Published in final edited form as:

Chem Biol Interact. 2015 June 5; 234: 12–17. doi:10.1016/j.cbi.2014.12.022.

Aldehyde dehydrogenase homologous folate enzymes: evolutionary switch between cytoplasmic and mitochondrial localization

Natalia I. Krupenko[a], **Roger S. Holmes**[b], **Yaroslav Tsybovsky**[c], and **Sergey A. Krupenko**[a,*]

[a]Department of Nutrition, UNC-Chapel Hill, UNC Nutrition Research Institute, Kannapolis, NC 28081, United States

[b]The Eskitis Institute for Drug Discovery and School of Natural Sciences, Griffith University, Nathan 4111 Brisbane, Queensland, Australia

[c]Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106, United States

Abstract

Cytosolic and mitochondrial 10-formyltetrahydrofolate dehydrogenases are products of separate genes in vertebrates but only one such gene is present in invertebrates. There is a significant degree of sequence similarity between the two enzymes due to an apparent origin of the gene for the mitochondrial enzyme (*ALDH1L2*) from the duplication of the gene for the cytosolic enzyme (*ALDH1L1*). The primordial ALDH1L gene originated from a natural fusion of three unrelated genes, one of which was an aldehyde dehydrogenase. Such structural organization defined the catalytic mechanism of these enzymes, which is similar to that of aldehyde dehydrogenases. Here we report the analysis of *ALDH1L1* and *ALDH1L2* genes from different species and their phylogeny and evolution. We also performed sequence and structure comparison of ALDH1L enzymes possessing aldehyde dehydrogenase catalysis to those lacking this feature in an attempt to explain mechanistic differences between cytoplasmic ALDH1L1 and mitochondrial ALDH1L2 enzymes and to better understand their functional roles.

Keywords

Folate metabolism; ALDH1L enzymes; Mitochondria; Aldehyde dehydrogenases; Enzyme mechanism; Evolution

*Corresponding author. Tel. + 1 704 250 5053, sergey_krupenko@unc.edu.

1. Introduction

Folate metabolism is crucial for several biosynthetic processes including *de novo* purine and thymidylate generation, synthesis of methionine from homocysteine and biosynthesis of glycine from serine (1,2). It is also involved in the degradation of histidine and glycine and metabolism of betaine and dimethylglycine, which donate carbon groups into folate pool (1,2). Enzymes involved in folate pathways are compartmentalized in the cell between cytoplasm and mitochondria (2). Of note, several folate-dependent reactions take place in both compartments and are catalyzed by cytoplasmic and mitochondrial isozymes. Corresponding cytosolic and mitochondrial forms of folate enzymes are products of separate genes, which have likely arisen from gene duplication (3). In recent years, presence of several folate enzymes in the nucleus has also been established (4,5). This phenomenon, however, is the result of translocation of cytosolic enzymes under certain conditions to allow thymidylate generation directly at DNA replication sites (6,7). Overall, folate-dependent nucleotide and methionine biosynthesis takes place outside of mitochondria and it has been proposed that the mitochondrial folate metabolism plays a supportive role providing cytoplasmic folate metabolism with additional one-carbon groups derived from glycine degradation and betaine/dimethylglycine conversion (8,9).

One of the folate reactions duplicated between cytosol and mitochondria is the conversion of 10-formyltetrahydrofolate to tetrahydrofolate (THF) and CO_2 (10). This reaction is catalyzed by two similar enzymes, cytosolic and mitochondrial 10-formyl-THF dehydrogenases, which are products of separate genes (11). While the precise roles of these enzymes are not clear at present, the cytosolic isoform is likely to serve as a regulator of the overall folate metabolism since it irreversibly removes one-carbon groups from folate pool thus restricting the capacity of folate-dependent biosynthetic reactions (12,13). In agreement with this regulatory function, *ALDH1L1* is ubiquitously silenced in human cancers apparently as a mechanism favoring limitless proliferation (14–16). The function of ALDH1L2 enzyme is even less clear, but it could be involved in the production of formate instead of CO_2 (17).

The cloning of *ALDH1L1* gene in 1991 immediately revealed the fact that it is the product of natural fusion of three unrelated genes (18). One of these genes was an aldehyde dehydrogenase (ALDH) and another was similar to two 10-formyl-THF utilizing enzymes, GARFT and FMT (19). Such gene organization results in the enzyme with two distinct catalytic domains, the amino-terminal folate-binding domain and carboxyl-terminal ALDH domain (20,21). These domains are connected by a short (about a 100 amino acid residues) linker, which is not a part of either domain. We later demonstrated that the linker domain is a structural and functional homolog of acyl carrier proteins (22). The characteristic feature of these proteins, the 4′-phosphopantetheinyl prosthetic group, allows the transfer of the intermediate of the ALDH1L1 catalytic reaction from folate binding site to the ALDH catalytic center (19,23).

Compared to canonical *ALDHs*, which are ancient genes and present in all kingdoms of life, the *ALDH1L1* gene appeared later in evolution: it is not found in plant, bacteria or yeast (3). Our previous phylogenetic analysis pointed to the conclusion that mitochondrial ALDH1L2

has emerged after cytosolic ALDH1L1 and the appearance of the former was traced to bony fish (3). The annotation of additional genomes in recent years indicated the necessity to re-evaluate the evolutionary relationship between *ALDH1L1* and *ALDH1L2*. Here we performed the extended phylogenetic analysis of ALDH1L1 and ALDH1L2 enzymes and compared structures of the enzymes to understand differences in their catalytic abilities.

