Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family

Vasilis Vasiliou* and Daniel W. Nebert

1 Molecular Toxicology and Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Health Sciences Center, Denver, CO, USA
2 Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH, USA
* Correspondence to: Tel: +1 303 315 6153; Fax: +1 303 315 6281; E-mail: vasilis.vasiliou@uchsc.edu

Date received: 23rd February 2005

Abstract
The aldehyde dehydrogenase (ALDH) gene superfamily encodes enzymes that are critical for certain life processes and detoxification via the NAD(P)⁺-dependent oxidation of numerous endogenous and exogenous aldehyde substrates, including pharmaceuticals and environmental pollutants. Analysis of the ALDH gene superfamily in the latest databases showed that the human genome contains 19 putatively functional genes and three pseudogenes. A number of ALDH genes are upregulated as a part of the oxidative stress response and inexplicably overexpressed in various tumours, leading to problems during cancer chemotherapy. Mutations in ALDH genes cause inborn errors of metabolism —such as the Sjögren–Larsson syndrome, type II hyperprolinaemia and γ-hydroxybutyric aciduria— and are likely to contribute to several complex diseases, including cancer and Alzheimer’s disease. The ALDH gene products appear to be multifunctional proteins, possessing both catalytic and non-catalytic properties.

Keywords: human genome, aldehyde dehydrogenase gene family, genetic polymorphism, evolution, crystallins

Introduction
Aldehyde dehydrogenases (ALDHs; EC1.2.1.3) represent a group of enzymes that oxidise a wide range of endogenous and exogenous aldehydes to their corresponding carboxylic acids. Endogenous aldehydes are formed during the metabolism of amino acids, carbohydrates, lipids, biogenic amines, vitamins and steroids. Biotransformations of a large number of drugs and environmental chemicals also generate aldehydes. Aldehydes are highly reactive electrophilic compounds which interact with thiol and amino groups, the resulting effects vary from physiological and therapeutic to cytotoxic, mutagenic or carcinogenic. In this respect, ALDHs efficiently oxidise and, in most instances, detoxify a significant number of chemically diverse aldehydes which otherwise would be harmful to the organism. Strong evidence supporting this notion comes from the fact that mutations in ALDH genes cause inborn errors of metabolism associated with clinical phenotypes—such as Sjögren–Larsson syndrome (SLS), type II hyperprolinaemia and γ-hydroxybutyric aciduria. In addition, mutations in ALDH genes contribute to clinically relevant diseases such as cancer and Alzheimer’s disease.

There are instances, however, in which ALDHs catalyse reactions yielding chemically reactive or bioactive metabolites that are essential to the organism. Several ALDH enzymes—including ALDH1A1, ALDH1A2 and ALDH1A3—catalyse the irreversible oxidation of retinal to retinoic acid. Whereas the light-absorbing properties of retinal are an essential element for vision, the carboxylic acid isomers, all-trans-retinoic acid and/or 9-cis-retinoic acid, serve as ligands for the retinoic receptor (RAR) and the retinoid X receptor (RXR) that mediate gene expression for growth and development.

The importance of ALDH enzymes in retinoic acid formation became evident from the fact that homozygous disruption of the mouse Aldh1a2 gene results in an embryonic lethal phenotype due to defects in early heart morphogenesis, whereas Aldh1a3 null mice die shortly after birth, due to respiratory distress caused by choanal atresia.

Formation of retinoic acid and γ-aminobutyric acid (GABA) are among the most intriguing functions of ALDHs regarding bioactivation. GABA is implicated in the regulation of the GABAergic, dopaminergic and opioid systems. Even though the main pathway for GABA synthesis is the decarboxylation of L-glutamate, this neurotransmitter can also be
formed from putrescine by direct oxidative deamination to give γ-aminobutyraldehyde, which is then converted into GABA by an ALDH. All in all, the ALDH gene family represents a truly diverse group of proteins which are critical to metabolism.

