human and opossum ALDH3A1; CA4-human and opossum ALDH3A2; CA5-human and chick ALDH3A2; CA6-human and rat ALDH3B1/2; CA7-human and opossum ALDH3B1; CA8-human and chicken ALDH3B1.