2. Materials and methods

2.1. ALDH1L1 and ALDH1L2 gene and protein identification

ALDH1L1 and ALDH1L2 sequences for representative vertebrate and invertebrate species were retrieved from ExPASy (http://www.expasy.org) (24) and NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) databases using human (*Homo sapiens*) (11) and zebrafish (*Danio rerio*) (3) ALDH1L1 and ALDH1L2 sequences to seed searches. Identification of these genes was based on high predictive scores (>850) and sequence coverage (>98%) for ALDH1L-like protein sequences listed by NCBI, in each case (Table 1). BLAT searches were performed using relevant ALDH1L1 and ALDH1L2 protein sequences to confirm the presence or absence of these genes among the species examined using the UCSC Genome Browser (25). Predicted gene structures, gene locations and ALDH1L1 and ALDH1L2 amino acid sequences were obtained for each protein identified (Table 1). Prediction of the ALDH1L-like protein N-terminal sequence that may serve as a mitochondrial targeting peptide, and the cleavage site for this peptide, was undertaken using MITOPROT (26).

2.2. Amino acid sequence alignments and phylogenetic analyses

Vertebrate and invertebrate ALDH1L-like sequences were subjected to phylogenetic analysis using the http://www.phylogeny.fr/ portal to enable alignment (MUSCLE), curation (Gblocks), phylogeny (PhyML) and tree rendering (TreeDyn) to reconstruct phylogenetic relationships (27). Vertebrate sequences were identified as members of the ALDH1L1 (cytosolic) or ALDH1L2 (mitochondrial) groups, whereas invertebrate sequences were identified as a members of a single group, designated as ALDH1L1.

2.3. Homology modeling

Homology models of the C-terminal domains of human mtFDH and zebrafish cytosolic FDH (zFDH) were generated using the SWISS-MODEL server as described earlier (17).

3. Results and Discussion

3.1. Predicted gene locations, exonic structures and amino acid sequences for ALDH1L1 and ALDH1L2 genes and proteins

Table 1 summarizes the predicted localization, sizes and numbers of exons for vertebrate *ALDH1L1* and *ALDH1L2* genes and invertebrate *ALDH1L1* genes examined, and for encoded ALDH1L1 and ALDH1L2 subunit amino acid sequences. These predictions were based on BLAST interrogations of ALDH1L1 and ALDH1L2 sequences (http://blast.ncbi.nlm.nih.gov/protein) using human ALDH1L1 and ALDH1L2 (11), and zebra fish ALDH1L1 and ALDH1L2 (3) sequences and from BLAT analyses of vertebrate and invertebrate genomes using the UC Santa Cruz Genome Browser (http://genome.ucsc.edu/

cgi-bin/hgBlat) (25). For every vertebrate genome examined, there were 23 exons observed for *ALDH1L1* and *ALDH1L2* genes, which encode the cytosolic and mitochondrial enzymes, respectively (Table 1). However, in *ALDH1L1* genes the first exon is non-translatable. Alignments of N-terminal amino acid sequences for several vertebrate ALDH1L1 and ALDH1L2 enzymes demonstrated that the additional translatable *ALDH1L2* exon encodes for the N-terminus containing mitochondrial leader sequence, which is responsible for the mitochondrial localization of the enzyme (11) (Fig. 1). Moreover, alignments of invertebrate ALDH1L1 N-terminal sequences demonstrated that these were predominantly consistent with the exonic structure for the vertebrate *ALDH1L1* genes, although with a reduced number of exons overall. The fruit fly (*Drosophila melanogaster*) *ALDH1L1* gene, for example, contained only 2 coding exons, whereas the sea hare *ALDH1L1* gene contained 23 coding exons (Table 1). It is of considerable interest that two isoform sequences were observed for sea squirt (*Ciona intestinalis*) *ALDH1L1* (designated as *ALDH1L1a* and *ALDH1L1b*), which are products of the same gene, but with an additional exon encoding a predicted mitochondrial leader sequence for the ALDH1L1b enzyme (Table 1 and Fig. 1). In addition, sea urchin (*Strongylocentrotus purpuratus)*, trichoplax (*Trichoplax adhaerens*) and termite (*Zootermopsis nevadensis*) ALDH1L1 N-terminal sequences exhibited mitochondrial leader sequence characteristics. It is not known at present, however, whether these proteins are actually localized in the cytoplasm or mitochondria.

3.2. Phylogeny and evolution of ALDH1L1 and ALDH1L2 sequences

A phylogram (Fig. 2) was calculated by the progressive alignment of 14 vertebrate ALDH1L1 and ALDH1L2 and 17 invertebrate ALDH1L1 amino acid sequences. ALDH1L2 sequences were identified for all vertebrate genomes examined, whereas ALDH1L1 sequences were notably absent from the chicken (*Gallus gallus*) and lizard (*Anolis carolensis*) genomes, which suggested that an *ALDH1L1* gene loss event may have taken place during avian and reptilian evolution. Bioinformatic analyses of other bird (zebra finch and turkey) genomes confirmed the apparent absence of the *ALDH1L1* gene from these species (results not shown). These results are consistent with a previous report on vertebrate *ALDH1L1* and *ALDH1L2* gene evolution (3). The phylogram also demonstrated separation of these sequences into two distinct groups during vertebrate evolution (*ALDH1L1* and *ALDH1L2* sequences) and suggested that these genes were derived from an ancestral invertebrate *ALDH1L1* gene, given the consistent presence of this gene among all invertebrate genomes examined.