Multiple function(s) of the ALDH enzymes

Although the major function of ALDH enzymes is the NAD(P)⁺-dependent aldehyde oxidation, it has become increasingly clear that some, if not most, ALDHs exhibit multiple functions (Figure 1). For example, ALDH1A1, ALDH2, ALDH3A1 and ALDH4A1 are known to catalyse the formation of 1,2-glyceryl dinitrate and nitrite from nitroglycerin within mitochondria, leading to the production of cGMP and vasorelaxation. Indeed, it has recently been suggested that ALDH2 also possesses nitrate reductase activity, which catalyses the formation of 1,2-glyceryl dinitrate and nitrite from nitroglycerin within mitochondria, leading to the production of cGMP and vasorelaxation.

Aside from their catalytic properties, ALDH proteins are capable of non-catalytic interactions with chemically diverse endogenous compounds and chemotherapeutic agents. In this context, ALDH1A1 has been identified as an androgen-binding protein prominently expressed in human genital fibroblasts; as a cholesterol-binding protein in bovine lens epithelium; and as a cytosolic thyroid hormone-binding protein in Xenopus. ALDH1A1 has also been identified as a flavopyridol-binding protein in non-small cell lung carcinomas and as a daunorubicin binding protein in rat liver. Similar to ALDH1A1, ALDH2 also displays binding capabilities with exogenous compounds, which became evident from its identification as an acetaminophen binding protein.

In addition, it has been suggested that some ALDHs may play a critical role in cellular homeostasis by maintaining redox balance. For example, it has been proposed that ALDH3A1 may scavenge hydroxyl radicals via the −SH groups of Cys and Met residues, and that both ALDH3A1 and ALDH1A1 may contribute to the antioxidant capacity of the cell by generating NADPH and/or NADH. The enzymatic activity of ALDH3A1 generates NADPH, which is linked to the regeneration of reduced glutathione (GSH) from its oxidised form (GSSG) via the glutathione reductase/peroxidase system. NAD(P)H may also function as a direct antioxidant by reducing glutathyl radicals (GS• or GSH•) or tyrosyl radicals.

The expression of ALDH3A1 and ALDH1A1 at very high concentrations in the mammalian cornea and lens (crystallins) has led to additional hypotheses regarding the multifunctional properties of these proteins—including a structural function contributing to transparency. Finally, the ALDH1A1 gene product is similar to the green garden pea ‘26g protein’ involved in the regulation of turgor pressure, suggesting that the ALDH1A1 protein might have osmoregulatory properties.

Evolution of the ALDH genes

ALDHs have a wide distribution in nature, ranging from bacteria and yeasts to plants and animals. Sequence comparisons indicate extensive similarity between bacterial and human ALDHs and suggest that the superfamily has a common ancestral gene, dating back to ~3 billion years ago. A systematic nomenclature scheme for the ALDH gene superfamily (in animals, plants, bacteria and yeasts) has been developed, based on evolutionary divergence, which has been implemented with biannual updates and is available via the internet (http://www.aldh.org).

ALDH proteins are conveniently classified into families and subfamilies based on the percentage of amino acid identity. Proteins sharing ≥40 per cent identity are assigned to a particular family designated by an Arabic numeral, whereas those sharing ≥60 per cent identity are classified in the same subfamily designated by a letter. These cut-off values follow the original recommendations by Margaret Dayhoff and were first applied to the cytochrome P-450 superfamily. At present, more than 130 additional gene superfamilies and large gene families follow this same format.

