Chemico-Biological Interactions 234 (2015) 80–84

Contents lists available at ScienceDirect

Chemico-Biological Interactions

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

The mammalian alcohol dehydrogenase genome shows several gene duplications and gene losses resulting in a large set of different enzymes including pseudoenzymes

Linus J. Östberg [a], Bengt Persson [b,c], Jan-Olov Höög [d,*]

[a] Department of Medical Biochemistry and Biophysics, Science for Life Laboratory, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[b] Department of Cell and Molecular Biology, Science for Life Laboratory, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[c] Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-751 24 Uppsala, Sweden
[d] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden

ARTICLE INFO

Article history:
Available online 3 December 2014

Keywords:
Alcohol dehydrogenase
Gene duplications
Phylogenetic tree
Protein sequences
Pseudoenzyme

ABSTRACT

Mammalian alcohol dehydrogenase (ADH) is a protein family divided into six classes and the number of known family members is increasing rapidly. Several primate genomes are completely analyzed for the ADH region, where higher primates (human and hominoids) have seven genes of classes ADH1–ADH5. Within the group of non-hominoids apes there have been further duplications and species with more than the typical three isozymic forms for ADH1 are present. In contrast there are few completely analyzed ADH genomes in the non-primate group of mammals, where an additional class has been identified, ADH6, that has been lost during the evolution of primates. In this study 85 mammalian genomes with at least one ADH gene have been compiled. In total more than 500 ADH amino acid sequences were analyzed for patterns that distinguish the different classes.

For ADH1–ADH4 intensive investigations have been performed both at the functional and at structural levels. However, a corresponding functional protein to the *ADH5* gene, which is found in most ADH genomes, has never been detected. The same is true for ADH6, which is only present in non-primates.

The entire mammalian ADH family shows a broad spectrum of gene duplications and gene losses where the numbers differ from six genes (most non-primate mammals) up to ten genes (vole). Included in these sets are examples of pseudogenes and pseudoenzymes.

1. Introduction

The mammalian alcohol dehydrogenase (ADH) family is a set-up of different classes, isozymes and allelic variants with a wide range of alcohol oxidation functions. The set-up within mammals is divided into six classes, ADH1–ADH6 according to protein primary structure and function [1,2]. The only ADH that can be traced in almost all species is ADH3, which has been shown to be the ancestral ADH form from which the other ADH classes have emerged through gene duplications [3]. ADH3 is in addition the only ADH with a designated function as glutathione-dependent formaldehyde dehydrogenase [4] and in line with this the protein is found in species that harbor glutathione. In addition to this oxidative activity ADH3 harbors the specific reductive activity of S-nitrosoglutathione reductase. Both activities act as scavengers for formaldehyde and NO, respectively [5]. The formation of the six classes through gene duplications has been distinguished from alignment of amino acid sequences and phylogenetic tree generation [2,6].

The ADH "genome" can be seen as a cassette of genes where the gene transcription order is the same for all species, *ADH4–ADH1–ADH6–ADH5–ADH2–ADH3* (numbering is due to order of identification). The evolution of the mammalian ADHs shows both gene duplications and losses [6,7]. In the primate group there has been a loss of *ADH6*, probably in parallel with the duplication of *ADH1*gene [6,7]. For the higher classes, ADH5 and ADH6, no native proteins have been traced and for ADH6 no transcription product has been identified yet [8].

Within this report we focus on the gene duplications and gene losses within the mammalian class of ADH and the identification of sequence motifs that can distinguish the different classes. Furthermore, the denotation pseudoenzyme is introduced for the ADH family, i.e. an identified protein with an expected enzymatic

* Corresponding author. Tel.: +46 8 524 87740.
E-mail address: jan-olov.hoog@ki.se (J.-O. Höög).

http://dx.doi.org/10.1016/j.cbi.2014.11.020
0009-2797/

function but without any traced activity [9].This in contrast to pseudogenes that are non-functional genes within a gene family, which have lost their protein-coding ability [10].

2. Materials and methods

2.1. BLAST pipeline

The mammalian ADH sequences were obtained from the July 2014 releases of the UniProt and NCBI nr databases using an automated and reproducible strategy. The first step is a blastp [11] search for each of the sequences that are known in each of the six classes. All aligned sequences giving a local alignment with more than 70% query sequence coverage (local alignment longer than ~260 residues) and 68% query sequence identity were considered hits. Further filtering was performed to remove sequences from non-mammals. The sequence identity between ADH class I and IV are in multiple cases larger than 70%. In order to avoid misclassification, all sequences were confirmed to be located in the correct branch of a phylogenetic tree of all sequences. The tree was generated as described below.