Our analysis indicates that a single primordial *ALDH1L1* gene encodes for the protein, which can localize either in cytosol or mitochondria of invertebrates depending on species. Because the likely predecessor of the N-terminal domain of ALDH1L1 protein was an FMT-related protein, the mitochondrial localization should not be surprising. Indeed, FMT is a mitochondrial protein, encoded by nuclear DNA, and it has a typical mitochondrial leader sequence (28). Therefore, the fusion protein with the FMT-derived domain at the N-terminus would be expected to have a mitochondrial leader sequence as well, encoded by the FMT-related part of the gene. The alternative splicing *ALDH1L1* mRNA, which would exclude the first exon encoding for the leader sequence, enables the gene to produce enzyme localized to cytosol. Of note, such mechanism is known for another folate enzyme,

mitochondrial serine hydroxymethyltransferase (29). It is likely that this mechanism can be responsible for producing either the cytosolic or mitochondrial enzyme in invertebrate genomes.

While the precise evolutionary path for the *ALDH1L* gene is not clear at present, we hypothesize that the primordial gene encoded the cytoplasmic enzyme, which is the feature of the gene and enzyme observed for the fruit fly (Table 1). Subsequently, the addition of an exon encoding the mitochondrial leader sequence may have provided the option for a mitochondrial form of this enzyme within other invertebrate genomes. The later event, the gene duplication in vertebrates, enabled both cytoplasmic and mitochondrial forms to be encoded by separate genes, rather than arise from distinct isoforms of a single *ALDH1L1* gene, as reported here for the sea squirt genome (Table 1). The loss of translation of the first exon during evolution may have changed the subcellular localization of the enzyme to cytoplasm. The later event, gene duplication in vertebrates, was apparently accompanied by re-activation of translation of the first exon coding the mitochondrial leader sequence thus allowing the mitochondrial localization of the protein. As noted above, our analyses indicated the loss of the *ALDH1L1* gene encoding the cytoplasmic enzyme, in birds and reptiles. This, however, does not necessarily indicate the lack of the enzyme from cytoplasm if alternative splicing takes place in these species. If the cytosolic enzyme indeed is not present in birds and reptiles, this raises the question of functional significance of such evolutionary event. Potential implications of the presence of only mitochondrial enzyme are discussed below.

3.3. Catalytic and structural differences of human ALDH1L1 and ALDH1L2 enzymes

We have previously reported that human mitochondrial ALDH1L2 enzyme lacks ALDH activity while still possesses 10-formyl-THF dehydrogenase activity (17). Interestingly, it has been reported that the cytosolic enzyme from zebrafish, similar to human ALDH1L2, does not catalyze the ALDH reaction (30). Here we have analyzed sequence alignment between ALDH domains (C_t-FDH) of four ALDH1L enzymes, human, rat and zebrafish ALDH1L1 and human ALDH1L2. Residues identical in zebrafish ALDH1L1 and human ALDH1L2 but different from those in human/rat ALDH1L1 were cross-referenced with human ALDH1 and ALDH2. If identified residues were identical between human/rat ALDH1L1 and ALDH1/2, their potential role in catalysis was further analyzed by comparison of structure of rat and zebrafish cytosolic ALDH1L1.

Homology models of the ALDH domains of cytosolic zebrafish and mitochondrial human enzymes are very similar to the structure of rat C_t-FDH. In particular, the residues forming the catalytic center are identical in all three proteins (17,31). The majority of the residues that are different between the two enzymes with no ALDH activity and rat C_t-FDH are located on the surface of the molecule far from catalytic and nucleotide-binding regions. Generally, such residues are unlikely to affect the enzymatic properties. However, we have shown previously that long-range communications are involved in regulating coenzyme binding in the case of C_t-FDH (32,33). Thus, the influence of distant solvent-exposed residues that leads to the absence of ALDH activity in cytosolic zebrafish and mitochondrial human enzymes cannot be completely excluded. A more plausible hypothesis, however, is

that a glutamate residue at position 889 (the number corresponds to zebrafish enzyme) may interfere with shuttling of the proton abstracted by the catalytic glutamate from the catalytic cysteine to the bulk solvent (Fig. 3). According to our previous studies, a network of water molecules and main chain carbonyl groups leading from Glu673 of C_t-FDH (the carboxyl terminal domain of rat ALDH1L1) to the milieu serves to release this proton, which is required for the subsequent hydrolysis of the thiohemiacetal intermediate (31). Ala887 of C_t-FDH is located in the immediate vicinity to this network. The large side chain of Glu889 in cytosolic zebrafish and mitochondrial human enzymes clashes with several water molecules forming the shuttling chain, and this counterproductive interaction may disturb deprotonation of the catalytic glutamate, thereby interfering with the ALDH catalysis.

It is not clear at present whether the difference in the ability to catalyze ALDH reaction is an isolated phenomenon or whether all mitochondrial enzymes and perhaps cytosolic enzymes from lower species lack this feature. Of note, the sequence alignment of ALDH1L1 and ALDH1L2 proteins from different species indicated that the alanine residue proposed to enable the catalysis is found only in mammalian cytosolic enzymes (Fig. 3B). Also, the question of whether ALDH1L1 ALDH activity has an independent functional significance, or whether the enzyme *in vivo* utilizes naturally occurring aldehydes as substrates, has never been addressed. It is an open question what the selective advantage would be for the human mitochondrial enzyme to lack the ALDH activity? Of note, this enzyme has higher 10-formyl-THF hydrolase activity and lower 10-formyl-THF dehydrogenase activity compared to the cytosolic enzyme (17). Therefore, it is possible that the main function of mitochondrial ALDH1L2 is to produce formate rather than CO_2 from folate-bound one-carbon groups. This would be consistent with the hypothesis that mitochondrial one-carbon metabolism serves the function of supplying one-carbon groups, in the form of formate, for the cytosolic one-carbon metabolism (9). Thus, the two ALDH1L enzymes, cytosolic and mitochondrial, are likely to serve different functions in the cell, one is regulatory (ALDH1L1) while the other is biosynthesis supporting (ALDH1L2) (schematically shown in Fig. 4). In agreement with such rather opposite functions, ALDH1L1 is ubiquitously silenced in cancer cell lines while ALDH1L2 is easily detectable there (11,14). The related question is why reptiles and birds lack the gene for cytosolic enzyme? One possible explanation is that the loss of cytosolic ALDH1L1 would be beneficial if these species require a higher rate of purine biosynthesis. Of note, in contrast to other animals, both classes excrete nitrogen as uric acid, which is the product of purine degradation (34,35). Such pathway would require constant biosynthesis of purines as the mechanism of clearance of nitrogen produced by amino acid metabolism. In this case, the lack of Aldh1l1 would allow the utilization of 10-formyl-THF strictly for the *de novo* purine pathway.