Endogenous functions of ALDHs

Antioxidants and oxidative stress increase the expression of certain ALDH genes, leading to increased protection of the cell against insult by environmental chemicals and drugs. Increased expression of certain ALDHs in tumour cells, however, leads to decreased cellular sensitivity to cyclophosphamide and other oxazaphosphorines and, thus, to clinical problems in the treatment of cancer patients. The reason for certain ALDHs—and other non-P450 members of the [4h]
Latest genes in the ALDH database

A search of the Human Gene Nomenclature Committee (HGNC) database using ‘ALDH’ produced 20 ‘hits’: 19 ALDH genes plus AGPS (encoding alkylglycerone phosphate synthase). This latter entry appears in the database because one of its aliases is ‘ALDHPSY’. A search of the HGNC database using ‘aldehyde dehydrogenase’ produced 21 hits — the 19 putatively functional ALDH genes plus two others, AASDH-PPT and ADH5 (Table 1). After various analyses, it was concluded that these latter two, evolutionarily, do not belong to the ALDH gene superfamily. AASDH-PPT was found to belong to the 4′-phosphopantetheinyl transferase superfamily (pfam01648; ACP5) and ADH5 belongs to the alcohol dehydrogenase family (pfam00107; ADH_zinc_N). These two genes appear in the HGNC database in response to the cue ‘aldehyde dehydrogenase’, because these two words are included within their names: aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase and form-aldehyde dehydrogenase. Interestingly, the Enzyme Commission (EC) database gives L-aminoadipate-semialdehyde dehydrogenase the number EC 1.2.1.31 — meaning that it is closely related functionally to the other ALDH activities (EC 1.2.1.3).

The two most recently discovered ALDH genes are ALDH1L2 and ALDH16A1. The ALDH1L2 protein is very similar to ALDH1L1, which is better known as 10-formyl-tetrahydrofolate dehydrogenase (TFDH), a bifunctional enzyme formed from the fusion of two unrelated genes; TFDH is highly expressed in human liver, kidney and pancreas. The deduced amino acid sequence of ALDH1L1 contains three domains — including the amino terminal (residues 1–203), which is approximately 30 per cent identical to phosphoribosylglycinamid formyltransferase (EC 2.1.2.2), and the carbonyl terminal (residues 417–902), which belongs to the ALDH superfamily. The intermediate domain (residues 204–416) does not appear to have any known catalytic function, although it shows significant homology with the structural domain of a calmodulin-like protein. This intermediate domain is apparently an essential structural element that aligns the two functional domains together for 10-formyl-tetrahydrofolate (10-FTHF) dehydrogenase activity — the primary function of this enzyme. This multidomain enzyme catalyses: (a) the NADP+-dependent oxidation of 10-FTHF to tetrahydrofolate (THF); (b) the NADP+-dependent oxidation of 2-propanol and acetaldehyde and; (c) the NADP+-independent hydrolysis of 10-FTHF to formate and THF. ALDH1L1 is involved in formate metabolism, as well as the regulation of 10-FTHF and THF, which are principal sources of folate in the cell.

The ALDH1L2 gene encodes a protein that is 72.3 per cent identical with the ALDH1L1 and is also a fusion gene, comprising three domains: (a) the formyl-trans-N-formyl transferase (pfam00551) at the amino terminal (residues 23–202); (b) the formyltransferase carboxyl terminal domain (pfam02911) in the middle (residues 226–327); and (c) the aldehyde dehydrogenase domain at the carbonyl terminal (residues 451–910). No functional data have yet been reported for the ALDH1L2 protein. It is worth mentioning that the BLAST scores for the ALDH domain in both the ALDH1L1 and ALDH1L2 genes are much higher than those for the other two domains, which is strong evidence to support the notion that these two genes should be listed as ALDH genes.

The ALDH16A1 gene, located at 19q13.33, was identified recently by the National Institutes of Health Mammalian Gene Collection (MGC) Program, which represents a multi-institutional effort to identify and sequence a cDNA clone containing a complete open reading frame for each human and mouse gene. The ALDH16A1 gene encodes a protein of 802 amino acids (listed in databases as ‘hypothetical protein MGC10204’), which is ~35 per cent identical to putatively membrane-anchored ALDHs found in bacterial species such as Sinorhizobium meliloti 102. Putative orthologues
of ALDH16A1 are found in mouse, rat and chimpanzee, and exhibit around 72–74 per cent amino acid identity with the human ALDH16A1.

