SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature

Hans Jörnvall*, Jan-Olov Höög, Bengt Persson

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden

Abstract Short-chain dehydrogenases/reductases (SDR) and medium-chain dehydrogenases/reductases (MDR) are protein families originally distinguished from characterisations of alcohol dehydrogenase of these two types. Screening of completed genome sequences now reveals that both these families are large, wide-spread and complex. In *Escherichia coli* alone, there are no fewer than 17 MDR forms, identified as open reading frames, considerably extending previously known MDR relationships in prokaryotes and including ethanol-active alcohol dehydrogenase. In entire databanks, 1056 SDR and 537 MDR forms are currently known, extending the multiplicity further. Complexity is also large, with several enzyme activity types, subgroups and evolutionary patterns. Repeated duplications can be traced for the alcohol dehydrogenases, with independent enzymogenesis of ethanol activity, showing a general importance of this enzyme activity.

© 1999 Federation of European Biochemical Societies.

Key words: Short-chain dehydrogenase/reductase; Medium-chain dehydrogenase/reductase; Protein family; Genome screening; Duplication; Enzyme multiplicity; Alcohol dehydrogenase

1. Introduction

Knowledge of short-chain dehydrogenase/reductase (SDR, with subunits typically of 250-odd residues) and medium-chain dehydrogenase/reductase (MDR, with subunits typically of 350-odd residues) enzymes as two separate protein families initially emerged from observations on alcohol dehydrogenases. These observations showed the *Drosophila* and mammalian liver proteins to be clearly different [1,2], although the mammalian and yeast proteins were related [3,4]. This alcohol dehydrogenase split into two types was soon joined by corresponding knowledge of similarly split polyol dehydrogenases [5], thus establishing the SDR and MDR families as containing activities in common but working with different mechanisms and structures. For a long time, both families were regarded as special, with SDR forms known only from prokaryotes and insects [5], and with MDR forms representing only enzymes with active-site zinc in a widely spaced ligand pattern ([6], cf. [7]). However, at the start of the present decade, both these restrictions were removed: characterisation of human 15-hydroxyprostaglandin dehydrogenase showed it to be an SDR form [8], establishing SDR also in vertebrates; and characterisation of ζ-crystallin showed it to be an MDR lacking typical ligands to a catalytic zinc, hence without that zinc arrangement at the active site [9]. Soon, an 'explo-sion' of further characterisations extended the families, defined the general characteristics and showed both SDR [10] and MDR [11] to be of wide occurrence and to represent many activities.

The SDR enzymes have an N-terminal coenzyme-binding pattern [12] of typically GXXXGXG, and an active-site pattern of YXXXK. The SDR family is highly divergent with a typical pairwise residue identity of 15–30%. The enzymes cover a wide range of substrate specificity, including steroids, alcohols and aromatic compounds. There also exist SDR forms with 350-odd residue subunits that are distantly related, exhibiting dehydrogenase, dehydratase, epimerase or isomerase activity [10]. Tertiary structures of SDR forms [13] confirmed the distinct family assignment and showed SDR typically to be one-domain proteins with other subunit interactions than the two-domain MDR proteins. Still, however, the MDR form was considered somewhat special, with several proteins successively formed recently, and hence a different evolutionary pattern from that of the apparently more distantly related SDR forms [14]. The completed genome sequences now available allow further distinctions of the two families, establishing both SDR and MDR as ancient and of surprisingly common occurrence. In addition, they are both complex in nature and represent many activities and properties.

2. Materials and methods

All open reading frames (ORFs) of the 18 completed genomes available [15] were compared with FASTA3 [16] against the SwissProt database (release 36, July 1998) [17]. The top matches were analysed for the occurrence of known MDR or SDR forms. MDR forms were extracted searching for SwissProt ID containing ADH, DHSO or QOR. Classical SDR forms were extracted searching for sequences having the ADH_SHORT pattern in Prosite [18]. Extended SDR forms were identified using the 22 sequences known in 1995 [10] and later sequences clearly related as detected by running FASTA3 with each of these 22 sequences against SwissProt.