4. Conclusion

Our analysis indicated that invertebrate genomes contain a single *ALDH1L*-like gene. Such a gene, however, may result in either cytoplasm or mitochondria-localized protein. Indeed, the primordial *ALDH1L* originated as the fusion with an *FMT* gene, the fact indicating a potential presence of the mitochondrial leader sequence. Of note, a single *ALDH1L* gene may produce both proteins simultaneously as exemplified by the presence of two isoforms in sea squirt. These isoforms are identical except for an additional N-terminal sequence, which

appears to define the mitochondrial localization. We hypothesized that most likely the primordial *ALDH1L* gene originally encoded for a cytoplasmic ALDH1L enzyme while its alternative splicing resulted in the mitochondrial localization of the enzyme in particular species. Later on, the duplication of *ALDH1L* gene resulted in the separation of *ALDH1L* into *ALDH1L1* and *ALDH1L2* forms among vertebrate genomes. Thus, it is possible that this invertebrate *ALDH1L* gene is able to form potential cytosolic and mitochondrial isoforms whereas for vertebrates, separate genes (*ALDH1L1* and *ALDH1L2*) are present. Of note, an additional twist in *ALDH1L* evolution took place when the gene for cytosolic enzyme was apparently lost in reptiles and birds. It is also probable that during evolution the cytosolic ALDH1L1 enzyme acquired the ability to catalyze ALDH reactions, though more data on the activity of ALDH1L enzymes from different species are required to support this hypothesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the National Institutes of Health grant DK054388 and CA095030 (SAK).

Abbreviations

ALDH	aldehyde dehydrogenase
C_t-FDH	carboxyl terminal domain of 10-formyltetrahydrofolate dehydrogenase
THF	tetrahydrofolate

References

1. Fox JT, Stover PJ. Folate-mediated one-carbon metabolism. Vitam Horm. 2008; 79:1–44. [PubMed: 18804690]
2. Tibbetts AS, Appling DR. Compartmentalization of Mammalian folate-mediated one-carbon metabolism. Annu Rev Nutr. 2010; 30:57–81. [PubMed: 20645850]
3. Strickland KC, Holmes RS, Oleinik NV, Krupenko NI, Krupenko SA. Phylogeny and evolution of aldehyde dehydrogenase-homologous folate enzymes. Chem Biol Interact. 2011; 191:122–128. [PubMed: 21215736]
4. Woeller CF, Anderson DD, Szebenyi DM, Stover PJ. Evidence for small ubiquitin-like modifier-dependent nuclear import of the thymidylate biosynthesis pathway. The Journal of biological chemistry. 2007; 282:17623–17631. [PubMed: 17446168]
5. Anderson DD, Eom JY, Stover PJ. Competition between sumoylation and ubiquitination of serine hydroxymethyltransferase 1 determines its nuclear localization and its accumulation in the nucleus. The Journal of biological chemistry. 2012; 287:4790–4799. [PubMed: 22194612]
6. Anderson DD, Woeller CF, Chiang EP, Shane B, Stover PJ. Serine hydroxymethyltransferase anchors de novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. The Journal of biological chemistry. 2012; 287:7051–7062. [PubMed: 22235121]
7. MacFarlane AJ, Anderson DD, Flodby P, Perry CA, Allen RH, Stabler SP, Stover PJ. Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. The Journal of biological chemistry. 2011; 286:44015–44022. [PubMed: 22057276]

8. Barlowe CK, Appling DR. In vitro evidence for the involvement of mitochondrial folate metabolism in the supply of cytoplasmic one-carbon units. Biofactors. 1988; 1:171–176. [PubMed: 2475123]

9. Appling DR. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. Faseb J. 1991; 5:2645–2651. [PubMed: 1916088]

10. Garcia-Martinez LF, Appling DR. Characterization of the folate-dependent mitochondrial oxidation of carbon 3 of serine. Biochemistry. 1993; 32:4671–4676. [PubMed: 8485144]

11. Krupenko NI, Dubard ME, Strickland KC, Moxley KM, Oleinik NV, Krupenko SA. ALDH1L2 is the mitochondrial homolog of 10-formyltetrahydrofolate dehydrogenase. J Biol Chem. 2010; 285:23056–23063. [PubMed: 20498374]

12. Strickland KC, Krupenko NI, Krupenko SA. Molecular mechanisms underlying the potentially adverse effects of folate. Clin Chem Lab Med. 2012:1–10.