2.2. Phylogenetic tree

As an outgroup sequence, salmon ADH3 (C0H868) was obtained from UniProt and added to the list of all the ADHs identified above. The sequences were aligned using Kalign 2.04 [12]. The JalView 2.8.1 [13] neighbor-joining algorithm based on sequence identity was used to generate an unrooted phylogenetic tree from the alignment. The tree was visualized and rerooted to the salmon sequence using Dendroscope 3.2.10 [14].

3. Results

Scanning the databases NCBI and UniProt for mammalian ADHs using blastp identified a total of 85 mammalian species harboring at least one ADH (Table 1). The final set, after removal of redundant and non-mammalian sequences, consisted of 584 sequences divided into the six mammalian ADH classes. This is an overestimate of the actual number of ADH sequences in the databases of species with multiple forms of the same protein sequence (e.g. length variants, allelic forms, mutations etc.).

The number of identified ADH1 is by far the largest: totaling 102 different amino acid sequences (where the absolute majority is translated genomic sequences). Gene duplications are identified for several ADH1s that increased the total number substantially. All primates show gene duplications for ADH1, where human and human-like apes all show three isoforms of ADH1 (A, B and C). In other monkeys, e.g. Rhesus macaque, *ADH1* is duplicated into five genes with the two additional genes classified as *ADH1-like*, whereas one is postulated to be a pseudoform as no initiation codon has been identified [7]. Horse was the first species with characterized isoforms for ADH (ADH E and S [15]), which were identified before the class system was introduced. In this compilation of data, vole is the only non-primate species except horse that harbor isoforms for ADH1 (Fig. 1). Four *ADH1* genes have been detected in the vole genome where all show a high sequence identity to each other, 93–95%.

Hare animals are the only species that show gene duplications for ADH2, but for several species the *ADH2* gene has not been identified. Rodents have Pro at position 47, which results in an inactive enzyme (Table 2). No ADH3 gene duplication has been identified so far in any species. Platypus is the only species with identified ADH4 isoforms [16]. ADH5 seems to be a general class in mammals, but in the mouse genome this form is identified as a true pseudogene due to the lack of several exons [17]. In the platypus no ADH5 has been identified, but at the localization for ADH5 a pseudogene was identified.

ADH6 is a further puzzling ADH class. No transcript or protein has been reported and from an evolutionary point of view two subgroups can be distinguished (Fig. 2). Gene duplications have been identified for rodents (rat, mouse and vole with A- and B-forms) and for hare animals (rabbit and pika). The vole is the species showing most ADH forms of the mammalian genomes analyzed with a total of ten genes. All of these genes code for ADH proteins of 374–377 amino acid residues and all six ADH classes are represented where ADH1 is divided in four isoforms and ADH6 in two.

Table 1
Compilation of mammalian species with an identified alcohol dehydrogenase. Taxonomy order names are shown in bold.

Primates	**Artiodactyla**	**Rodents**	**Chiroptera**
Human	Cattle	Rat	Brandt's bat
Chimpanzee	River buffalo	Mouse	Big brown bat
Bonobo (Pygmy chimpanzee)	Yak	Guinea pig	Little brown bat
Gorilla	Tibetan antelope	Chinchilla	Mouse-eared bat
Orangutan (Sumatran)	Sheep	Chinese hamster	Vampire bat
White-cheeked gibbon	Goat	Golden hamster	Black flying fox
Rhesus macaque	Alpaca	Egyptien Jerboa	**Soricomorpha/Scandentia**
Crab-eating macaque	Camel	Naked mole-rat	Shrew, European
Baboon (olive)	Pig	Prairie vole	Cape elephant shrew
Hamadryas baboon	**Cetacea**	Degu	Common tree shrew
Squirrel monkey (Bolivian)	Minke whale	Deer-mouse	Chinese tree shrew
White-tufted ear Marmorset	Chinese river dolphin	Ground squirrel, 13-lined	Northen tree shrew
Ring-tailed lemur	Bottlenose dolphin	Pocket gopher, Attwater's	Cape golden mole
Collared brown lemur	Sperm whale	Pocket gopher, Plains	Star-nosed mole
Green monkey	Killer whale	Pocket gopher, Knox Jone's	**Sirenia**
Coquerel's Sifaka	Weddell seal	Pocket gopher, Ilano	West indian manatee
Small-eared galago	**Carnivora**	**Lagomorpha**	**Eulipotyphla**
Tarsier	Giant Panda	Rabbit	Hedgehog, European
Galago	Dog	Pika, American	Lesser hedgehog tenrec
Perissodactyla	Cat	**Tubulidentata**	**Didelphimorphia**
Horse	Tiger, amur	Aardvark	Opossum
Przewalski's horse	Ferret	**Dermoptera**	**Dasyuromorphia**
White rhinoceros	Mink, American	Sunda flying lemur	Tasmanian devil
Proboscidea	Walrus, Pacific	**Cingulata**	**Monotreme**
Elephant, African		Armadillo, 9-banded	Platypus