Alcohol dehydrogenases

HANS JORNVALL, BENGT PERSSON, MARIA KROOK and RUDOLF KAISER
Department of Chemistry I, Karolinska Institutet, S-10401 Stockholm, Sweden

Alcohol dehydrogenases constitute enzymes of at least three different protein types. They illustrate general aspects of protein relationships, indicate the presence of repeated duplicatory events, and show the effects of successive structural changes.

Long-chain type

The initially characterized, traditional liver enzyme is a member of the 'long-chain', zinc-containing alcohol dehydrogenase family [1]. An increasing number of forms and subforms have been characterized, representing isoenzymes and different alcohol dehydrogenase classes. Combined, they constitute the complex liver alcohol dehydrogenase system (reviewed in [2]). Further proteins have also been found to belong to this family, including sorbitol dehydrogenase [1], threonine dehydrogenase [3] and an eye lens crystallin [4]. The diversity between these proteins is extensive and the crystallins, although structurally a clear member of the long-chain alcohol dehydrogenase family, does not appear to contain zinc and has not yet been proven even to have a dehydrogenase activity [4].

Short-chain type

The second type of alcohol dehydrogenase family is the 'short-chain' non-metalloenzyme type, initially defined from the structures of insect alcohol dehydrogenase and prokaryotic dehydrogenases of sugar metabolism [1, 5]. This group has subsequently been discovered to be large and is now known to be represented also by mammalian dehydrogenases in the form of 17β-hydroxysteroid dehydrogenase and cytosolic 15-hydroxyprostaglandin dehydrogenase [6, 7].

Additional forms

A third enzyme family of alcohol dehydrogenases was initially detected as a prokaryotic iron-dependent enzyme type [8], now also known to be represented by a yeast enzyme [9] and further forms.

Differences

Alcohol dehydrogenases of these three families differ extensively, exhibiting separate domain arrangements, different enzymatic mechanisms, and, except for the coenzyme-binding regions [5, 10], show little sequence similarity and little further relationships. However, at least one protein member in each of the three families has ethanol dehydrogenase activity, giving the common nomenclature of alcohol dehydrogenase (EC 1.1.1.1). The properties are discussed below in relation to several characteristics for both the long-chain and short-chain dehydrogenase families.

Multiplicity

Gene duplications at different levels are frequent. Thus far, this has been traced in greatest detail for the long-chain zinc dehydrogenases, revealing at least four levels of complexity (Fig. 1), explaining the multiplicity of enzymes, enzyme classes and isoenzymes.

Long-chain family. Ancient duplications explain the presence of different enzymes within the family. They diverged early and have now reached a stage (about 25% residue identity) where they have diverged to such an extent that they are not immediately recognized as similar. In this way, four functionally separate proteins are structurally related, constituting alcohol, sorbitol and threonine dehydrogenases, apart from ζ-crystallin (Fig. 1). These enzymes all have different EC numbers (1.1.1.1 for alcohol dehydrogenase, 1.1.1.14 for sorbitol dehydrogenase, 1.1.1.103 for threonine dehydrogenase). Thus, the end-numbers do not reflect the structural relationships, illustrating the difficulties.

One protein (one gene) ε Isoenzymes ± Classes $ Different enzymes

<table>
<thead>
<tr>
<th>Examples</th>
<th>α-Chain</th>
<th>β-Chain</th>
<th>γ-Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH</td>
<td>II</td>
<td>III</td>
<td>ζ-Crystallin</td>
</tr>
</tbody>
</table>

Residue identities (%)
- 90
- 60
- 25

Fig. 1. Successive development of increasing multiplicity in the long-chain alcohol dehydrogenase enzyme family is created by repeated gene duplications

Each arrow indicates a level of duplications and subsequent accumulation of mutational differences. Presumably, this links the presence of isoenzymes with the emergence of new functions, eventually leading to highly divergent forms. In this manner, isoenzyme development may be looked upon as a natural step in the development of new enzyme functions. Abbreviations: ADH, alcohol dehydrogenase; SDH, sorbitol dehydrogenase; ThrDH, threonine dehydrogenase.
with enzyme nomenclature when structural aspects are considered.

A later level of duplication explains the presence of separate classes within the alcohol dehydrogenase line (Fig. 1). Similarities here are still quite distant and the enzymes have different sub-specificities. Interestingly, recent evidence from species variations within each class indicates that the classes differ in speed of evolutionary divergence [11]. This suggests that the classes represent stages toward the emergence of new enzyme functions (Fig. 1).

