Published in Biochemical Genetics 48: 287-303 (2010) Prepublication PDF Version

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.,

2006). 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; Vasilidou 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>] (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>) (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>) (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 (Miyachi 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 α D 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>]). 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; Vasiliou 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

- Abedinia M, Pain T, Algar EM, Holmes RS (1990) Bovine corneal aldehyde dehydrogenase: the major soluble protein with a possible dual protective role for the eye. *Exp Eye Res* 51:419.
- Algar EM, Holmes RS (1989) Purification and properties of mouse stomach aldehyde dehydrogenase. Evidence for a role in the oxidation of peroxidic and aromatic aldehydes. *Biochim Biophys Acta* 995:168.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497.
- De Laurenzi V, Rogers GR, Hamrock DJ, Marekov LN, Steinhart PM, Compton JN, Markova N, Rizzo WB (1996) Sjogren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nature Genet* 12:52.
- Fitch DH, Bailey WJ, Tagle DA, Goodman M, Saieue L, Slightom JL (1991) Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an earlier ancestor of simian primates. *Proc Natl Acad Sci USA* 88:7396.
- Hempel J, Harper K, Lindahl R (1989) Inducible class 3 aldehyde dehydrogenase from rat carcinoma and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin treated liver: distant relationship from mammalian liver/mitochondrial. *Biochem* 28:1160.
- Hempel J, Liu Z-J, Perozich J, Rose J, Lindahl R, Wang B-C (1996) Conserved residues in the aldehyde dehydrogenase family. Locations in the class 3 tertiary structure. *Adv Exp Med Biol* 414:9.

Holmes RS, Popp RA, VandeBerg JL (1988) Genetics of ocular NAD-dependent alcohol dehydrogenase and aldehyde dehydrogenase in the mouse: evidence for genetic identity with stomach isozymes and localization of *Ahd-4* on chromosome 11 near Trembler Biochem Genet 26:191.

Holmes RS, van Oorschot RAH, VandeBerg JL (1990). Genetics of alcohol dehydrogenase and aldehyde dehydrogenase from *Monodelphis domestica* cornea: further evidence for identity of corneal aldehyde dehydrogenase with a major soluble protein. Genet Res 56:259.

Holmes RS, van Oorschot RAH, Vandeberg JL (1991) Aldehyde dehydrogenase isozymes in the gray short-tailed opossum (*Monodelphis domestica*): tissue and subcellular distribution and biochemical genetics of ALDH3. Biochem Genet 29:163.

Holmes RS (2009) Opossum aldehyde dehydrogenases. Evidence for four *ALDH1A1*-like genes on chromosome 6 and *ALDH1A2* and *ALDH1A3* genes on chromosome 1. Biochem Genet 47:609.

Hsu LC, Chang WC, Hiraoka L, Hsieh CL (1994a) Molecular cloning, genomic organization and chromosomal localization of an additional human aldehyde dehydrogenase, ALDH6. Genomics 24:333.

Hsu L, Chang W-C, Yoshida A (1994b) Sequencing and expression of the human ALDH8 encoding a new member of the aldehyde dehydrogenase family. Gene 151:285.

Hsu L, Chang W-C, Yoshida A (1997) Human aldehyde dehydrogenase genes, *ALDH7* and *ALDH8*: Genomic organization and gene structure comparison. Gene 189:89.

Jones DE, Brennan MD, Hempel J, Lindahl R (1988) Cloning and complete nucleotide sequence of a full-length cDNA encoding a catalytically functional tumor-associated aldehyde dehydrogenase. Proc. Natl. Acad. Sci. USA 85:1782.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12:994.

King G, Holmes RS (1998) Human ocular aldehyde dehydrogenase isozymes: distribution and properties as major soluble proteins in cornea and lens. J Exp Zool 282:12.

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917.

