Chemico-Biological Interactions 234 (2015) 59–74

Contents lists available at ScienceDirect

Chemico-Biological Interactions

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

Residues that influence coenzyme preference in the aldehyde dehydrogenases

Lilian González-Segura [a], Héctor Riveros-Rosas [b], Adriana Julián-Sánchez [b], Rosario A. Muñoz-Clares [a,*]

[a] Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, México D. F. 04510, Mexico
[b] Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, México D. F. 04510, Mexico

ARTICLE INFO

Article history:
Available online 16 January 2015

Keywords:
NAD(P)$^+$ specificity
Dual-coenzyme specificity
Conformational changes
Kinetic parameters
Structural and multiple sequence alignments
Logos analysis

ABSTRACT

To find out the residues that influence the coenzyme preference of aldehyde dehydrogenases (ALDHs), we reviewed, analyzed and correlated data from their known crystal structures and amino-acid sequences with their published kinetic parameters for NAD(P)$^+$. We found that the conformation of the Rossmann-fold loops participating in binding the adenosine ribose is very conserved among ALDHs, so that coenzyme specificity is mainly determined by the nature of the residue at position 195 (human ALDH2 numbering). Enzymes with glutamate or proline at 195 prefer NAD$^+$ because the side-chains of these residues electrostatically and/or sterically repel the 2′-phosphate group of NADP$^+$. But contrary to the conformational rigidity of proline, the conformational flexibility of glutamate may allow NADP$^+$-binding in some enzymes by moving the carboxyl group away from the 2′-phosphate group, which is possible if a small neutral residue is located at position 224, and favored if the residue at position 53 interacts with Glu195 in a NADP$^+$-compatible conformation. Of the residues found at position 195, only glutamate interacts with the NAD$^+$-adenosine ribose; glutamine and histidine cannot since their side-chain points are opposite to the ribose, probably because the absence of the electrostatic attraction by the conserved nearby Lys192, or its electrostatic repulsion, respectively. The shorter side-chains of other residues—aspartate, serine, threonine, alanine, valine, leucine, or isoleucine—are distant from the ribose but leave room for binding the 2′-phosphate group. Generally, enzymes having a residue different from Glu bind NAD$^+$ with less affinity, but they can also bind NADP$^+$ even sometimes with higher affinity than NAD$^+$, as do enzymes containing Thr/Ser/Gln195. Coenzyme preference is a variable feature within many ALDH families, consistent with being mainly dependent on a single residue that apparently has no other structural or functional roles, and therefore can easily be changed through evolution and selected in response to physiological needs.

1. Introduction

The aldehyde dehydrogenase (ALDH) superfamily includes NAD$^+$- and NADP$^+$-specific as well as dual-coenzyme specificity enzymes, and the mode of binding of the coenzymes has been determined at atomic detail in several of them. In ALDHs the coenzyme binds in a five-stranded open α/β domain—the Rossmann fold [1]—, which in ALDHs has differences with the "classical" domain of other pyridine nucleotide-dependent dehydrogenases [2,3]. The study of the structural reasons behind the coenzyme preference of ALDHs was addressed long ago by site-directed mutagenesis and X-ray crystallography [4–7]. These early studies showed the importance of having a Glu at a position in the Rossmann fold similar to the acidic residue (Asp or Glu) in other pyridine nucleotide-linked dehydrogenases for discriminating against NADP(H) by sterically and electrostatically repelling the 2′-phosphate group of their adenosine ribose [5]. In some other NAD$^+$-dependent ALDHs it was found that a Pro in the place of this Glu prevents the binding of the 2′-phosphate group [8], which was allowed in enzymes having a small and neutral residue, such as Thr or Ser [6,7]. But ALDH enzymes that have Glu at this critical position and are able to bind NADP(H) were also found [5,9]. Therefore, in addition to Glu, other residues are involved in determining the coenzyme preference in these enzymes. In order to find out these residues, and to learn more about how the enzymes of this important superfamily bind the coenzymes, here we review, analyze,

Abbreviations: ALDH, aldehyde dehydrogenase; *Hs*ALDH2, aldehyde dehydrogenase 2 from human; PDB, protein data bank; *Pa*BADH, betaine aldehyde dehydrogenase from *Pseudomonas aeruginosa*.

* Corresponding author. Tel.: +52 55 56225276; fax: +52 55 56225329.
E-mail address: clares@unam.mx (R.A. Muñoz-Clares).

http://dx.doi.org/10.1016/j.cbi.2014.12.039
0009-2797/

summarize and correlate the available crystallographic, amino acid sequence and kinetic data of a significant number of ALDH enzymes.

2. Methods

2.1. Structural comparisons

Structural comparisons of the ALDH crystal structures available in the PDB were made using PyMOL (http://www.pymol.org) and Coot [10].

2.2. Sequence analyses

ALDH amino acid sequences were retrieved by Blastp searches at the UniProt site [11] (http://www.uniprot.org./blast). To identify protein sequences as members of the different reported ALDH families we carried out phylogenetic analyses using the MEGA6 software [12] (http://www.megasoftware.net). Progressive multiple protein sequence alignments were performed with ClustalX version 2 [13] (http://www.clustal.org). Analysis of residues at selected positions in the alignment was performed with ProfileGrid (http://www.profilegrid.org). The unrooted phylogenetic tree was inferred from 500 replicates, using the Maximum Likelihood method. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories; +G, parameter = 1.0567). Sequence logos were constructed using the WebLogo server (http://weblogo.threeplusone.com). Each logo consists of stacks of amino acid letters, one stack for each position in the sequence. The height of letters within the stack indicates the relative frequency of each amino acid at that position [14].

