Biochemistry

 Volume 34, Number 18 May 09, 1995

*Perspectives in Biochemistry*

# Short-Chain Dehydrogenases/Reductases (SDR)[†]

Hans Jörnvall,*[,‡] Bengt Persson,[‡,§] Maria Krook,[‡] Silvia Atrian,[∥] Roser Gonzàlez-Duarte,[∥] Jonathan Jeffery,[⊥] and Debashis Ghosh[#]

*Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden, European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, D-69012 Heidelberg, Germany, Department of Genetics, University of Barcelona, E-08071 Barcelona, Spain, Department of Molecular and Cell Biology, University of Aberdeen, Marischal College, Scotland, U.K., and Hauptman-Woodward Medical Research Institute, Buffalo, New York 14203-1196*

*Received December 13, 1994; Revised Manuscript Received March 8, 1995*[⊗]

ABSTRACT: Short-chain dehydrogenases/reductases (SDR) constitute a large protein family. Presently, at least 57 characterized, highly different enzymes belong to this family and typically exhibit residue identities only at the 15−30% level, indicating early duplicatory origins and extensive divergence. In addition, another family of 22 enzymes with extended protein chains exhibits part-chain SDR relationships and represents enzymes of no less than three EC classes. Furthermore, subforms and species variants are known of both families. In the combined SDR superfamily, only one residue is strictly conserved and ascribed a crucial enzymatic function (Tyr 151 in the numbering system of human $NAD^+$-linked prostaglandin dehydrogenase). Such a function for this Tyr residue in SDR enzymes in general is supported also by chemical modifications, site-directed mutagenesis, and an active site position in those tertiary structures that have been characterized. A lysine residue four residues downstream is also largely conserved. A model for catalysis is available on the basis of these two residues. Binding of the coenzyme, NAD(H) or NADP(H), is in the N-terminal part of the molecules, where a common GlyXXXGlyXGly pattern occurs. Two SDR enzymes established by X-ray crystallography show a one-domain subunit with seven to eight $\beta$-strands. Conformational patterns are highly similar, except for variations in the C-terminal parts. Additional structures occur in the family with extended chains. Some of the SDR molecules are known under more than one name, and one of the enzymes has been shown to be susceptible to native, chemical modification, producing reduced Schiff base adducts with pyruvate and other metabolic keto derivatives. Most SDR enzymes are dimers and tetramers. In those analyzed, the area of major subunit contacts involves two long $\alpha$-helices ($\alpha$E, $\alpha$F) in similar and apparently strong subunit interactions. Future possibilities include verification of the proposed reaction mechanism and tracing of additional relationships, perhaps also with other protein families. Short-chain dehydrogenases illustrate the value of comparisons and diversified research in generating unexpected discoveries.

Knowledge about short-chain dehydrogenases has increased during the last few years from just limited findings concerning the properties of a few dehydrogenases to a substantial body of structural information and characterized relationships. The first enzymes of this type were analyzed already in the 70's, giving first partial and later complete

[†] Supported by the Swedish Medical Research Council (projects 13X-3532 and 13F-10248), the Swedish Alcohol Research Fund (project 88/12:6), the Swedish Society of Medicine, the Foundation Blanceflor Boncompagni-Ludovisi, née Bildt, the Swedish Society for Medical Research, a Spanish CICYT grant (BI092-0591-C02-02), and the National Institutes of Health (DK26546).

* To whom correspondence should be addressed.
[‡] Karolinska Institutet.
[§] European Molecular Biology Laboratory.
[∥] University of Barcelona.
[⊥] University of Aberdeen.
[#] Hauptman-Woodward Medical Research Institute.

[⊗] Abstract published in *Advance ACS Abstracts,* April 15, 1995.