Lassen N, Bateman JB, Estev T, Kuszak JR, Nees DW, Piatigorsky J, Deuster G, Day BJ, Huang J, Hines LM, Vasiliou V (2007) Multiple and additive functions of ALDH3A1 and ALDH1A1: cataract phenotype and ocular oxidative damage in *aldh3a1(-/-) /aldh1a1(-/-)* knockout mice. J Biol Chem 282:25668.

Liu Z-J, Sun YJ, Rose JP, Chung Y-J, Hsiao CD, Chang W-R, Kuo I, Perozich J, Lindahl R, Hempel J, Wang B-C (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. Nature Struct Biol 4:317.

Mammalian Gene Collection Team (2004) The status, quality and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection. Genome Res 14:2121.

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koin E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL,

Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Marshall Graves JA, Ponting CP, Breen M, Samollow PB, ELander ES, Lindblad-Toh J (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167.

Miyauchi K, Masaki R, Taketani S, Yamamoto A, Akayama M, Tashiro Y (1991) Molecular cloning, sequencing and expression of cDNA for rat liver microsomal aldehyde dehydrogenase. *J Biol Chem* 266:19536.

Moreb JS, Baker HV, Chang L-J, Amaya M, Lopez MC, Ostmark B, Chou W (2008) ALDH isozymes downregulation affects cell growth, cell motility and gene expression in lung cancer cells. *Mol Cancer* 7:87.

Nilsson MA, Arnason U, Spencer PB, Janke A (2004) Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene* 340:1189.

Pappa A, Estey T, Manzer R, Brown D, Vasiliou V (2003) Human aldehyde dehydrogenase 3A1 (ALDH3A1): biochemical characterization and immunohistochemical localization in the cornea. *Biochem J* 376:615.

Pappa A, Brown D, Koutalos Y, DeGregory J, White C, Vasilou, V (2005) Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of human corneal epithelial cells. *J Biol Chem* 280:27998.

Rainwater DL, Kammerer CM, Singh ATK, Moore PM, Pousesh M, Shelledy WR, Vandeberg JL, Robinson ES (2001) Genetic control of lipoprotein phenotypes in the laboratory opossum. *Genescreen* 1:117.

Rogers GR, Marlova NG, De Laurenzi V, Rizzo WB, Compton JG (1997) Genomic organization and expression of the fatty aldehyde dehydrogenase gene (*FALDH*). *Genomics* 39:127.

Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modelling server. *Nucleic Acids Res* 31:3383.

Sladek NE (2003) Human aldehyde dehydrogenases: potential pathological, pharmacological and toxicological impact. *J Biochem Mol Toxicol* 17:7.

Sophos NA, Black WJ, Vasiliou V (2006) An update of the ALDH gene family. In: *Enzymology and Molecular Biology of Carbonyl Metabolism*. Eds: H Weiner, B Plapp, R Lindahl E Maser Purdue University Press, Indiana 12:3.

Timms GP, Holmes RS (1982) Genetics and ontogeny of aldehyde dehydrogenase isozymes in the mouse; evidence for a locus controlling the inducibility of the liver microsomal isozymes. *Biochem Genet* 19:1223.

VandeBerg JL, Williams-Blangero S, Hubbard GB, Robinson ES (1994) Susceptibility to ultra-violet induced corneal sarcomas is highly heritable in a laboratory opossum model. *Int. J. Cancer* 56:119.

Vasiliou V, Reuter SF, Williams S, Puga A, Nebert DW (1999) Mouse cytosolic class 3 aldehyde dehydrogenase (Aldh1a1): gene structure and regulation of constitutive and dioxine-inducible expression. *Pharmogenet* 9:569.

Vasiliou V, Nebert DW (2005) Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. *Human Genomics* 2:138.

Woodburne MO, Rich TH, Springer MS (2003) The evolution of tribospheny and the antiquity of mammalian clades. *Mol Phylogenet Evol* 28:360

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.