Finally, there are three pseudogenes—ALDH7A1P1, ALDH7A1P2 and ALDH7A1P3—located at Chr 5q14, 7q36 and 10q21, respectively, of which only ALDH7A1P1 meets the HGNC criteria for a pseudogene (at least 50 per cent amino acid identity across 50 per cent of the open reading frame); however, the names ALDH7A1P2 and ALDH7A1P3 are proposed because sequence homology with the ALDH7A1 gene is significantly higher than that with any other ALDH gene. An alternative nomenclature system for naming four various types of pseudogenes has recently been proposed. As is commonly seen with pseudogenes in the mammalian genome, all three of these pseudogenes are found at chromosomal locations that differ from that of the ALDH7A1 functional gene from which the pseudogenes clearly originated.
Conclusions

The authors’ current analysis of the ALDH genes within the human genome is now probably complete, and it can be concluded that the human ALDH gene superfamily comprises 19 genes in 11 families and four subfamilies (Figure 2). The ALDH1 family contains six functional genes: the cytosolic ALDH1A1 and the mitochondrial ALDH1B1 may be involved in acetaldehyde metabolism; ALDH1A1 also participates in retinal oxidation and the detoxification of cyclophosphamide; the ALDH1A2 and ALDH1A3 proteins are integral to the oxidation of retinal to retinoic acid; the ALDH1L1 gene codes for 10-FTHF dehydrogenase; the ALDH1L2 gene product is very similar to that of ALDH1L1, but no functional data are available yet. The ALDH2 family has a single member, encoding the mitochondrial ALDH that exhibits the highest affinity for acetaldehyde and is critical in ethanol metabolism. Although ALDH2 officially qualifies as a seventh member of the ALDH1 family, its longstanding name of “ALDH2” associated with ethanol metabolism has been grandfathered into the more recent nomenclature system based on evolutionary divergence. The ALDH3A subfamily contains the dioxin-inducible ALDH3A1 and ALDH3A2, which are primarily involved in the oxidation of medium- and long-chain aliphatic and aromatic aldehydes. The ALDH3B subfamily consists of two structurally related genes, ALDH3B1 and ALDH3B2; as yet, there are no functional data for either gene product. ALDH5A1 encodes the succinic semialdehyde dehydrogenase. ALDH6A1 encodes the acetyl CoA-dependent methylmalonate semialdehyde dehydrogenase. The ALDH7A1 gene product, also known as ‘antiquitin’, is similar to the green garden pea 26-g protein involved in the regulation of turgor pressure. ALDH8A1 appears to metabolise retinal. ALDH9A1 codes for an enzyme that participates in the metabolism of γ-aminobutyraldehyde and aminoaldehydes derived from polyamines. The ALDH16A1 gene encodes an 802-amino acid protein with as-yet unknown function. Finally, the ALDH18A1 gene encodes D1-pyrroline-5-carboxylate synthetase, which qualifies for classification in the ALDH superfamily based on the sequence homology of one of the protein domains (residues 361–772).

Acknowledgments

We thank Dr Elspeth Bruford and Tia Estey for valuable discussions and a critical reading of this manuscript. The writing of this paper was funded, in part, by NIH grants R01 EY11490 (VV) and P30 ES06096 (DWN).

