CMLS, Cell. Mol. Life Sci. 59 (2002) 1070–1075
1420-682X/02/061070-06 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

Differential multiplicity of MDR alcohol dehydrogenases: enzyme genes in the human genome versus those in organisms initially studied

E. Nordling[a], B. Persson[a,b] and H. Jörnvall[a,*]

[a] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm (Sweden), Fax + 46 8 337462, e-mail: Hans.Jornvall@mbb.ki.se
[b] Stockholm Bioinformatics Centre, Karolinska Institutet, 171 77 Stockholm (Sweden)

Received 3 May 2002; accepted 3 May 2002

Abstract. Screens were made for alcohol dehydrogenase (ADH) of the classical type (the MDR superfamily) in translations of human and other relevant genomes, corresponding to the organism types from which the enzyme was initially purified. Considerable multiplicities were detected in the dimeric enzymes from higher eukaryotes: seven forms in the human (plus three pseudogenes), all genes on chromosome 4, in the order class IV → class Iγ → class Iβ → class Iα → class V → class II → class III, and eight forms in *Arabidopsis thaliana* (plus one pseudogene). These multiplicity patterns, and the species variability in the animal (human/mouse) and plant (*Arabidopsis*/pea) lines, suggest parallel but separate duplicatory events, giving rise to three families of dimeric MDR-ADH: class III, the animal non-class III, and the plant non-class III enzymes, with functions in formaldehyde elimination, in alcohol/aldehyde detoxication and in special pathways in higher eukaryotes. Multiplicity, although to a lesser extent, was also noted in tetrameric MDR-ADH, suggesting functional divergence between the dimeric and tetrameric enzymes. Combining these observations, at least five levels of divergence are reflected in the present ADH forms, corresponding to nodes at the SDR/MDR, the dimer/tetramer, the class III/non-class III, the class I/P, and the more recent class splits, each branch associated with separate functional patterns.

Key words. Alcohol dehydrogenase; MDR; SDR; enzyme multiplicity; genome; duplicatory event; formaldehyde elimination; alcohol/aldehyde detoxication; dimer/tetramer divergence.

Introduction

Alcohol dehydrogenases (ADHs) were detected early and have the lowest EC number, EC 1.1.1.1. Four types, in particular, were initially studied and long served as different models and markers [1]. They were yeast ADH (the largest), crystallized as early as 1937 [2]; *Drosophila* ADH (the smallest), purified in 1968 [3], liver ADH, and plant ADH (both intermediate), crystallized or purified in 1948 and 1968, respectively [4, 5]. Later, all four types were found to occur in multiple forms [1], and later still to represent different evolutionary lines [6], although they are now known to be related in a more diverse manner than initially visualized.

With the completion of the sequencing of the human and other genomes, the genetic representation of these four ADH types has been established, allowing comparisons not only of the ADHs, but of the entire enzyme systems to which they belong, i.e., MDR (medium-chain dehydrogenases/reductases) [Nordling E., Jörnvall H. and Persson B., submitted] and SDR (short-chain dehydrogenases/reductases) [7]. Such overall comparisons have revealed at least eight different MDR families and at least five different SDR families. In the human genome, they

* Corresponding author.

cover, minimally, 23 MDR and 63 SDR genes [Nordling E., Jörnvall H. and Persson B., submitted]. Notably, additional families with ADH activity also occur, such as iron-dependent ADH [8], whose gene has been identified in the human genome, long bacterial ADH [9], and AKR (aldo-keto reductases) ADH [10].

Recent comparisons [Nordling E., Jörnvall H. and Persson B., submitted] have concered group distinctions rather than direct correlations within each enzyme activity. In the present work, therefore, we centered on the classical ADH activity in relation to the four ADH types originally defined (as above) and the human genome. The results show differential multiplicities and allow functional conclusions.

Materials and methods

Dimeric and tetrameric ADH members were identified from protein sequences translated from the genomes of *Homo sapiens* [11, 12], *Drosophila melanogaster* [13], *Caenorhabditis elegans* [14], *Arabidopsis thaliana* [15], *Saccharomyces cerevisiae* [16], and *Escherichia coli* [17] using FASTA3 [18] with previously identified members of these groups [19] as query sequences. Hits with an expect value below 10^{-10} were extracted and aligned with ClustalW [20]. ADH data were also collected from SwissProt for pea and mouse [21]. Evolutionary trees were calculated from the alignments using the neighbor-joining technique implemented in ClustalW. The certainty of each branch point was assessed with bootstrap tests of the different trees. The trees with comparisons of mammalian and plant ADHs were rooted using the prokaryotic *Haemophilus influenzae* ADH (ADH3_HAEIN) as outgroup. The resulting evolutionary trees are displayed with TreeView [22]. All genomic sequences have been checked versus the SwissProt database [21] for functional annotation, and the SwissProt identifier is given when known.