ORFs with an expect value of less than 10^{-5} were aligned using ClustalW [19]. Alignments were screened for conserved sequence motifs typical of the superfamilies. Sequence motifs used for SDR were the N-terminal $TGX_{2-3}GXG$ pattern, and the mid-chain patterns NNAG and YXXXK [10], and for MDR the GHE pattern 60-odd residues from the N-terminus, and the mid-chain $GX_{1-3}GX_{1-3}G$ pattern [11]. Single exchanges were also allowed for, provided remaining overall patterns were conclusive.

3. Results and discussion

3.1. Abundance of MDR and SDR enzymes corresponding to ORFs of completed genome sequences

Eighteen completed genome sequences now available were searched for the SDR and MDR family motifs as given in Section 2. Results are given in Table 1. It is obvious that both

*Corresponding author. Fax: (46) (8) 337 462.
E-mail: hans.jornvall@mbb.ki.se

Table 1
Number of SDR and MDR members of the 18 hitherto completed genomes

Organism	Number of ORFs	SDR, classical	SDR, extended	MDR
Archaea				
Methanococcus jannaschii	1733	0	2	0
Methanobacterium thermoautotrophicum	1868	0	6	0
Pyrococcus horikoshii	2061	0	4	1
Archaeoglobus fulgidus	2407	1	4	1
Bacteria				
Mycoplasma genitalium	468	0	1	0
Mycoplasma pneumoniae	676	0	1	1
Rickettsia prowazekii	837	2	0	0
Borrelia burgdorferi	850	0	1	0
Chlamydia trachomatis	894	1	0	0
Treponema pallidum	1030	0[a]	0	0
Aquifex aeolicus	1522	2	4	1
Helicobacter pylori	1577	3	5	1
Haemophilus influenzae	1678	3	4	2
Synechocystis sp.	3158	8	11	2
Mycobacterium tuberculosis	3924	46	11	14
Bacillus subtilis	4100	28	4	10
Escherichia coli	4273	17	7	17
Eukaryota				
Saccharomyces cerevisiae	6266	14	6	17

[a]In *Treponema pallidum*, no extended SDR family member was found with our cut-off criteria, but a capsular polysaccharide biosynthesis protein is present, distantly related to the epimerases of the extended SDR family.

SDR and MDR sequences are common. The majority of the MDR forms are hypothetical zinc-binding forms also in prokaryotes. However, absolute numbers vary considerably, and some organisms, such as *Mycobacterium tuberculosis*, have especially large numbers of these enzymes, in particular of SDR (Table 1). It is further seen that organisms with the smallest numbers of ORFs typically only have SDR forms of the extended type, i.e. SDR forms with 350-odd residue subunits, and those that include dehydratases, epimerases and isomerases, rather than just oxidoreductases [10]. The presence of these extended SDR forms in organisms lacking classical SDR and MDR forms (Table 1) may suggest that even a non-oxidoreductase role could constitute an original function of the whole family. Alternatively, and perhaps more likely, the oxidoreductase function is original also in the extended SDR forms but has later emerged to further functions.