13. Oleinik NV, Krupenko NI, Reuland SN, Krupenko SA. Leucovorin-induced resistance against FDH growth suppressor effects occurs through DHFR up-regulation. Biochem Pharmacol. 2006; 72:256–266. [PubMed: 16712799]

14. Krupenko SA, Oleinik NV. 10-formyltetrahydrofolate dehydrogenase, one of the major folate enzymes, is down-regulated in tumor tissues and possesses suppressor effects on cancer cells. Cell Growth Differ. 2002; 13:227–236. [PubMed: 12065246]

15. Rodriguez FJ, Giannini C, Asmann YW, Sharma MK, Perry A, Tibbetts KM, Jenkins RB, Scheithauer BW, Anant S, Jenkins S, Eberhart CG, Sarkaria JN, Gutmann DH. Gene expression profiling of NF-1-associated and sporadic pilocytic astrocytoma identifies aldehyde dehydrogenase 1 family member L1 (ALDH1L1) as an underexpressed candidate biomarker in aggressive subtypes. J Neuropathol Exp Neurol. 2008; 67:1194–1204. [PubMed: 19018242]

16. Chen XQ, He JR, Wang HY. Decreased expression of ALDH1L1 is associated with a poor prognosis in hepatocellular carcinoma. Med Oncol. 2011

17. Strickland KC, Krupenko NI, Dubard ME, Hu CJ, Tsybovsky Y, Krupenko SA. Enzymatic properties of ALDH1L2, a mitochondrial 10-formyltetrahydrofolate dehydrogenase. Chem Biol Interact. 2011; 191:129–136. [PubMed: 21238436]

18. Cook RJ, Lloyd RS, Wagner C. Isolation and characterization of cDNA clones for rat liver 10-formyltetrahydrofolate dehydrogenase. The Journal of biological chemistry. 1991; 266:4965–4973. [PubMed: 1848231]

19. Krupenko SA. FDH: an aldehyde dehydrogenase fusion enzyme in folate metabolism. Chem Biol Interact. 2009; 178:84–93. [PubMed: 18848533]

20. Krupenko SA, Wagner C, Cook RJ. Expression, purification, and properties of the aldehyde dehydrogenase homologous carboxyl-terminal domain of rat 10-formyltetrahydrofolate dehydrogenase. J Biol Chem. 1997; 272:10266–10272. [PubMed: 9092577]

21. Krupenko SA, Wagner C, Cook RJ. Domain structure of rat 10-formyltetrahydrofolate dehydrogenase. Resolution of the amino-terminal domain as 10-formyltetrahydrofolate hydrolase. J Biol Chem. 1997; 272:10273–10278. [PubMed: 9092578]

22. Donato H, Krupenko NI, Tsybovsky Y, Krupenko SA. 10-formyltetrahydrofolate dehydrogenase requires a 4′-phosphopantetheine prosthetic group for catalysis. The Journal of biological chemistry. 2007; 282:34159–34166. [PubMed: 17884809]

23. Strickland KC, Hoeferlin LA, Oleinik NV, Krupenko NI, Krupenko SA. Acyl carrier protein-specific 4′-phosphopantetheinyl transferase activates 10-formyltetrahydrofolate dehydrogenase. J Biol Chem. 2010; 285:1627–1633. [PubMed: 19933275]

24. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. Nucleic acids research. 2012; 40:W597–603. [PubMed: 22661580]

25. Karolchik D, Bejerano G, Hinrichs AS, Kuhn RM, Miller W, Rosenbloom KR, Zweig AS, Haussler D, Kent WJ. Comparative genomic analysis using the UCSC genome browser. Methods in molecular biology. 2007; 395:17–34. [PubMed: 17993665]

26. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. European journal of biochemistry / FEBS. 1996; 241:779–786. [PubMed: 8944766]

27. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic acids research. 2008; 36:W465–469. [PubMed: 18424797]

28. Takeuchi N, Kawakami M, Omori A, Ueda T, Spremulli LL, Watanabe K. Mammalian mitochondrial methionyl-tRNA transformylase from bovine liver. Purification, characterization, and gene structure. The Journal of biological chemistry. 1998; 273:15085–15090. [PubMed: 9614118]

29. Anderson DD, Stover PJ. SHMT1 and SHMT2 are functionally redundant in nuclear de novo thymidylate biosynthesis. PloS one. 2009; 4:e5839. [PubMed: 19513116]

30. Chang WN, Lin HC, Fu TF. Zebrafish 10-formyltetrahydrofolate dehydrogenase is similar to its mammalian isozymes for its structural and catalytic properties. Protein expression and purification. 2010; 72:217–222. [PubMed: 20381623]

31. Tsybovsky Y, Donato H, Krupenko NI, Davies C, Krupenko SA. Crystal structures of the carboxyl terminal domain of rat 10-formyltetrahydrofolate dehydrogenase: implications for the catalytic mechanism of aldehyde dehydrogenases. Biochemistry. 2007; 46:2917–2929. [PubMed: 17302434]

32. Tsybovsky Y, Krupenko SA. Conserved Catalytic Residues of the ALDH1L1 Aldehyde Dehydrogenase Domain Control Binding and Discharging of the Coenzyme. J Biol Chem. 2011; 286:23357–23367. [PubMed: 21540484]

33. Tsybovsky Y, Malakhau Y, Strickland KC, Krupenko SA. The mechanism of discrimination between oxidized and reduced coenzyme in the aldehyde dehydrogenase domain of Aldh1l1. Chem Biol Interact. 2013

34. Wright PA. Nitrogen excretion: three end products, many physiological roles. The Journal of experimental biology. 1995; 198:273–281. [PubMed: 7699310]

35. Singer MA. Do mammals, birds, reptiles and fish have similar nitrogen conserving systems? Comparative biochemistry and physiology Part B, Biochemistry & molecular biology. 2003; 134:543–558.