Results and discussion

Overall properties

Screens were made for the originally purified ADH enzyme types (yeast, *Drosophila*, liver, and plant) in the genomes of human, *A. thaliana, S. cerevisiae*, and *D. melanogaster*, and were compared, where relevant, to corresponding forms in mouse, pea, *E. coli*, and *C. elegans*. The screenings revealed the presence of ten human, ten *Arabidopsis*, five yeast, and one *Drosophila* MDR-ADH genes. Of the human genes, three had properties like pseudogenes (recognized by unusual exon/intron patterns, combined with a lack of upstream promoter elements such as TATA, CAAT or GC boxes, and with deviating chromosome localizations), while the remaining seven genes correspond to the enzymes of classes I (three genes) and II–V (one each) [23, 24]. The seven genes are all on chromosome 4 (fig. 1), in the order class IV → class Iγ → class Iβ → class Iα → class V → class II → class III, compatible with early assignments [25]. This total number suggests that additional classes, sometimes named in other species [cf. 23, 24], do not occur in humans and may in some cases be just species variants of human orthologues. The screenings, of course, also revealed several properties already known for ADH enzymes. First, that *Drosophila* ADH is different from all other ADHs, illustrating the fact that they belong to different superfamilies [26], SDR versus MDR, respectively. Furthermore *Drosophila* also has an MDR form, the glutathione-dependent formaldehyde dehydrogenase [27] (equivalent to ADH class III, but with low ethanol activity). However, of these organisms, only *Drosophila* has an SDR-ADH, demonstrating that the insect line and the other lines handle alcohol metabolism by different enzyme forms. Since this study concerns MDR enzymes, *Drosophila* ADH will not be further considered here, except for the class III MDR-ADH type.

The screenings also showed that ethanol-active MDR-ADH, when present, appears to occur multiply. The multiplicity [6] and the general occurrence of ethanol activity [28] have been noted before, but can now be correlated with enzyme type and with eukaryotic organism type. These and other functional conclusions are most clearly revealed when the MDR-ADH enzymes are divided into the two constituent families of dimeric and tetrameric ADHs, initially represented by the liver and yeast ADHs, respectively.

Dimeric ADHs

Deduction of relationships for the dimeric ethanol-active ADHs from evolutionary trees constructed with the ClustalW method reveals divergence into at least three

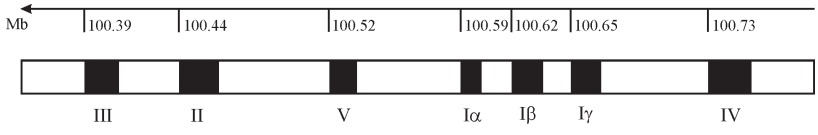

Figure 1. Genomic representation of ADH classes on human chromosome 4 [12]. All genes are on the lagging strand, are displayed as a continuous sequence without indications for the introns, and with start points given in megabases (Mb), numbered according to the leading strand.