Genes encoding SDR and MDR enzymes were found in all genomes with more than 2061 ORFs (Fig. 1). The only organism where it appears to be absent using present screening models is *Treponema pallidum*. However, further analysis of structures in this organism reveals the presence of TP0077, a 538-residue capsular polysaccharide biosynthesis protein (cap5D), the C-terminal two thirds of which are distantly related to the extended SDR forms. Combined, all results establish that not only SDR, in agreement with initial results when the family was defined [5], but also MDR is of common occurrence in prokaryotes. Furthermore, the screening results reveal that the enzymes identified correspond to those in higher organisms. An MDR form deduced from the *Escherichia coli* genome corresponds to sorbitol dehydrogenase, and yet another to the MDR-type of ζ-crystallin [11] described as quinone oxidoreductase in *E. coli* [20]. These results identify oxidoreductases familiar from eukaryotes as common also in prokaryotes. Although perhaps expected, this has not been clear from the literature. For example, ethanol dehydrogenase activity, previously shown in *E. coli* [21], has not often been
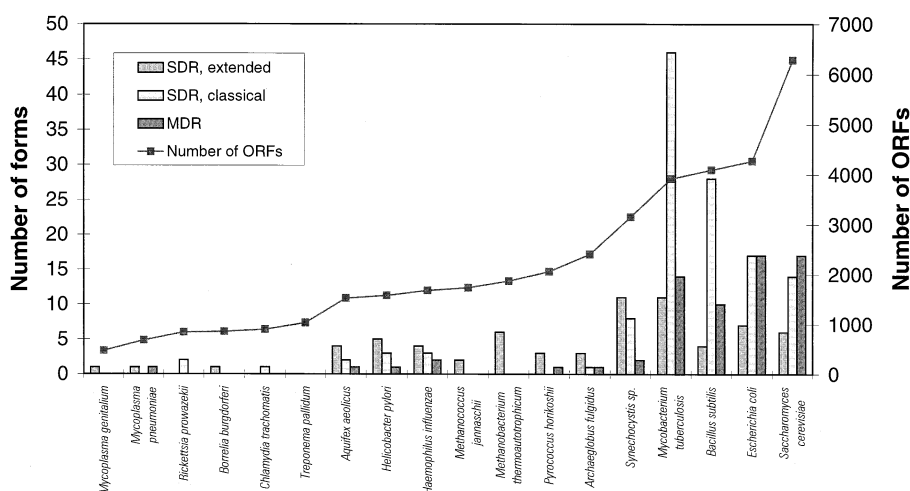

Fig. 1. Number of SDR and MDR forms in all 18 genomes ordered after the number of ORFs.
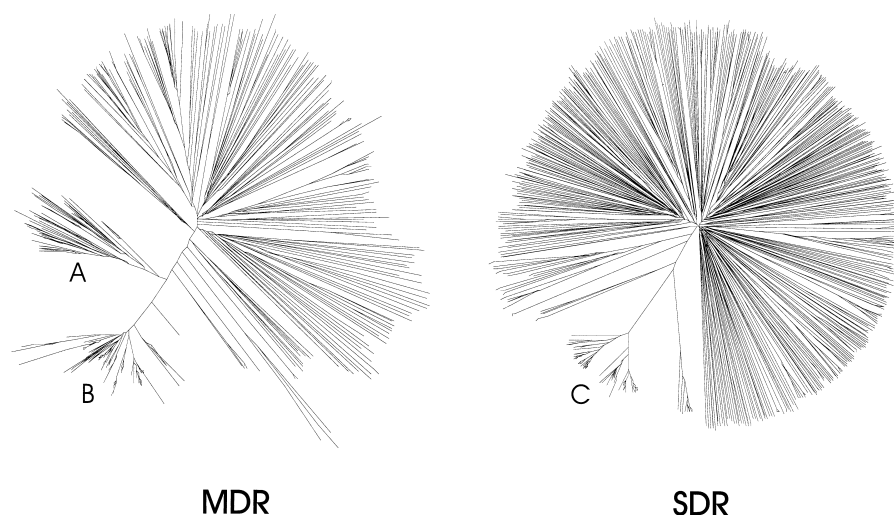
MDR   SDR

Fig. 2. Evolutionary patterns of MDR and classical SDR enzymes currently characterised. Structures in databank screenings for MDR and SDR enzymes were collected, aligned and assembled into a phylogenetic tree as previously described [14]. The three clustered branchings labelled A, B and C are explained by many species variants. Thus, A represents the animal alcohol dehydrogenases, B the plant alcohol dehydrogenases, and C the *Drosophila* alcohol dehydrogenases, but excluding these clusters of species variants recent branching from enzyme variants is still of greater occurrence in the MDR than the SDR family.

reported, even in *E. coli* expression of recombinant alcohol dehydrogenases, and has previously been identified as another enzyme type [22] rather than as an MDR alcohol dehydrogenase. We now see that the zinc-type ethanol dehydrogenase activity is present also in prokaryotes. This has recently also been verified by purification of the *E. coli* enzyme [23].