Animal	ALDH Gene	RefSeq mRNA XP: Prediction	No of Amino Acids	Chromosome Location	Strand	Gene Size	Exons	Uniprot ID	Alternate Gene Name
Human	AL3A1	NP_000682.3	453	17: 19,582,228-19,589,034	Negative	6,807	9	P30838	ALDH3A1
	AL3A2	NP_000373.1	485	17: 19,492,877-19,515,861	Positive	22,985	9	P51648	ALDH3A2
	AL3B1	NP_000685.1	468	11: 67,539,344-67,551,956	Positive	12,613	10	P43353	ALDH3B1
	AL3B2	NP_000686.2	385	11: 67,187,265-67,190,741	Negative	3,477	7	P48448	ALDH3B2
Rat	AL3A1	NP_114178.1	453	10: 47,368,564-47,374,585	Positive	6,022	9	P11883	ALDH3A1
	AL3A2	NP_113119.2	484	10: 47,404,346-47,420,947	Negative	16,602	9	P30839	ALDH3A2
	AL3B1	NP_001006999.1	468	1: 206,438,748-206,449,646	Negative	10,899	9	Q5XI42	ALDH3B1
	AL3B2	XP_001068348.1	469	1: 206,549,738-206,553,421	Positive	3,684	9	na	ALDH3B2
Opossum	AL3A1	XP_001372749.1	462	2: 518,942,352-518,961,160	Negative	18,809	9	na	ALDH3A1
	AL3A3	XP_001372729.1*	408*	2: 518,892,342-518,907,548*	Positive	15,207*	8*	na	ALDH3A3
	AL3A2	XP_001365258.1	507	2: 519,102,420-519,137,640	Negative	35,221	10	na	ALDH3A2
	AL3B1	XP_001362651.1	465	Un: 78,602,810-78,619,018	Negative	16,209	8	na	ALDH3B1
Chicken	AL3A2	NP_001006223.1	490	19: 6,791,454-6,794,935	Positive	3,482	9	na	ALDH3A2
	AL3B1	XP_426371.2	496	5: 79,353-85,060	Positive	5,708	10	na	ALDH3B1

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.

ALDH Gene	hu3A1	ra3A1	op3A1	op3A3	hu3A2	ra3A2	op3A2	ch3A2	hu3B1	hu3B2	ra3B1	ra3B2	op3B1	ch3B1
hu3A1	100	85	78	79	68	68	71	68	55	51	55	54	50	51
ra3A1	85	100	73	74	67	68	69	67	56	51	55	55	51	49
op3A1	78	73	100	98	67	66	69	65	54	51	54	53	50	48
op3A3	79	74	98	100	68	66	70	66	55	52	55	54	51	49
hu3A2	68	67	67	68	100	87	83	76	57	52	58	56	52	56
ra3A2	68	68	66	66	87	100	80	76	56	51	57	54	50	52
op3A2	71	69	69	79	83	80	100	77	55	50	54	54	51	54
ch3A2	68	67	65	66	76	76	77	100	56	51	55	52	51	51
hu3B1	55	56	54	55	57	56	55	56	100	84	87	81	74	68
hu3B2	51	51	51	52	52	51	50	51	84	100	79	80	70	63
ra3B1	55	55	54	55	58	57	54	55	87	79	100	85	75	68
ra3B2	54	55	53	54	56	54	54	52	81	80	85	100	71	64
op3B1	50	51	50	51	52	50	51	51	74	70	75	71	100	68
ch3B1	51	49	48	49	56	52	54	51	68	63	68	64	68	100

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.