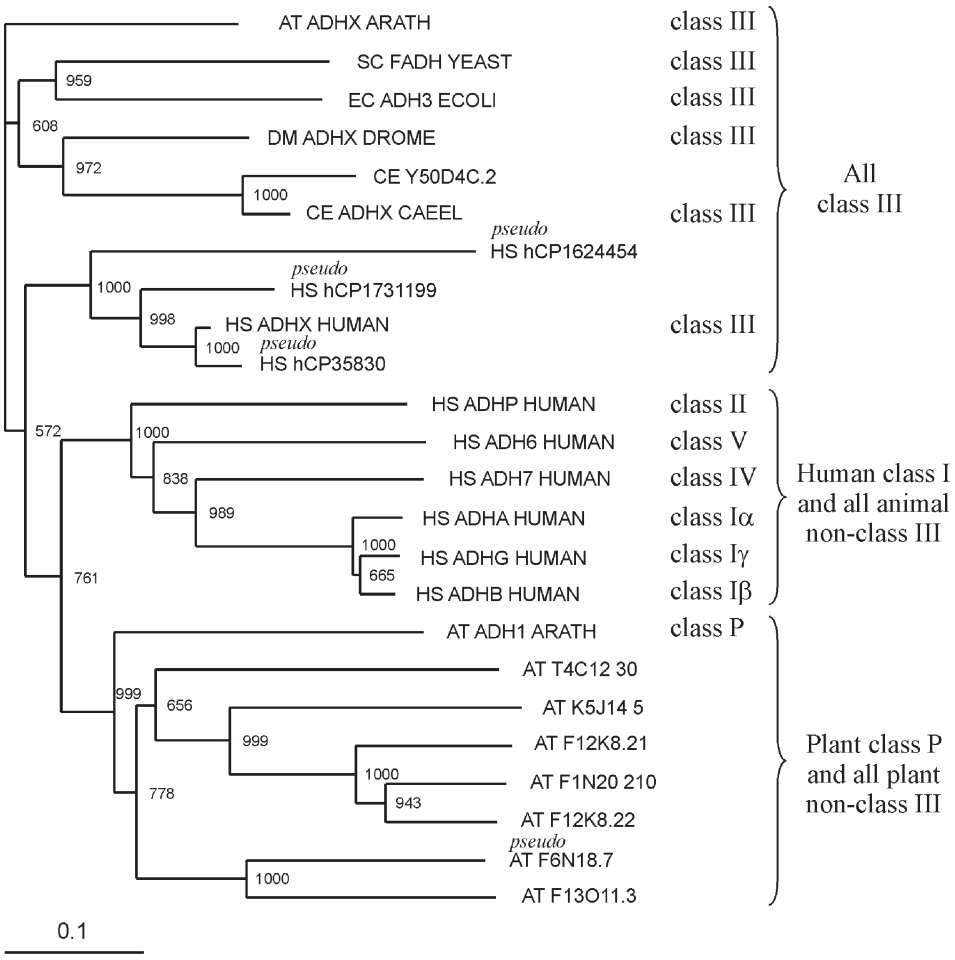

Figure 2. Structural relationships reflected by evolutionary trees for dimeric ADH of the MDR type as deduced from the completed genomes studied. Bootstrap values are given for all nodes, and line lengths represent calculated structural divergence. Designations immediately to the right of all lines show the species (AT, *A. thaliana*; HS, *H. sapiens*; EC, *E. coli*; DM, *D. melanogaster*; CE, *C. elegans*; SC, *S. cerevisiae*) and the SwissProt identifier or the genomic identifier [21]. The word *pseudo* immediately above the code designates pseudogenes. Designations to the left of the brackets show commonly used class names [23, 24], and those to the right, the three groups now discerned.

groups (fig. 2), corresponding to the formaldehyde-active class III enzyme from all species, the ethanol-active non-class III enzymes from animals, and the ethanol-active non-class III enzymes from plants. This tri-partite grouping of dimeric MDR-ADH suggests several functional conclusions regarding these enzymes. One is that only plants and vertebrates appear to have the ethanol-active dimeric MDR-ADHs. Another is that in both the vertebrate and plant groups of these enzymes, the multiplicity is extensive, derived from a number of gene duplications. And a third, that the patterns in the non-class III groups from vertebrates and plants show some similarities, with both early and late branching frequently at similar levels (fig. 2). To illustrate this similarity in patterns from the animal and plant lines of MDR-ADH, we also included species variants in each case (fig. 3), in a manner similar to that previously undertaken [29] but now comparing mouse to the human form, and with time estimates for the branch points. We then see a largely symmetrical pattern between the human/mouse and the *Arabidopsis*/pea variability regarding both the formaldehyde-active and the ethanol-active lines (fig. 3). This supports a common function for non-class III ADHs in higher eukaryotes. However, the time estimates, calculated in two different manners, show that the species separation values for class III are too distant, falsely suggesting that the class III species separations are more distant than those for classes I and P (fig. 3) and also more distant than accepted radiation nodes of higher vertebrates. Hence, the constant speed expected for class III evolution may be slightly higher than previously calculated [30], or class III may evolve faster in higher eukaryotes, perhaps giving them novel enzymogenesis and additional functions, as already known for class I with its recent isozyme emergence [23, 24].