Highlights

We have performed a phylogenetic analysis of ALDH1L1 and ALDH1L2 enzymes

Single invertebrate *ALDH1L* gene can produce cytosolic and mitochondrial enzymes

ALDH1L2 gene is a likely result of the duplication of *ALDH1L1* gene

During evolution cytosolic ALDH1L1 enzyme acquired ALDH catalysis

Fig. 1.

Alignment of N-terminal amino acid sequences for vertebrate and invertebrate Aldh1l proteins ("*" indicates identical residues; ":" similar alternate residues; "." less similar alternate residues; known or predicted exon junctions are shown in boldface; predicted mitochondrial leader sequences are shaded; two isoforms for sea squirt ALDH1L1, 1L1a and 1L1b, were examined).

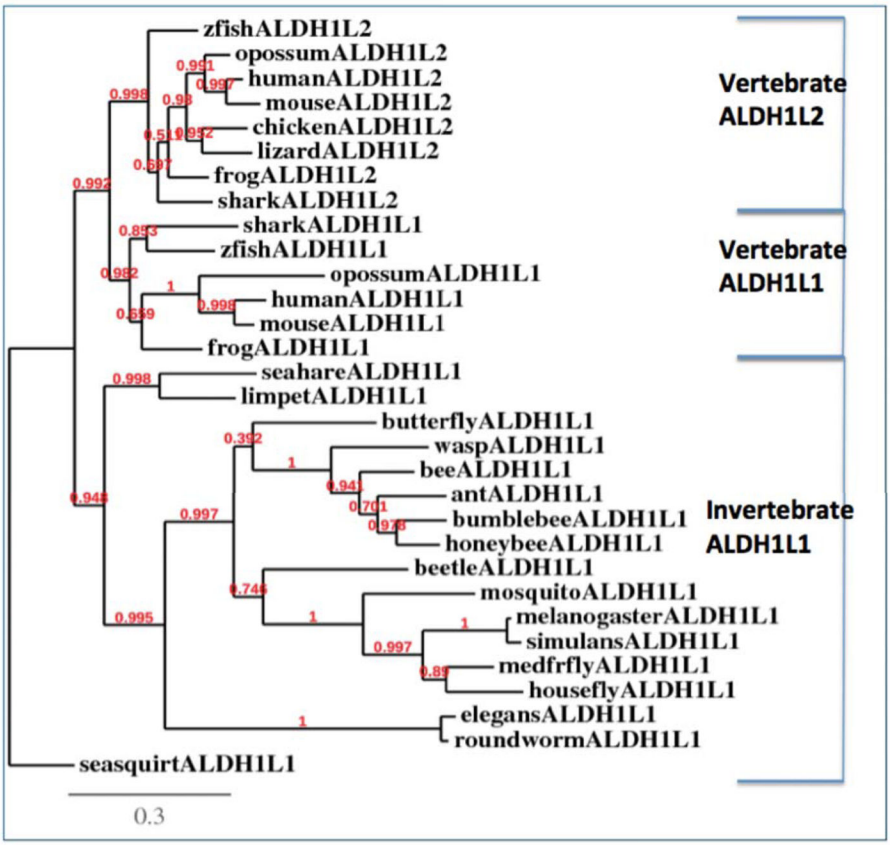
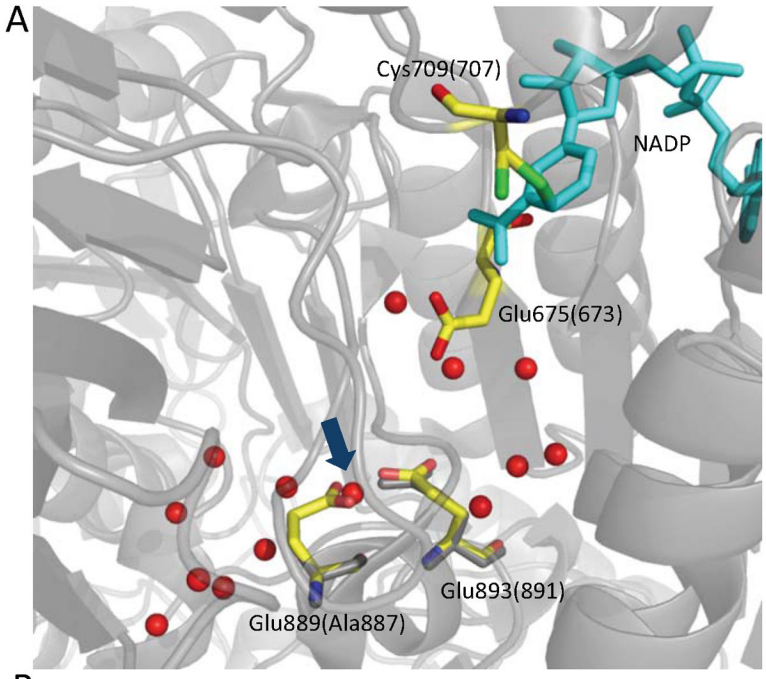


Fig. 2.

Phylogenetic tree for invertebrate ALDH1L1 and vertebrate ALDH1L1 and ALDH1L2 sequences. The tree is labeled with the gene name and the name of the species. Note the major clusters for the invertebrate ALDH1L1 and for vertebrate ALDH1L1 and ALDH1L2 sequences. Note the absence of ALDH1L2 for invertebrate species and ALDH1L1 for chick and lizard genomes. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates is represented as a fraction out of 100. Only replicate values of 0.9 or more are highly significant, with 100 bootstrap replicates performed in each case.