0006-2960/95/0434-6003$09.00/0 

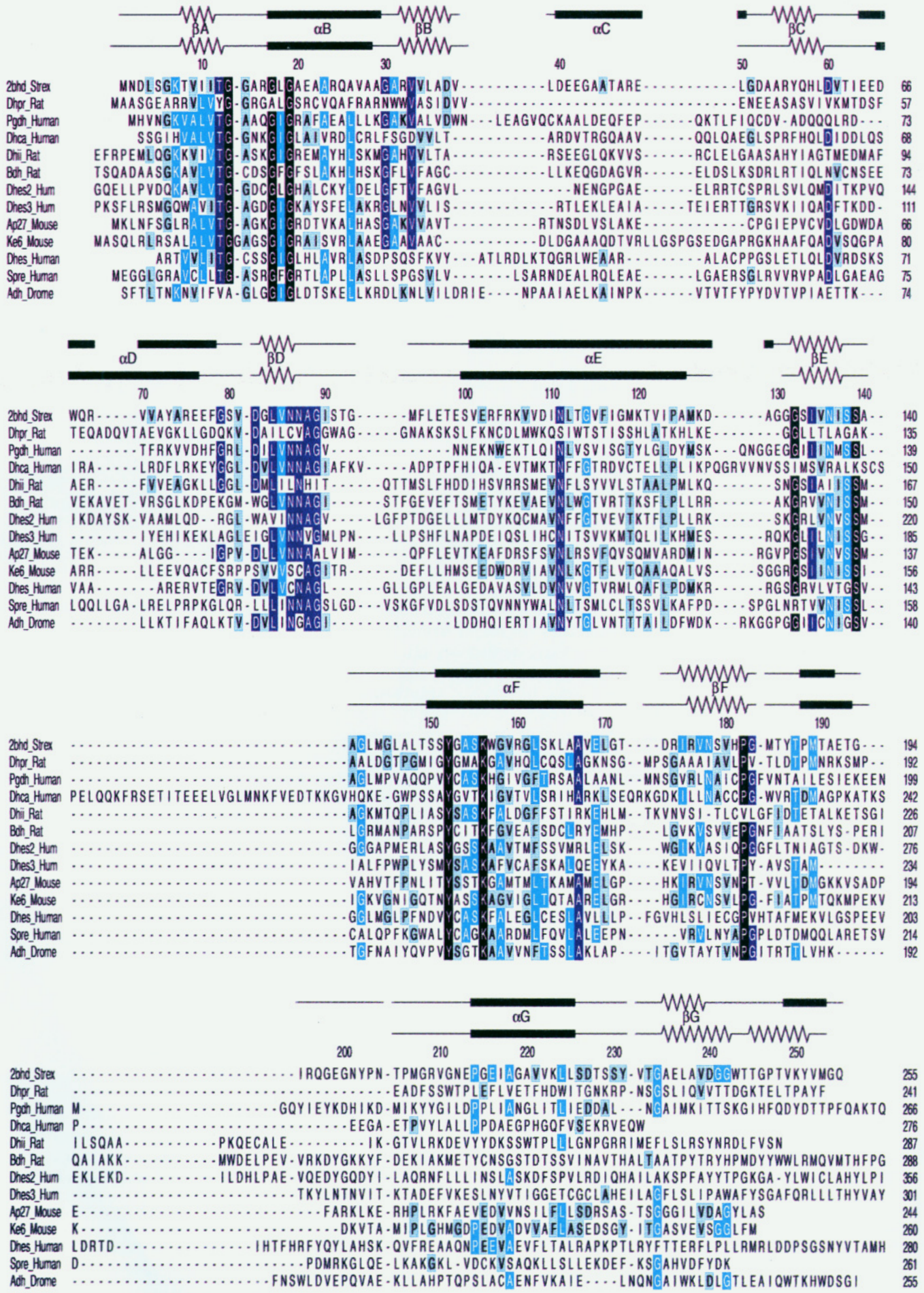
βA αB βB αC βC
10 20 30 40 50 60
2bhd_Strex MNDLSGKTVIITG-GARGLGAEAARQAVAAGARVVLADV----------LDEEGAATARE----------LGDAARYQHLDVTIEED 66
Dhpr_Rat MAASGEARRVLVYG-GRGALGSRCVQAFRARNWWVASIDVV----------------------------ENEEASASVIVKMTDSF 57
Pgdh_Human MHVNGKVALVTG-AAQGIGRAFAEALLKGAKVALVDWN---LEAGVQCKAALDEQFEP------QKTLFIQCDV-ADQQQLRD--- 73
Dhca_Human SSGIHVALVTG-GNKGIGLAIVRDLCRLFSGDVVLT----------ARDVTRGQAAV------QQLQAEGLSPRFHQLDIDDLQS 68
Dhii_Rat EFRPEMLQGKKVIVTG-ASKGIGREMAYHLSKMGAHVVLTA----------RSEEGLQKVVS------RCLELGAASAHYIAGTMEDMAF 94
Bdh_Rat TSQADAASGKAVLVTG-CDSGFGFSLAKHLHSKGFLVFAGC----------LLKEQGDAGVR------ELDSLKSDRLRTIQLNVCNSEE 73
Dhes2_Hum GQELLPVDQKAVLVTG-GDCGLGHALCKYLDELGFTVFAGVL----------NENGPGAE------ELRRTCSPRLSVLQMDITKPVQ 144
Dhes3_Hum PKSFLRSMGQWAVITG-AGDGIGKAYSFELAKRGLNVVLIS----------RTLEKLEAIA-----TEIERTTGRSVKIIQADFTKDD- 111
Ap27_Mouse MKLNFSGLRALVTG-AGKGIGRDTVKALHASGAKVVAVT----------RTNSDLVSLAKE----------CPGIEPVCVDLGDWDA 66
Ke6_Mouse MASQLRLRSALALVTGGAGSGIGRAISVRLAAEGAAVAAC----------DLDGAAAQDTVRLLGSPGSEDGAPRGKHAAFQADVSQGPA 80
Dhes_Human ARTVVLITG-CSSGIGLHLAVRLASDPSQSFKVY---ATLRDLKTQGRLWEAAR---------ALACPPGSLETLQLDVRDSKS 71
Spre_Human MEGGLGRAVCLLTG-ASRGFGRTLAPLLASLLSPGSVLV-------LSARNDEALRQLEAE------LGAERSGLRVVRVPADLGAEAG 75
Adh_Drome SFTLTNKNVIFVA-GLGGIGLDTSKELLKRDLKNLVILDRIE----NPAAIAELKAINPK------VTVTFYPYDVTVPIAETTK--- 74
αD βD αE βE
70 80 90 100 110 120 130 140
2bhd_Strex WQR-----VVAYAREEFGSV-DGLVNNAGISTG------MFLETESVERFRKVVDINLTGVFIGMKTVIPAMKD------AGGGSIVNISSA-- 140
Dhpr_Rat TEQADQVTAEVGKLLGDQKV-DAILCVAGGWAG-----GNAKSKSLFKNCDLMWKQSIWTSTISSHLATKHLKE--------GGLLTLAGAK-- 135
