Chemico-Biological Interactions 234 (2015) 12–17

Contents lists available at ScienceDirect

Chemico-Biological Interactions

journal homepage: www.elsevier.com/locate/chembioint

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

Natalia I. Krupenko [a], Roger S. Holmes [b], Yaroslav Tsybovsky [c], 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

ARTICLE INFO

Article history:
Available online 27 December 2014

Keywords:
Folate metabolism
ALDH1L enzymes
Mitochondria
Aldehyde dehydrogenases
Enzyme mechanism
Evolution

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.

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

Abbreviations: ALDH, aldehyde dehydrogenase; C_1-FDH, carboxyl terminal domain of 10-formyltetrahydrofolate dehydrogenase; THF, tetrahydrofolate.

* Corresponding author. Tel.: +1 704 250 5053.
E-mail address: sergey_krupenko@unc.edu (S.A. Krupenko).

http://dx.doi.org/10.1016/j.cbi.2014.12.022
0009-2797/

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].

Table 1
Invertebrate *ALDH1L1* and vertebrate *ALDH1L1* and *ALDH1L2* genes and proteins.

Animal	Species	Gene	RefSeq ID [a]Prediction	GenBank ID	[b]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[a]	NA	23 (+ve)	27,506	922	Mitochondria	1…20
Lizard	*Anolis carolinensis*	*ALDH1L2*	XP_003220962.1[a]	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[a]	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[a]	NA	22 (+ve)	19,873	923	Mitochondria	1…44
Shark	*Callorhinchus milii*	*ALDH1L1*	XP_007888551.1[a]	NA	22 (−ve)	15,784	901	Cytosol	NA
Shark	*Callorhinchus milii*	*ALDH1L2*	XP_007907882.1[a]	JW862169	23 (+ve)	27,862	922	Mitochondria	1…19
Sea squirt	*Ciona intestinalis*	*ALDH1L1a*	XP_002130073.1	NA	17 (+ve)	7083	898	Cytosol	NA
Sea squirt	*Ciona intestinalis*	*ALDH1L1b*	XP_002130073.2	NA	18 (+ve)	7339	921	Mitochondria	1…12
Sea urchin	*Strongylocentrotus purpuratus*	*ALDH1L1*	XP_784777.3[a]	NA	22 (−ve)	25,243	927	Mitochondria	1…18
Sea hare	*Aplysia californica*	*ALDH1L1*	XP_005090853.1[a]	NA	23 (+ve)	19,195	900	Cytosol	NA
Trichoplax	*Trichoplax adhaerens*	*ALDH1L1*	XP_002111368.1[a]	NA	NA	NA	921	Mitochondria	1…18
Worm	*Caenorhabditis elegans*	*ALDH1L1*	NM_069653.6	NA	7 (+ve)	3128	908	Cytosol	NA
Round worm	*Caenorhabditis brenneri*	*ALDH1L1*	GL379933.1[a]	EGT36278.1	7 (−ve)	3102	908	Cytosol	NA
Fruit fly	*Drosophila melanogaster*	*ALDH1L1*	NP_610107.1	CG8665	2 (+ve)	3149	913	Cytosol	NA
Mosquito	*Anopheles gambiae*	*ALDH1L1*	XP_318614.3[a]	NA	2 (−ve)	2820	916	Cytosol	NA
House fly	*Musca domestica*	*ALDH1L1*	XP_005181895.1[a]	NA	NA	NA	912	Cytosol	NA
Bee	*Apis mellifera*	*ALDH1L1*	XM_623795[a]	NA	6 (−ve)	3404	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[a]	NA	NA	NA	902	Cytosol	NA
Beetle	*Tribolium castaneum*	*ALDH1L1*	XP_969916.1[a]	NA	NA	NA	915	Cytosol	NA
Ant	*Camponotus floridanus*	*ALDH1L1*	ENF71966.1[a]	NA	NA	NA	900	Cytosol	NA
Termite	*Zootermopsis nevadensis*	*ALDH1L1*	KDR07781.1[a]	NA	NA	NA	922	Mitochondria	1…21

NA, not available; "bp" refers to base pairs of nucleotide sequence; the length of the predicted mitochondrial leader sequence is shown.
RefSeq refers to the NCBI reference sequence;
[a] Predicted NCBI sequence.
[b] The number of translatable exons is shown.