Although ClustalW does not sensitively distinguish distant relationships, and although some of the branchings

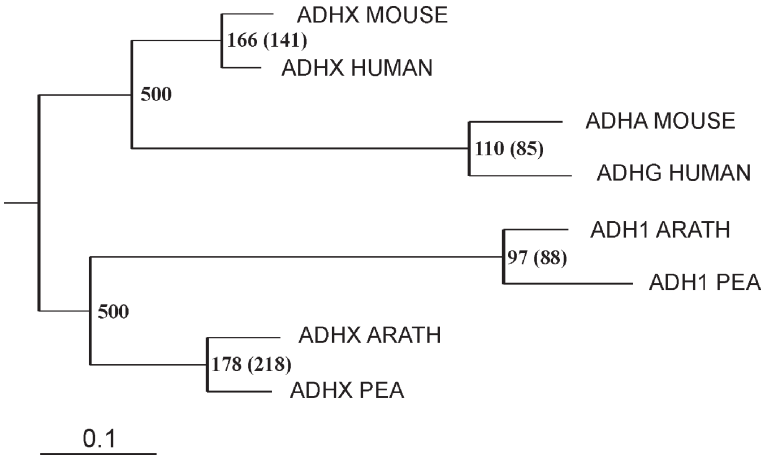

Figure 3. Comparison of mammalian (human/mouse) and plant (*Arabidopsis*/pea) ADH types. SwissProt identifiers are given [21]. Trees are rooted using the prokaryotic *H. influenzae* ADH (ADH3_ HAEIN) as outgroup. Estimated separation times are given at each node in million years ago (MYA) assessed by branch lengths and the known separation time of class I and class III ADH [30]. For comparison, separation times are also given (within parentheses) from the number of amino acid differences and the known mutation rates of the different enzyme lines [30].

have low bootstrap numbers, the overall similarities in patterns appear significant. If so, these patterns must reflect functional relationships in the corresponding organisms. The general conclusion would then be that most organisms have evolved an MDR-ADH and that in higher eukaryotes, this is dimeric MDR-ADH, which has then frequently undergone considerable multiplication: to seven coding genes in the human genome and to eight in the *Arabidopsis* genome (fig. 2). This multiplicity was previously concluded to suggest a protective function for ADH towards toxic effects of small alcohols and aldehydes [6], much like the function of cytochrome P450, also a multiply occurring enzyme. The added advantage with ADH would then be that it participates in dehydrogenations/reductions without generation of reactive oxygen species. Such a protective function still seems possible, but with the added conclusion that ADH would now also be expected to have more specific functions for higher vertebrates, since they have additional MDR-ADH forms. The lack in lower eukaryotes of MDR-ADH activity was previously explained by a marine life [28], i.e., excretion of toxic alcohols/aldehydes into seawater rather than elimination by an ADH enzyme reaction. The present pattern, however, makes the distinction between higher and lower eukaryotes in general rather than between terrestrial and marine life forms, and with partly parallel patterns in plants and animals may suggest that a specific function is associated with ADH in higher eukaryotes. If so, dimeric ADH may have key functions in higher eukaryotes rather than just general detoxifications.

Tetrameric ADHs

A similar visualization of the tetrameric family shows considerably less multiplicity and places yeast ADH with

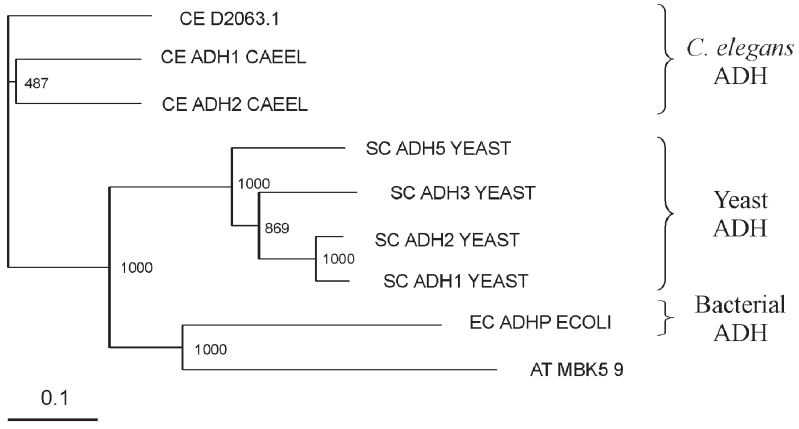

Figure 4. Evolutionary tree of tetrameric ADH of the MDR type within the completed genomes studied. Designations are SwissProt or genomic identifiers as in fig. 2.