References

1. Vasiliou, V., Pappa, A. and Petersen, D.R. (2000), ‘Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism’, Chem. Biol. Interact. Vol. 129, pp. 1–19.
2. Vasiliou, V. and Pappa, A. (2000), ‘Polymorphisms of human aldehyde dehydrogenases. Consequences for drug metabolism and disease’, Pharmacology Vol. 61, pp. 192–198.
3. Duester, G. (2000), ‘Families of retinoid dehydrogenases regulating vitamin A function: Production of visual pigment and retinoic acid’, Eur. J. Biochem. Vol. 267, pp. 4315–4324.
4. Niederreither, K., Fraulob, V., Garnier, J.M. et al. (2002), ‘Differential expression of retinoic acid-synthesizing (RALDH) enzymes during fetal development and organ differentiation in the mouse’, Mech. Dev. Vol. 110, pp. 165–171.
5. Niederreither, K., Vermot, J., Schuhbaur, B. et al. (2000), ‘Retinoic acid synthesis and hindbrain patterning in the mouse embryo’, Development Vol. 127, pp. 75–85.
6. Niederreither, K., Subbarayan, V., Dolle, P. et al. (1999), ‘Embryonic retinoic acid synthesis is essential for early mouse post-implantation development’, Nat. Genet. Vol. 21, pp. 444–448.
7. Dupe, V., Matt, N., Garnier, J.M. et al. (2003), ‘A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment’, Proc. Natl. Acad. Sci. USA Vol. 100, pp. 14036–14041.
8. Seiler, N. and Eichentopf, B. (1975), ‘4-aminobutyrate in mammalian putrescine catabolism’, Biochem. J. Vol. 152, pp. 201–210.
9. Sladek, N.E. (2003), ‘Human aldehyde dehydrogenases: Potential pathological, pharmacological, and toxicological impact’, J. Biochem. Mol. Toxicol. Vol. 17, pp. 7–23.
10. Sydow, K., Daiber, A., Oelze, M. et al. (2004), ‘Central role of mitochondrial aldehyde dehydrogenase and reactive oxygen species in nitroglycerin tolerance and cross-tolerance’, J. Clin. Invest. Vol. 113, pp. 482–489.
11. Vasiliou, V., Pappa, A. and Estey, T. (2004), ‘Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism’, Drug Metab. Rev. Vol. 36, pp. 279–299.
12. Nebert, D.W., Roe, A.L., Dieter, M.Z. et al. (2000), ‘Role of the aromatic hydrocarbon receptor and [AH] gene battery in the oxidative stress response, cell cycle control, and apoptosis’, Biochem. Pharmacol. Vol. 59, pp. 65–85.
13. Uma, L., Hariparan, J., Sharma, Y. et al. (1996), ‘Corneal aldehyde dehydrogenase displays antioxidant properties’, Exp. Eye Res. Vol. 63, pp. 117–120.
14. Kirsch, M. and De, G.H. (2001), ‘NAD(P)H, a directly operating antioxidant?’, FASEB J. Vol. 15, pp. 1569–1574.
15. Piatigorsky, J. (2001), ‘Enigma of the abundant water-soluble cytoplasmic proteins of the cornea: The “refraction hypothesis”’, Cornea Vol. 20, pp. 853–858.
16. Jester, J.V., Moller-Pedersen, T., Huang, J. et al. (1999), ‘The cellular basis of corneal transparency: Evidence for “corneal crystallins”’, J. Cell Sci. Vol. 112, pp. 613–622.
17. Vasiliou, V. (1997), ‘Aldehyde dehydrogenase genes’, Adv. Exp. Med. Biol. Vol. 414, pp. 595–600.
18. Vasiliou, V., Baroosh, A., Tipton, K.F. et al. (1999), ‘Eukaryotic aldehyde dehydrogenase (ALDH) genes: Human polymorphisms, and recommended nomenclature based on divergent evolution and chromosomal mapping’, Pharmacogenetics Vol. 9, pp. 421–434.
19. Sophos, N.A. and Vasiliou, V. (2003), ‘Aldehyde dehydrogenase gene superfamily: The 2002 update’, Chem. Biol. Interact. Vol. 143–144, pp. 5–22.
20. Sophos, N.A., Pappa, A., Ziegler, T.L. et al. (2001), ‘Aldehyde dehydrogenase gene superfamily: The 2000 update’, Chem. Biol. Interact. Vol. 130–132, pp. 325–337.