The classical liver enzyme belongs to the class I branch [2] and has long been considered typical of the mammalian and vertebrate liver alcohol dehydrogenases. Recent data from the crystallin structure [4] and from sub-mammalian alcohol dehydrogenases (unpublished work) suggest that classes I and II are the most dissimilar, while class III shows some intermediate properties. Consequently, class I and III alcohol dehydrogenases are those of greatest interest in relation to the original metabolic roles of the present-day, typical, liver-type of alcohol dehydrogenase.

More recent gene duplications explain the presence of typical isoenzymes as illustrated by the first arrow set in Fig. 1. In this case, three genes exist. They are still separately controlled as evidenced by non-identical expression in different tissues and at different periods of life. The α-chain of the protein represents the fetal form of alcohol dehydrogenase, while functions (Fig. 1) with separate frequencies of representation in different populations (cf. [2]).

Short-chain family. The different members of this dehydrogenase family also reflect many duplications, now giving rise to well over 10 known different enzymes ([5–7]; unpublished work). Most of these reflect distantly related forms, corresponding to the most ancient level of divergence in Fig. 1. However, the short-chain enzymes also show some more highly similar forms, illustrating that different levels of duplicatory events have occurred in this family as well.

Successive changes

The separate levels of multiplicity demonstrate the effects of successive changes on protein structures.

The long-chain type. Although only one of the structures in Fig. 1 is known from direct crystallographic analysis, a horse class I alcohol dehydrogenase, computer-graphics comparisons suggest that the separate isoenzymes have highly similar tertiary structures [12]. To a considerable extent, this may also apply to the separate classes and enzymes ([13]; unpublished work). Within this basic fold, it is possible to judge the effects of the various changes. Notably, at all multiplicity levels, changes appear concentrated in three segments of the molecule, one affecting the active site, one the area of subunit interactions and one the segment around the second zinc atom (cf. [2]). These three areas appear largely to reflect the functional differences between the alcohol dehydrogenases. Thus, the first type of variation (at the active site) presumably explains the different enzyme specificities, and the second type (subunit interactions) presumably reflects the fact that several of the alcohol/polyol dehydrogenases have non-identical quaternary structures and do not cross-hybridize (except the true isoenzymes). However, the third variation (around the second zinc atom) has thus far unknown functional consequences.

The latter variation makes the long-chain dehydrogenases a group of proteins well suited for the study of successive changes in metal interactions. As shown in Table 1, the separate enzymes and isoenzymes illustrate successive changes in both metal content and metal ligands. Few other protein systems appear to reveal clearly demonstrated metal differences of this type.

In a similar manner, the residue exchanges at the other variable segments allow direct correlation with substrate interactions, explaining the small differences in functional properties of the isoenzymes [12], and have also been interpreted in terms of the separate substrate specificities of alcohol and sorbitol dehydrogenases [13].

Thus, characterization of all natural variants illustrates the relative importance of certain segments, defines strictly conserved residues [14], and suggests positions suitable for further studies by site-directed mutagenesis. In fact, many of the functional correlations thus far have first been traced by correlations with natural variants and are now successively accessible to confirmation and extended studies by selective mutants from directed alterations. In this manner, natural variability and directed mutagenesis can complement each other in a productive manner to arrive at still further functional conclusions.

The short-chain type. Successive changes in properties and direct correlations are thus far more difficult to trace for this family than for the long-chain family. This is because the short-chain forms presently known still represent highly divergent forms, and because no tertiary structure is known for any of these enzymes. However, both of these difficulties will probably soon be eliminated by discovery of further, interconnecting forms and by crystallizations of these short-chain dehydrogenase proteins. In particular, a Drosophila alcohol dehydrogenase would appear essential to crystallize, and an apparently successful crystallization of such an enzyme was one reason why the primary structure of an additional form of the protein was recently determined [15].

Future aspects

Two aspects of alcohol dehydrogenase are of particular interest. One is the functional role and biological activities. The other is the mechanistic means by which the protein structures fulfil these functions. The knowledge of both aspects has recently increased enormously. The value of studies of natural variants is that unexpected relationships have been discovered, such as, for example, those between eye lens ζ-crystallin and liver alcohol dehydrogenases [4]. Presumably, additional studies of still further proteins will complement existing data and gradually fill in the interconnecting gaps. At the same time, the functional correla-

Table 1. Successive changes in metal-binding properties within the long-chain zinc-dehydrogenase family

<table>
<thead>
<tr>
<th>Protein</th>
<th>Non-catalytic Zn atom</th>
<th>Ligands present</th>
<th>Segment present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol dehydrogenase</td>
<td>2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sorbitol dehydrogenase</td>
<td>1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ζ-Crystallin</td>
<td>0†</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

*+ and – signs refer to the presence and absence, respectively, of the four Cys ligands and the protein segment corresponding to the loop that binds the non-catalytic zinc atom in alcohol dehydrogenases.
†Zinc content not analytically determined, but concluded from structural properties to be absent.
ISOENZYMES

tions make it possible to ascribe specific roles to particular residues, such as the conserved glycine residues [14] at bends in the basic fold which appear to constitute alcohol dehydrogenase per se. Each of these concepts can now be further tested by directed mutagenesis.