ADH of *C. elegans*, and with some *E. coli* and *A. thaliana* forms. Thus, *C. elegans* and *A. thaliana* have both dimeric and tetrameric ADH enzymes. Of special interest is the occurrence of an *E. coli* ADH in this group (fig. 4). This is a bacterial ADH with activity toward ethanol [31]. It has recently been crystallized and its three-dimensional structure solved [32]. Yeast ADH was the first enzyme in this group to be crystallized [2], but its three-dimensional structure has still not been solved. The finding of an *E. coli* ADH in this group, and the fact that it is internally heterogeneous [32], may indicate special relationships for the tetrameric ADH structures.

ADH diversification

Having both the dimeric (figs. 2, 3) and tetrameric (fig. 4) patterns, and the timing at branch nodes [cf. ref. 30], figure 5 summarizes the SDR/MDR-ADH relationships as deduced from the genomes investigated. In particular, it distinguishes the tetramer/dimer duplication from the III/non-III duplication, and shows the presence of the ethanol-type activity in minimally five separate lines: SDR, tetrameric MDR, dimeric animal MDR, dimeric plant MDR, and formaldehyde dehydrogenase. ADH activity of the type that includes ethanol as substrate is concluded to be important and multiply divergent in a parallel manner in different higher eukaryotes (figs. 2, 3). This activity is fulfilled by dimeric ADH in higher eukaryotes, by tetrameric ADH, also multiple, although less divergently so in lower eukaryotes and prokaryotes (fig. 4), and is possibly absent in a set of intervening organisms [28]. Patterns are similar, whether derived from just ADH (this work) or all types of MDR [Nordling E., Jörnvall H. and Persson B., submitted].

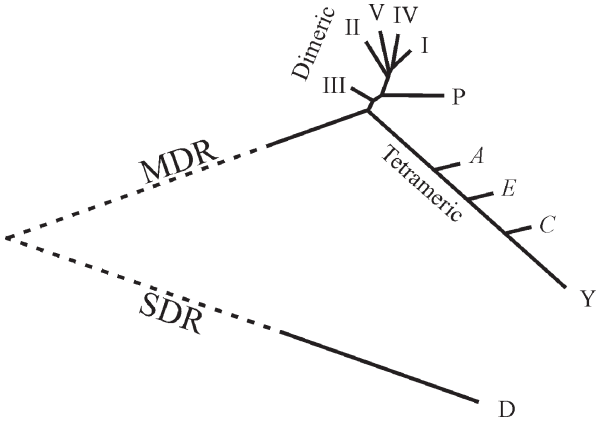

Figure 5. Schematic trace of ADH multiplicity in the forms now treated. Solid line lengths reflect the extent of divergence; dashed lines indicate non-characterized connections between the MDR and SDR superfamilies. In addition, a gene apparently corresponding to iron-dependent ADH is present in the human, as are AKR enzymes, one of which may have evolved ADH activity [cf. ref. 10]. I–V, the human ADH classes; P, plant ADH (class P form; all further plant ADH branches in fig. 2); Y, yeast ADH (type 1; all further yeast ADH branches in fig. 4); *A*, *E*, and *C* indicate the tetrameric ADH enzymes from *A. thaliana*, *E. coli* and *C. elegans*, respectively (with further tetrameric ADH branches in fig. 4).

Acknowledgements. Support by the Swedish Research Council, the Swedish Foundation for Strategic Research, and Karolinska Institutet is gratefully acknowledged.