B

```
humanALDH1L1          KQSGFGKDLGEAALNEYLRVKTVTFEY 902
ratALDH1L1            KQSGFGKDLGEAALNEYLRIKTVTFEY 902
mouseALDH1L1          KQSGFGKDLGEAALNEYLRIKTVTFEY 902
frogALDH1L1           KQSGFGKDLGEEALNEYLKTKAVTIEY 902
zfishALDH1L1          KQSGFGKDLGQEALNEYLKTKCVTVEY 904
seasquirtALDH1L1      KQSGFGKDLGEDALNEYLKTKAVTLEY 898
humanALDH1L2          KQSGFGKDLGEEALNEYLKTKTVTLEY 923
mouseALDH1L2          KQSGFGKDLGEEALNEYLKIKTVTLEY 923
frogALDH1L2           KQSGFGKDLGEEALHEYLRTKAVTVEY 922
zfishALDH1L2          KQSGFGKDLGEDALHEYLRTKAVTVEY 923
                      **********: **.***: * **.**
```

Fig. 3.

(A) Superposition of the model of the ALDH domain of zebrafish ALDH1L1 and the structure of rat C_t-FDH (PDB 202Q, shown in gray). The amino acid numbering is according to zebrafish ALDH1L1 (corresponding numbers for rat ALDH1L1 are shown in parentheses). The proton shuttling chain (31) consists of water molecules (red spheres) and Glu893(891). The side chain of Glu889 of zebrafish ALDH1L1 clashes with water molecules forming the chain (*arrow*). (B) Sequence alignment of the carboxyl terminus of ALDH1L enzymes ("*" indicates identical residues; ":" similar alternate residues; "." less similar alternate residues; the residue suggested as the discriminator for the aldehyde dehydrogenase activity is in *boldface*).

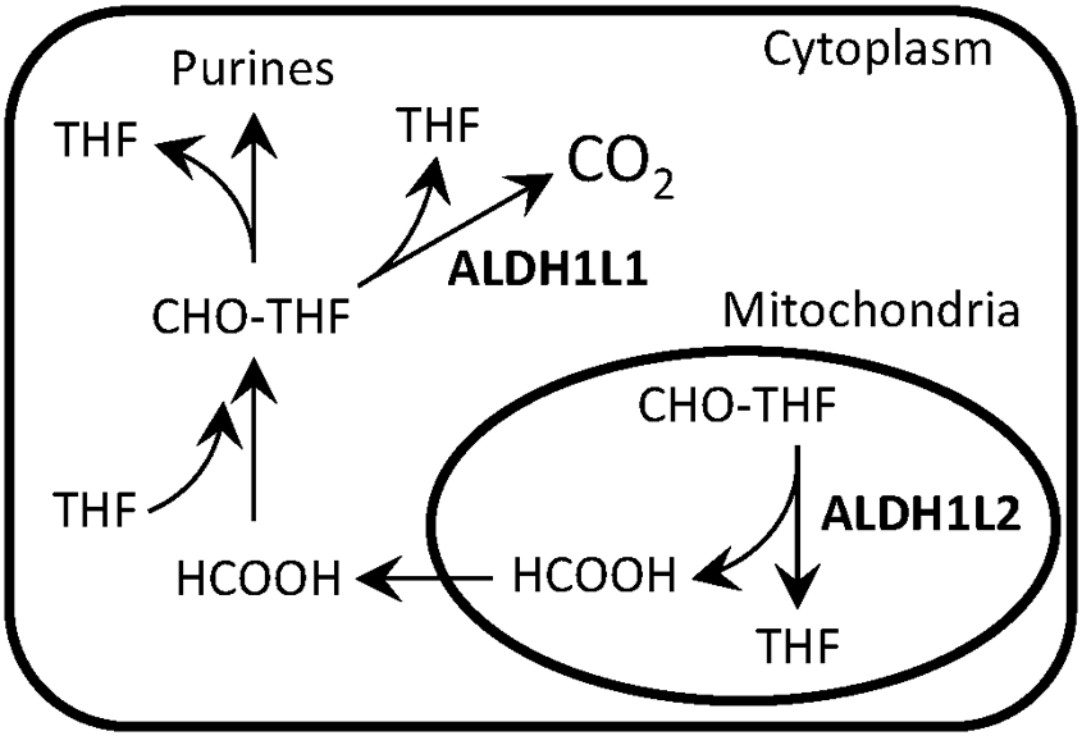


Fig. 4.

Schematic depicting proposed function for cytosolic and mitochondrial ALDH1L enzymes. Cytosolic ALDH1L1 converts 10-formyl-THF (CHO-THF) to THF and CO_2, thus serving a catabolic function of removing one-carbon groups from folate pool. Mitochondrial ALDH1L2 generates formate, which is transported to cytoplasm and upon incorporation into the cytoplasmic folate pool can be used for biosynthetic reactions.

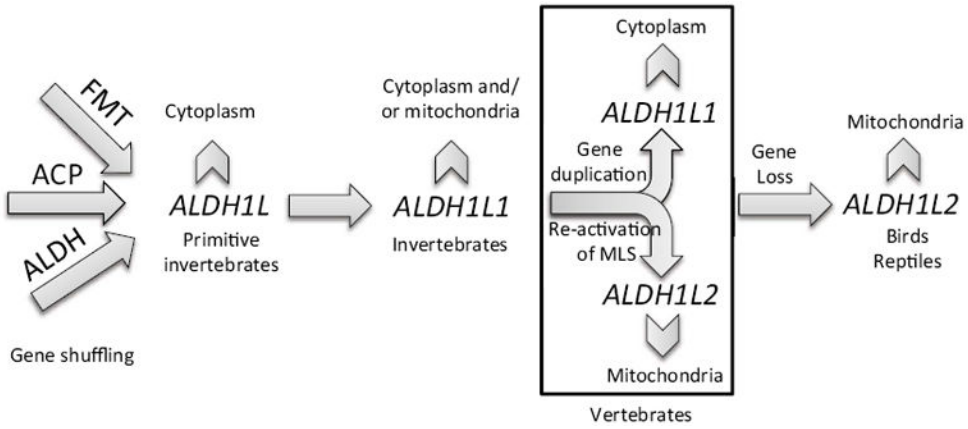


Fig. 5.