3.2. Evolutionary patterns of known MDR and SDR enzymes

Upon extension of the screenings, inclusion of not only the completed genomes as given above but also MDR and SDR sequences from any organism gives a total of 1056 SDR and 537 MDR structures. Alignment of these and construction of corresponding evolutionary trees give the patterns shown in Fig. 2. Again, these patterns of long branches show that both families are of old origin. However, close inspection shows that late sub-branching with separate enzyme activities is more frequent in MDR than in SDR. Hence, the pattern noticed before in MDR, with repeated gene duplications along a branch and successive evolution of new activities, is still valid in several cases. This is the pattern that constitutes the steps of enzymogenesis giving the different classes of vertebrate alcohol dehydrogenase [24]. However, the MDR pattern is now more complex than before, showing that MDR, in addition to the successive steps previously deduced, also has a distant ancestral origin and central functions in all organisms, just as SDR has.

3.3. Extensive alcohol dehydrogenase multiplicity and special properties

All data emphasise an extensive multiplicity for alcohol dehydrogenase. The *E. coli* genome [25] proves the existence of an MDR alcohol dehydrogenase multiplicity in prokaryotes, as previously shown for yeast alcohol dehydrogenase [26], plant alcohol dehydrogenase [27] and vertebrate alcohol dehydrogenase [24]. In each case, the most ethanol-active form appears to be of more recent origin than the life forms itself. Hence, the pattern with repeated alcohol dehydrogenase gene duplications in eukaryotic lines [11] is now substantiated by independent duplications also in the prokaryotic line. Ob-

viously, the emergence of alcohol dehydrogenases active on ethanol and other small alcohols/aldehydes has been repeated in all major lines of life. These properties suggest important functions not only for the ubiquitous GSH-dependent formaldehyde dehydrogenase (= class III alcohol dehydrogenase) [28] but also for the ethanol-active forms of other alcohol dehydrogenase types, giving a complex pattern in all lines. In addition, some of the prokaryotic ethanol-active forms show evidence of being inducible, while all Gram-positive prokaryotes do not appear to possess even the otherwise ubiquitous, GSH-dependent enzyme (class III alcohol dehydrogenase). Instead a mycothiol-dependent MDR form of related structure has been characterised from a Gram-positive prokaryote [29]. Further details on the origin and patterns of early alcohol dehydrogenases cannot yet be judged beyond the facts that both ethanol-active and GSH-dependent formaldehyde-active forms are common, that the pattern is complex, that the ethanol activity has been formed repeatedly, and that the formaldehyde dehydrogenase activity is old although not necessarily the oldest form.

3.4. Functional properties and general dehydrogenase/reductase importance

As shown above, three different lines of evidence (multiple occurrence in prokaryotes, early origins as evaluated by evolutionary trees (Fig. 1), and repeated duplications along a line) suggest that both SDR and MDR enzymes have important functions. Significantly, although many inborn errors of metabolism are known in humans and animals, absence of class III alcohol dehydrogenase, or of more than just one ethanol-active alcohol dehydrogenase class [30], is unknown in vertebrates. Apparently, both these activities are essential to eukaryotic life (with the possible exception of marine invertebrates [31], in which small substrates may anyway be eliminated to the environment). Notably, oxidations carried through by dehydrogenases in general can often be replaced by cytochrome P450 functions [32], but in reality this is apparently not the end solution, since the MDR/SDR enzymes appear to be essential in all multicellular organisms. Why then

this difference between the facts that replacement by other enzymes such as cytochrome P450 appears functionally possible but is still apparently not known to have occurred? One answer could be that dehydrogenases, in contrast to cytochromes, carry through substrate oxidations without generation of free radicals, and therefore do the elimination reactions in a manner less harmful to the cell. If this is the explanation for the common occurrence of the dehydrogenase functions of the SDR and MDR families, it would support the notion that these enzymes constitute major components in cellular defense reactions. This would mean a function compatible with their pattern of occurrence, and would still not exclude additional functions in special life forms, such as in retinoid metabolism regulating vertebrate differentiation as suggested for the class IV form of MDR alcohol dehydrogenase [33].