1 Brändén C.-I., Jörnvall H., Eklund H. and Furugren B. (1975) The Enzymes, vol. 11, 3rd edn., pp. 103–190, Academic Press, New York
2 Negelein E. and Wulff H.-J. (1937) Kristallisation des Proteins der Acetaldehydreduktase. Biochem. Z. **289:** 436–437
3 Sofer W. and Ursprung H. (1968) *Drosophila* alcohol dehydrogenase. J. Biol. Chem. **243:** 3110–3115
4 Bonnichsen R. K. and Wassén A. M. (1948) Crystalline alcohol dehydrogenase from horse liver. Arch. Biochem. Biophys. **18:** 361–363
5 Pattee H. E. and Swaisgood H. E. (1968) Peanut alcohol dehydrogenase. I. Isolation and purification. J. Food Sci. **33:** 250–253
6 Jörnvall H., Höög J.-O., Persson B. and Parés X. (2000) Pharmacogenetics of the alcohol dehydrogenase system. Pharmacology **61:** 184–191
7 Kallberg Y., Oppermann U., Jörnvall H. and Persson B. (2002) Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, and plant genomes. Protein Sci. **11:** 636–641
8 Williamson V. M. and Paquin C. E. (1987) Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. Mol. Gen. Genet. **209:** 374–381
9 Inoue T., Sunagawa M., Mori A., Imai C., Fukuda M., Takagi M. et al. (1989) Cloning and sequencing of the gene encoding the 72-kilodalton dehydrogenase subunit of alcohol dehydrogenase from *Acetobacter aceti*. J. Bacteriol. **171:** 3115–3122
10 Crosas B., Cederlund E., Torres D., Jörnvall H., Farrés J. and Parés X. (2000). A vertebrate aldo-keto reductase active with retinoids and ethanol. J. Biol. Chem. **276:** 19132–19140
11 Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Smith H. O. et al. (2001) The sequence of the human genome. Science **291:** 1304–1351
12 Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J. et al. (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860–921
13 Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G. et al. (2000) The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195
14 The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018
15 The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815
16 Goffeau A., Barrell B. G., Bussey H., Davis R. W., Dujon B., Feldmann H. et al. (1996) Life with 6000 genes. Science **274:** 563–567
17 Blattner F. R., Plunkett G. III, Bloch C. A., Perna N. T., Burland V., Riley M. et al. (1997) The complete genome sequence of *Escherichia coli* K-12. Science **277:** 1453–1474
18 Pearson W. R. and Lipman D. J. (1988) Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:** 2444–2448
19 Persson B., Zigler J. S. Jr and Jörnvall H. (1994) A super-family of medium-chain dehydrogenases/reductases (MDR): sublines including ζ-crystallin, alcohol and polyol dehydroge-

nases, quinone oxidoreductase, enoyl reductases, VAT-1 and other proteins. Eur. J. Biochem. **226:** 15–22

20 Thompson J. D., Higgins D. G. and Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–680

21 Bairoch A. and Apweiler R. (2000) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 2000. Nucleic Acids Res. **28:** 45–48

22 Page R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:** 357–358

23 Jörnvall H. and Höög J.-O. (1995) Nomenclature of alcohol dehydrogenases. Alcohol Alcoholism **30:** 153–161

24 Duester G., Farrés J., Felder V. C., Höög J.-O., Parés X., Plapp B. et al. (1999) Alcohol dehydrogenase nomenclature. Biochem. Pharmacol. **58:** 389–395

25 Duester G., Hatfield G. W. and Smith M. (1985) Molecular genetic analysis of human alcohol dehydrogenase. Alcohol **2:** 53–56

26 Jörnvall H., Persson M. and Jeffery J. (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc. Natl. Acad. Sci. USA **78:** 4226–4230

27 Luque T., Atrian S., Danielsson O., Jörnvall H. and Gonzàlez-Duarte R. (1994) Structure of the *Drosophila melanogaster* glutathione-dependent formaldehyde dehydrogenase/octanol dehydrogenase gene (class III alcohol dehydrogenase): evolutionary pathway of the alcohol dehydrogenase genes. Eur. J. Biochem. **225:** 985–993

28 Fernández M. R., Biosca J. A., Norin A., Jörnvall H. and Parés X. (1995) Class III alcohol dehydrogenase from *Saccharomyces cerevisiae*: structural and enzymatic features differ toward the human/mammalian forms in a manner consistent with functional needs in formaldehyde detoxication. FEBS Lett. **370:** 23–26

29 Shafqat J., El-Ahmad M., Danielsson O., Martínez M.C., Persson B., Parés X. et al. (1996) Pea formaldehyde-active class III alcohol dehydrogenase: common derivation of the plant and animal forms but not of the corresponding ethanol-active forms (classes I and P). Proc. Natl. Acad. Sci. USA **93:** 5595–5599

30 Cañestro C., Albalat R., Hjelmqvist L., Godoy L., Jörnvall H. and Gonzàlez-Duarte R. (2002). Ascidian and amphioxus *Adh* genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. J. Mol. Evol. **54:** 81–89

31 Shafqat J., Höög J.-O., Hjelmqvist L., Oppermann U. C. T., Ibanez C. and Jörnvall H. (1999) An ethanol-inducible MDR ethanol dehydrogenase/acetaldehyde reductase in *Escherichia coli*: structural and enzymatic relationships to the eukaryotic protein forms. Eur. J. Biochem. **263:** 305–311

32 Eklund H., Karlsson A., El-Ahmad M., Johansson K., Shafqat J., Jörnvall H. et al. (in press) Two types of zinc-coordination in tetrameric NAD-dependent alcohol dehydrogenase. Proc. Enzymol. Mol. Biol. Carbonyl Metab. 9