Predicted Function	rat 3A1	opossum 3A1	opossum 3A3	rat 3A2	opossum 3A2	rat 3B1	rat 3B2	opossum 3B1
s binding	Glu209	Glu209	Glu155	Glu206	Glu206	Glu209	Glu212	Glu209
	Cys243	Cys243	Cys189	Cys240	Cys240	Cys243	Cys246	Cys243
catalysis	Asn116	Asn116	Asn62	Asn113	Asn113	Asn116	Asn119	Asn116
glycine residues chain folding roles	Gly105	Gly105	Gly51	Gly102	Gly102	Gly105	Gly108	Gly105
	Gly131	Gly131	Gly77	Gly128	Gly128	Gly131	Gly134	Gly131
	Gly167	Gly167	Gly113	Gly164	Gly164	Gly167	Gly170	Gly166
	Gly187	Gly187	Gly134	Gly184	Gly184	Gly187	Gly190	Gly186
	Gly240	Gly240	Gly186	Gly237	Gly237	Gly240	Gly243	Gly236
	Gly305	Gly305	Gly251	Gly302	Gly302	Gly305	Gly308	Gly301
	Gly336	Gly336	Gly282	Gly333	Gly333	Gly336	Gly339	Gly332
	Gly383	Gly383	Gly329	Gly380	Gly380	Gly383	Gly336	Gly379
conserved proline residues	Gly403	Gly403	Gly349	Gly400	Gly400	Gly403	Gly406	Gly399
	Gly410	Gly410	Gly355	Gly407	Gly407	Gly410	Gly413	Gly406
	Pro103	Pro103	Pro49	Pro100	Pro100	Pro103	Pro106	Pro103
stem of active site funnel	Pro317	Pro317	Pro263	Pro314	Pro314	Pro317	Pro320	Pro313
	Pro337	Pro337	Pro284	Pro334	Pro334	Pro337	Pro340	Pro333
	Gly187	Gly187	Gly134	Gly184	Gly184	Gly187	Gly190	Gly186
	Gly211	Gly211	Gly157	Gly208	Gly208	Gly211	Gly214	Gly207
	Glu333	Glu333	Glu279	Glu330	Glu330	Glu333	Glu336	Glu329
	Phe335	Phe335	Phe281	Phe332	Phe332	Phe335	Phe338	Phe331

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-millions years ago. Dates for common ancestors were obtained from Kumar and Hedges, 1998; Woodburne et al, 2003; and Nilsson et al, 2004.

<i>AL3A</i>	Genetic	Common	%	Relative %
Common	Distance ¹	Ancestor	Substitution	Substitution
Ancestral Gene		MY ago	Rate/MY	Rate
CA1: hu-ra <i>AL3A1</i>	0.073±0.005	84-99 ²	0.07-0.09	1.0 ⁵
CA2: hu-ra <i>AL3A2</i>	0.064±0.01	84-99 ²	0.06-0.07	0.8
CA3: hu-op <i>AL3A1</i>	0.116±0.004	173-193 ³	0.07-0.08	0.9
CA4: hu-op <i>AL3A2</i>	0.089±0.005	173-193 ³	0.05	0.6
CA5: hu-ch <i>AL3A2</i>	0.112±0.008	300-320 ⁴	0.035	0.4
CA6: hu-ra <i>AL3B1/2</i>	0.089±0.02	84-99 ²	0.09-0.11	1.25
CA7: hu-op <i>AL3B1/2</i>	0.133±0.02	173-193 ³	0.07-0.08	0.9
CA8: hu-ch <i>AL3B1</i>	0.155±0.02	300-320 ⁴	0.05	0.6

Figure 1: Amino acid sequence alignments for rat and opossum aldehyde dehydrogenases (ALDH).

[illegible]

3B1 Rat	IEFINRREKPLALYAFSKR	SOVIKQVLAR	TSSGGFCGNDGFMHMTLSSLPFGGV	GTSGMGRYHG	KFSFDTF	S	NQRACLLRSPGMEKINDLRYPPY	446
3B1 Opossum	VAFINRKREKPLALYAFSNN	SEVVRQMLER	TSSGGFCGNDGFMHMTLTTL	PFGGV	GN	SGMGMYHG	KFTFDAFS	442
3B2 Rat	NVFINQREKPLALYAFSNN	GQVVTQMLECT	S	SGSFGGNDGFLYLTPALPLGGV	GN	SGMGRYHG	KFSFDTF	449
	*** * *	***** *	****	**		** * *	***** * *	