References

[1] Schwartz, M.F. and Jörnvall, H. (1976) Eur. J. Biochem. 68, 159–168.
[2] Thatcher, D.R. (1980) Biochem. J. 187, 875–883.
[3] Harris, I. (1964) Nature 203, 30–34.
[4] Jörnvall, H. (1973) Proc. Natl. Acad. Sci. USA 70, 2295–2298.
[5] Jörnvall, H., Persson, M. and Jeffery, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4226–4230.
[6] Eklund, H., Horjales, E., Jörnvall, H., Brändén, C.-I. and Jeffery, J. (1985) Biochemistry 24, 8005–8012.
[7] Vallee, B.L. and Auld, D.S. (1990) Proc. Natl. Acad. Sci. USA 87, 220–224.
[8] Krook, M., Marekov, L. and Jörnvall, H. (1990) Biochemistry 29, 738–743.
[9] Borrás, T., Persson, B. and Jörnvall, H. (1989) Biochemistry 28, 6133–6139.
[10] Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzàlez-Duarte, R., Jeffery, J. and Ghosh, D. (1995) Biochemistry 34, 6003–6013.
[11] Persson, B., Zigler Jr., J.S. and Jörnvall, H. (1994) Eur. J. Biochem. 226, 15–22.
[12] Wierenga, K., Maeyer, M.C.H. and Hol, W.G.J. (1985) Biochemistry 24, 1346–1357.
[13] Ghosh, D., Weeks, C.M., Grochulski, P., Duax, W.L., Erman, M., Rimsay, R.L. and Orr, J.C. (1991) Proc. Natl. Acad. Sci. USA 88, 10064–10068.
[14] Jörnvall, H., Danielsson, O., Hjelmqvist, L., Persson, B. and Shafqat, J. (1996) in: Gene Families: Structure, Function, Genetics and Evolution (Holmes, R.S. and Lim, H.A., Eds.), pp. 35–41, World Scientific, Singapore.
[15] http://www.tigr.org
[16] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[17] Bairoch, A. and Apweiler, R. (1998) Nucleic Acids Res. 26, 38–42.
[18] Bairoch, A., Bucher, P. and Hofmann, K. (1997) Nucleic Acids Res. 25, 217–221.
[19] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[20] Thorn, J.M., Barton, J.D., Dixon, N.E., Ollis, D.L. and Edwards, K.J. (1995) J. Mol. Biol. 249, 785–799.
[21] Clark, D. and Cronan Jr., J.E. (1980) J. Bacteriol. 141, 177–183.
[22] Goodlove, P.E., Cunningham, P.R., Parker, J. and Clark, D.P. (1989) Gene 85, 209–214.
[23] Shafqat, J., Höög, J.-O., Hjelmqvist, L., Oppermann, U., Ibáñez, C. and Jörnvall, H. (1999) Eur. J. Biochem. (submitted).
[24] Jörnvall, H. and Höög, J.-O. (1995) Alcohol Alcoholism 30, 153–161.
[25] Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) Science 277, 1453–1474.
[26] Young, E.T. and Pilgrim, D. (1985) Mol. Cell. Biol. 5, 3024–3034.
[27] Shafqat, J., El-Ahmad, M., Danielsson, O., Martínez, M.C., Persson, B., Parés, X. and Jörnvall, H. (1996) Proc. Natl. Acad. Sci. USA 93, 5595–5599.
[28] Uotila, L. and Koivusalo, M. (1989) in: Coenzymes and Cofactors. Glutathione. Chemical, Biochemical and Medical Aspects, Vol. III, part A (Dolphin, D. et al., Eds.), pp. 517–551, John Wiley and Sons, New York.
[29] Norin, A., van Ophem, P.W., Piersma, S.R., Persson, B., Duine, J.A. and Jörnvall, H. (1997) Eur. J. Biochem. 248, 282–289.
[30] Burnett, K.G. and Felder, M.R. (1978) Biochem. Genet. 16, 443–454.
[31] Fernández, M.R., Jörnvall, H., Moreno, A., Kaiser, R. and Parés, X. (1993) FEBS Lett. 328, 235–238.
[32] Lewis, D.F., Watson, E. and Lake, B.G. (1998) Mutat. Res. 410, 245–270.
[33] Duester, G. (1996) Biochemistry 35, 12221–12227.