3. Results and discussion

3.1. Structural analysis

As a first step to identify the structural factors relevant to the binding of the adenosine ribose of NAD^+ and $NADP^+$ and to analyze their different modes of binding, we examined the X-ray structures of ALDHs so far deposited in the PDB, both as apo forms and in complex with different nucleotides. The structures examined constitute a significant sample that includes 67 different ALDH enzymes from different organisms, belonging to 14 different already classified families, as well as eight structures of five enzymes that belong to families not yet classified by the ALDH Gene Nomenclature Committee (http://www.aldh.org).

3.2. Kinetic and equilibrium binding data analysis

In order to correlate the structural data with the coenzyme preference, we retrieved from the literature the kinetic and binding data of 74 enzymes, which belong to 13 classified families and to nine unclassified (Table 1). When analyzing the kinetic data we took into account several considerations. First, we assumed that the differences in the kinetic parameters values for the NAD^+- and $NADP^+$-dependent reactions are mainly due to the absence or presence of the 2′-phosphate group in the adenosine ribose, given that, with the exception of ALDH3 enzymes, the structural analysis indicated that the rest of the molecule binds similarly, regardless of the nucleotide (not shown). Undoubtedly, the best parameter to assess binding is the dissociation constant (K_d) of the nucleotide from its complex with the enzyme, although the binding measured might not be kinetically relevant in some enzymes if non-productive binding occurs. For instance, a $K_d(NAD^+)$ of 10 μM was determined for the rat ALDH1L1, but the enzyme exhibited a low activity with NAD^+ as coenzyme [15]. Values of K_d (coenzyme) are known for only a few ALDH enzymes, and in even fewer of them these values have been determined for both NAD^+ and $NADP^+$. Alternatively, K_m values have been widely used as a measure of affinity. But in a Bi-Bi Ordered Steady-State mechanism, as the one followed by most ALDH enzymes, particularly the hydrolytic ones [9,16–18], K_m is determined by the rate constant of several steps in the mechanism, not only by the *on* and *off* rate constants of the binding step. Indeed, in some ALDHs for which K_d and K_m values are available there are important differences between them. The most striking reported differences between these two parameters are those of the above mentioned ALDH1L1 enzyme, which exhibits an apparent K_mNAD^+ of 2000 μM and a K_dNAD^+ of only 10 μM [15], and of the ALDH10 from the amaranth plant, which has a K_mNAD^+ of 39 μM and a K_dNAD^+ of 0.16 μM [18,19]. Also $K_m$nucleotide values may vary with the aldehyde substrate used. An example is the ALDH3I1 enzyme from *Arabidopsis thaliana*, which has a K_mNADP^+ of 1868 μM with hexanal as substrate but of 87 μM with nonenal [20]. Moreover, the interpretation of K_m values is often complicated because apparent values obtained at a fixed aldehyde concentration, not always saturating, are frequently reported. True $K_m$nucleotide values, or even better K_{ia} values, are very scarce in the literature of ALDH enzymes. To measure the affinity of the oxidized nucleotide k_{cat}/K_m should be determined, given that this kinetic parameter represents the bimolecular constant rate of the binding step of the first substrate to add to the enzyme in an ordered Bi-Bi mechanism, as the one generally found in ALDHs where the oxidized nucleotide binds first. But for several kinetically characterized enzymes k_{cat}/K_m is not known because V_{max} or k_{cat} values are not always reported even though K_m values are. Besides, the low $(k_{cat}/K_m)NAD(P)^+$ values for many ALDHs suggest that either there is an isomerization step associated with the binding of the coenzyme, or the kinetic mechanism is not ordered. In addition, the comparison of kinetic parameters of related enzymes is often complicated by the use of different assay conditions, mainly pH and temperature. In spite of these limitations, we tried to correlate the reported kinetic parameters with the known three-dimensional structures and/or amino acid sequences of ALDHs in an attempt to recognize the residues that may influence their preference for the coenzyme, or their ability to bind both. From the data available in the literature, we calculated the ratios K_mNAD^+/K_mNADP^+ and $(k_{cat}/K_m)NAD^+/(k_{cat}/K_m)NADP^+$, which were used as an indication of coenzyme preference.

3.3. Sequence analysis

A multiple sequence alignment of 1049 non-redundant ALDH protein sequences was performed to explore the degree of conservation of critical residues that influence coenzyme preference in the different ALDH families. In order to compare these residues among the different ALDH proteins, the sequence of the human ALDH2 mature protein (*Hs*ALDH2) was used as reference. This analysis allowed us to infer the distribution of NAD^+-, $NADP^+$- or dual coenzyme specific enzymes within different ALDH families. Because some characterized ALDHs with a resolved crystal structure and/or known coenzyme preference do not belong to any of the previously reported ALDH families [21–23], the protein family assigned by the Conserved Domain Database [24] was given. We used this database because the protein classification included in it agrees with the previously reported ALDH families and with the results of the phylogenetic analysis performed in this work.

3.4. Structural determinants of coenzyme preference

In Fig. 1 are shown the interactions made by the adenosine ribose of NAD^+ and $NADP^+$ with the ALDH protein, as exemplified