C-terminus helix AL3A2 membrane bound sequences

3A1 Rat	SPAKMPRH-----	452					
3A1 Opossum	FPAKVS	WIPGRSKLGGT-----	461				
3A3 Opossum	FPAKVS	WIPGRSKLGGT-----	407				
3A2 Rat	SESKVSWSKFLLKQFNKGRL	QLLLLVCLVAVAAV	IVKDQL-----	483			
3A2 Opossum	SQSKIDWTNFFILKR	FNKGRI	IGLFLMAFLGVLVAIT	IKKYQAVLR	RKALLVFLVIHRLH	WSNKQ	506
3B1 Rat	TSRNL	RVLLVAMEKR	CCSCTLL-----	467			
3B1 Opossum	AQ	RLGLLLSLVEIS	RKSQCTLL-----	464			
3B2 Rat	GTW	DQQLISWAMGSQ--	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  GGCAATGCTGTAGTCATCAAGCCTTCTGAGTTGAGCGAGCAGACAGCCATCATGTTGGCCACACTGATCCCTAAGTACCTGGACAAG 87
3A3  GGCAACGCTGTAGTCATCAAGCCTTCTGAGTTGAGCGAGCAGACAGCCATCATGTTGGCCACACTGATCCCTAAGTACCTGGACAAG 87
      *****
      Intron 3 90% identical
3A1  GTGAGTCACAGCCCCAAAGGATCTCTGTAGGGGAAAGAATGAGCAGAGCTGGAGGTGGGCAGGCCAAGGAGCTCCTTCTGG----- 168
3A3  GTGAGTCACAGCCCCAAAGGATCTCTGTAGGGGAAAGAATGAGCAGAGTGGAGGTGGGCAGGCCAAGGAGCTCCTTCTGGGCTGAGGGGCC 180
      *****

3A1  -----TGCTCTTTGGGGTTCTGGTGCCCTCCTAGGGAGGAGATCTCTCTAAAGGGATCTTATCTAAACTTCATGCCCTCCAGATCAATGATGTTTATTTCTGGGGCTCAGTGGTCATT 282
3A3  CAGGACTGCTCTTTGGGGTTCTGGTGCCCTCCTAGGGAGGAGATCTCTCTAAAGGGATCTTATCTAAACTTCATGCCCTCCAGATCAATGATGTTTATTTCTGGGGCTCAGTGGTCATT 300
      *****

3A1  TGCTGATTGGCAGATCCCCCGCTTTGGGATCTTAGGAAAGTCCCCTAACTCCTCTGGGCTCC-----GCTAGTTGTTGTTGTTGAGTCATTTCAGGGTGGCTGGCTCTTTGGGG 392
3A3  TGCTGATTGGCAGATCCCC--CTTTGGGATCTTAGGAAAGTCCCCTAACTCCTCTGGGCTCCAGTTTGTGTGCTAGTTGTTGTTGAGTCATTTCAGGGTGGCCGGCTCTTTGGGG 418
      *****

      Exon 4 >99% identical
3A1  GACTTGATCCAGTTATCAATGGTGGTATTCCTGAGACCACAGAAGTCTGAAGGAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2005
3A3  GACTTGATCCAGTTATCAATGGTGGTATTCCTGAGACCACAGAAGTCTGAAGGAGAAATTTGACCACATCATCTACACTGGGAGCACTGGAGTGGGCAAGATTGTCATGAC 2030
      *****

3A1  AGCAGCAGCCAAGCACCTGACTCCAGTGACCCTGGAAGTGGGAGGGAAGAGTCCCTGTTATATCGACAAGGACTGTGACCTGGATATTGCCTGC 2099
3A3  AGCTGCAGCCAAGCACCTGACTCCAGTGACCCTGGAAGTGGGAGGGAAGAGTCCCTGTTATATCGACAAGGACTGTGACCTGGATATTGCCTGC 2124
      ***

```


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.