Regarding the enzymic mechanisms, the basic properties of the zinc dehydrogenases have essentially been known for some time. However, the short-chain dehydrogenase type is largely unknown in mechanistic terms and studies of additional forms, complemented with extensive comparisons, should be valuable also for the short-chain dehydrogenase group, to define more constant segments and particularly conserved residues which can then be directly tested for in mechanistic and functional terms. For example, a segment around position 150 in the short-chain dehydrogenases appears likely to be close to the active site and has a few conserved residues of particular interest [6]. It is to be expected that further studies of natural variants will increase the reliability in the interpretational conclusions.

Support by grants from the Swedish Medical Research Council (project 03X-3532) and the Swedish Alcohol Research Fund is gratefully acknowledged.


Received 10 October 1989

Structure and expression of mammalian carbonic anhydrases

YVONNE EDWARDS

M.R.C. Human Biochemical Genes Unit, The Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London NWI 2HE, U.K.

Carbonic anhydrase (CA), is a ubiquitous enzyme involved in the transport of CO₂ between metabolizing tissue and the lungs, in many secretory processes, in ion transport and in the provision of bicarbonate for fatty acid synthesis, gluconoegenesis and urea-genesis (see [1] for a review and references). Despite the diversity of its metabolic involvements, the reaction that this enzyme catalyses is a simple one, namely the transfer of an OH⁻ moiety between H₂O and CO₂:

\[ \text{CO}_2 + 
\]

It was in the 1960s, with the advent of chromatographic techniques, that the first isoforms of CA were separated from erythrocytes, namely CAI which is present in high amounts, but with low activity, and CAII which is present in a smaller quantity, but with high specific activity. In higher vertebrates at least seven genetically distinct, but structurally similar, isoenzymes of CA are now recognized, each showing characteristic kinetic properties and tissue distribution (see [2] for a review).

Three of the mammalian CA genes are expressed predominantly in one cell type, cytoplasmic CAI and CAII in erythrocytes and skeletal muscle cells, respectively, and the secreted form, CAV, in parotid salivary gland. These genes show a greater degree of tissue specificity than the more ubiquitously expressed cytoplasmic CAII or the membrane-associated CAIV and mitochondrial CAV. It is not clear if any particular cell type expresses a single CA gene — most express two or more. So, for example, erythrocytes and skeletal muscle slow-twitch fibres each express CAI, CAII and CAIII, although the relative levels of expression of the three genes vary considerably between the two tissues with CAI > CAII > CAIII for erythrocytes and CAII > CAIII > CAI for slow-twitch fibres.

The CA isoforms show between 28% and 59% amino acid sequence similarity when pairwise comparisons are made, with CAI and CAII showing greatest similarity. Recently, Tashian [2] has used consensus amino acid sequences to construct a comparative matrix and a phylogenetic branching scheme for the CA isoforms. These data indicate an order of evolutionary relatedness: CAI, CAII, CAIII, CAVII, CAVI and CAIV and suggest that the salivary CAI gene diverged first from the ancestral root. This pattern of genetic ancestry is reflected in the chromosomal distribution of the mammalian CA genes. CAI, CAII and CAIII are clustered on the long arm of human chromosome 8 at q22 [3, 4], and are similarly clustered on mouse chromosomes near to the centromere [5, 6], while the more distantly related CAVI and CAVII have been mapped to human chromosomes 1 and 16, respectively [7, 8].

Why is a complex family of several CA genes required to encode a monomeric enzyme which catalyses a very simple reaction? Comparisons of these genes and their protein products suggest that the diversity is important both in the context of gene expression and protein function. Variation in those parts of the DNA which encode the CA confers characteristic kinetic properties on each of the isoforms and also provides the protein sequences necessary for particular subcellular localizations or secretion of the gene product, while variation in the non-coding regions provides for flexibility of expression among many different cell types.

Protein structure

Fig. 1 shows the 36 active-site residues for six CA isoforms, 17 of these are common to all CAs, while others are characteristic of particular isoforms. Insights into the role of such isoform-specific residues in active-site mechanisms are beginning to emerge from site-directed mutagenesis studies. For example, CAIII shows eight unique residues in all five