Proposed origin and evolution of the *ALDH1L* gene.

Table 1

Invertebrate *ALDH1L1* and vertebrate *ALDH1L1* and *ALDH1L2* genes and proteins.

Animal	Species	Gene	RefSeq ID [1] Prediction	GenBank ID	[2] Exons (strand)	Gene Size (bp)	Amino acids	Localization	Leader peptide
Human	*Homo sapiens*	*ALDH1L1*	NM_012190.3	AF052732	22 (−ve)	57,186	902	cytosol	NA
Human	*Homo sapiens*	*ALDH1L2*	NM_001034173.3	BC103934	23 (−ve)	60,010	923	mitochondria	1….20
Mouse	*Mus musculus*	*Aldh1l1*	NM_027406.1	BC024055	22 (+ve)	41,715	902	cytosol	NA
Mouse	*Mus musculus*	*Aldh1l2*	NM_153543.2	BC034531	23 (−ve)	43,433	923	mitochondria	1….20
Chicken	*Gallus gallus*	*ALDH1L2*	XP_416314.2[1]	NA	23 (+ve)	27,506	922	mitochondria	1….20
Lizard	*Anolis carolinensis*	*ALDH1L2*	XP_003220962.1[1]	NA	23 (−ve)	30,373	924	mitochondria	1….20
Frog	*Xenopus tropicalis*	*ALDH1L1*	NM_001011027.1	BC082822	22 (+ve)	16,496	902	cytosol	NA
Frog	*Xenopus tropicalis*	*ALDH1L2*	XP_002938116.1[1]	NA	23 (+ve)	27,966	924	mitochondria	1….33
Zebra fish	*Danio rerio*	*ALDH1L1*	NM_001198772.1	NA	22 (+ve)	27,616	904	cytosol	NA
Zebra fish	*Danio rerio*	*ALDH1L2*	XP_002661418.2[1]	NA	22 (+ve)	19,873	923	mitochondria	1….44
Shark	*Callorhinchus milii*	*ALDH1L1*	XP_007888551.1[1]	NA	22 (−ve)	15,784	901	cytosol	NA
Shark	*Callorhinchus milii*	*ALDH1L2*	XP_007907882.1[1]	JW862169	23 (+ve)	27,862	922	mitochondria	1….19
Sea squirt	*Ciona intestinalis*	*ALDH1L1a*	XP_002130073.1	NA	17 (+ve)	7,083	898	cytosol	NA
Sea squirt	*Ciona intestinalis*	*ALDH1L1b*	XP_002130073.2	NA	18 (+ve)	7,339	921	mitochondria	1….12
Sea urchin	*S. purpuratus*	*ALDH1L1*	XP_784777.3[1]	NA	22 (−ve)	25,243	927	mitochondria	1….18
Sea hare	*Aplysia californica*	*ALDH1L1*	XP_005090853.1[1]	NA	23 (+ve)	19,195	900	cytosol	NA
Trichoplax	*Trichoplax adhaerens*	*ALDH1L1*	XP_002111368.1[1]	NA	NA	NA	921	mitochondria	1….18
Worm	*Caenorhabditis elegans*	*ALDH1L1*	NM_069653.6	NA	7 (+ve)	3,128	908	cytosol	NA
Round worm	*Caenorhabditis brenneri*	*ALDH1L1*	GL379933.1[1]	EGT36278.1	7 (−ve)	3,102	908	cytosol	NA
Fruit fly	*Drosophila melanogaster*	*ALDH1L1*	NP_610107.1	CG8665	2 (+ve)	3,149	913	cytosol	NA
Mosquito	*Anopheles gambiae*	*ALDH1L1*	XP_318614.3[1]	NA	2 (−ve)	2,820	916	cytosol	NA
House fly	*Musca domestica*	*ALDH1L1*	XP_005181895.1[1]	NA	NA	NA	912	cytosol	NA
Bee	*Apis mellifera*	*ALDH1L1*	XM_623795[1]	NA	6 (−ve)	3,404	900	cytosol	NA
Butterfly	*Danaus plexippus*	*ALDH1L1*	NA	EHJ79154.1	NA	NA	927	cytosol	NA
Water flea	*Daphnia pulex*	*ALDH1L1*	NA	EFX71787.1	NA	NA	924	cytosol	NA
Wasp	*Nasonia vitripennis*	*ALDH1L1*	XP_001602871.1[1]	NA	NA	NA	902	cytosol	NA

Animal	Species	Gene	RefSeq ID [1] Prediction	GenBank ID	[2] Exons (strand)	Gene Size (bp)	Amino acids	Localization	Leader peptide
Beetle	*Tribolium castanaem*	*ALDH1L1*	XP_969916.1[1]	NA	NA	NA	915	cytosol	NA
Ant	*Camponotus floridanus*	*ALDH1L1*	ENF71966.1[1]	NA	NA	NA	900	cytosol	NA
Termite	*Zootermopsis nevadensis*	*ALDH1L1*	KDR07781.1[1]	NA	NA	NA	922	mitochondria	1….21

RefSeq refers to the NCBI reference sequence;

[1] predicted NCBI sequence;

[2] the number of translatable exons is shown; NA, not available; “bp” refers to base pairs of nucleotide sequence; the length of the predicted mitochondrial leader sequence is shown.