21. Nebert, D.W. and Wain, H.M. (2003), ‘Update on human genome completion and annotations: Gene nomenclature’, Hum. Genomics Vol. 1, pp. 66–71.
22. Pappa, A., Chen, C., Koutalos, Y. et al. (2003), ‘Aldh3a1 protects human corneal epithelial cells from ultraviolet- and 4-hydroxy-2-nonenal-induced oxidative damage’, Free Radiic. Biol. Med. Vol. 34, pp. 1178–1189.
23. Sladek, N.E. (1999), ‘Aldehyde dehydrogenase-mediated cellular relative insensitivity to the oxaazaphosphorines’, Curr. Pharm. Des. Vol. 5, pp. 607–625.
24. Nebert, D.W. and Gonzalez, E.J. (1987), ‘P450 genes: Structure, evolution and regulation’, Annu. Rev. Biochem. Vol. 56, pp. 945–993.
25. Kamino, K., Nagasaka, K., Imagawa, M. et al. (2000), ‘Deficiency in mitochondrial aldehyde dehydrogenase increases the risk for late-onset Alzheimer’s disease in the Japanese population’, Biochem. Biophys. Res. Commun. Vol. 273, pp. 192–196.
26. De Laurenti, L.V., Rogers, G.R., Hamrock, D.J. et al. (1996), ‘Sjögren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene’, Nat. Genet. Vol. 12, pp. 52–57.
27. Willemsen, M.A., Jilt, L., Steijlen, P.M. et al. (2001), ‘Clinical, biochemical and molecular genetic characteristics of 19 patients with the Sjögren-Larsson syndrome’, Brain Vol. 124, pp. 1426–1437.
28. Willemsen, M.A., Rotteveel, J.J., van Domburg, P.H. et al. (1999), ‘Preterm birth in Sjögren-Larsson syndrome’, Neuropediatrics Vol. 30, pp. 325–327.
29. Valle, D., Goodman, S.I., Applegarth, D.A. et al. (1976), ‘Type II hyper-prolactinemia. Delta1-pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes’, J. Clin. Invest. Vol. 58, pp. 598–603.
30. Pearl, P.L., Gibson, K.M., Acosta, M.T. et al. (2003), ‘Clinical spectrum of succinic semialdehyde dehydrogenase deficiency’, Neurology Vol. 60, pp. 1413–1417.
31. Roe, C.R., Struys, E., Kok, R.M. et al. (1998), ‘Methylmalonic semialdehyde dehydrogenase deficiency: Psychomotor delay and methylmalonic aciduria without metabolic decompensation’, Mol. Genet. Metab. Vol. 65, pp. 35–43.
32. Cook, R.J., Lloyd, R.S. and Wagner, C. (1991), ‘Isolation and characterization of cDNA clones for rat liver 10-formyltetrahydrofolate dehydrogenase’, J. Biol. Chem. Vol. 266, pp. 4965–4973.
33. Krupeenko, S.A., Wagner, C. and Cook, R.J. (1997), ‘Domain structure of rat 10-formyltetrahydrofolate dehydrogenase. Resolution of the aminoterminal domain as 10-formyltetrahydrofolate hydratase’, J. Biol. Chem. Vol. 272, pp. 10273–10278.
34. Xiao, C., Xin, H., Dong, A. et al. (1999), ‘A novel calmodulin-like protein gene in rice which has an unusual prolonged C-terminal sequence carrying a putative prenylation site’, DNA Res. Vol. 6, pp. 179–181.
35. Reuland, S.N., Vlasov, A.P. and Krupenko, S.A. (2003), ‘Disruption of a calmodulin central helix-like region of 10-formyltetrahydrofolate dehydrogenase impairs its dehydrogenase activity by uncoupling the functional domains’, J. Biol. Chem. Vol. 278, pp. 22894–22900.
36. Reuland, S.N., Vlasov, A.P. and Krupenko, S.A. (2003), ‘Disruption of a calmodulin central helix-like region of 10-formyltetrahydrofolate dehydrogenase impairs its dehydrogenase activity by uncoupling the functional domains’, J. Biol. Chem. Vol. 278, pp. 22894–22900.
37. Nelson, D.R., Zeldin, D., Hoffman, S. et al. (2004), ‘Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes including nomenclature recommendations for genes, pseudogenes, and alternative-splice variants’, Pharmacogenetics Vol. 14, pp. 1–18.