Pgdh_Human --------TFRKVVDHFGRL-DILVNNAGV----------NNEKNWEKTLQINLVSVISGTYLGLDYMSK---QNGGEGGIIINMSSL-- 139
Dhca_Human IRA-----LRDFLRKEYGGL-DVLVNNAGIAFKV-----ADPTPFHIQA-EVTMKTNFFGTRDVCTELLPLIKPQGRVVNVSSIMSVRALKSCS 150
Dhii_Rat AER-----FVVEAGKLLGGL-DMLILNHIT------QTTMSLFHDDIHSVRRSMEVNFLSYVVLSTAALPMLKQ-------SNGSIAIISSM-- 167
Bdh_Rat VEKAVET-VRSGLKDPEKGM-WGLVNNAGI------STFGEVEFTSMETYKEVAEVNLWGTVRTTKSFLPLLRR-------AKGRVVNISSM-- 150
Dhes2_Hum IKDAYSK-VAAMLQD--RGL-WAVINNAGV-----LGFPTDGELLLMTDYKQCMAVNFFGTVEVTKTFLPLLRK-------SKGRLVNVSSM-- 220
Dhes3_Hum --------IYEHIKEKLAGLEIGLVNNVGMLPN----LLPSHFLNAPDEIQSLIHCNITSVVKMTQLILKHMES------RQKGLILNISSG-- 185
Ap27_Mouse TEK-----ALGG----IGPV-DLLVNNAALVIM------QPFLEVTKEAFDRSFSVNLRSVFQVSQMVARDMIN-----RGVPGSIVNVSSM-- 137
Ke6_Mouse ARR-----LLEEVQACFSRPPSVVVSCAGITR------DEFLLHMSEEDWDRVIAVNLKGTFLVTQAAAQALVS-----SGGRGSIINISSI-- 156
Dhes_Human VAA-------ARERVTEGRV-DVLVCNAGL------GLLGPLEALGEDAVASVLDVNVVGTVRMLQAFLPDMKR------RGSGRVLVTGSV-- 143
Spre_Human LQQLLGA-LRELPRPKGLQR-LLLINNAGSLGD---VSKGFVDLSDSTQVNNYWALNLTSMLCLTSSVLKAFPD----SPGLNRTVVNISSL-- 158
Adh_Drome --------LLKTIFAQLKTV-DVLINGAGI----------LDDHQIERTIAVNYTGLVNTTTAILDFWDK---RKGGPGGIICNIGSV-- 140
αF βF
150 160 170 180 190
2bhd_Strex ----------AGLMGLALTSSYGASKWGVRGLSKLAAVELGT----DRIRVNSVHPG-MTYTPMTAETG--- 194
Dhpr_Rat ----------AALDGTPGMIGYGMAKGAVHQLCQSLAGKNSG--MPSGAAAIAVLPV-TLDTPMNRKSMP-- 192
Pgdh_Human ----------AGLMPVAQQPVYCASKHGIVGFTRSAALAANL--MNSGVRLNAICPGFVNTAILESIEKEEN 199
Dhca_Human PELQQKFRSETITEEELVGLMNKFVEDTKKGVHQKE-GWPSSAYGVTKIGVTVLSRIHARKLSEQRKGDKILLNACCPG-WVRTDMAGPKATKS 242
Dhii_Rat ----------AGKMTQPLIASYSASKFALDGFFSTIRKEHLM--TKVNVSI-TLCVLGFIDTETALKETSGI 226
Bdh_Rat ----------LGRMANPARSPYCITKFGVEAFSDCLRYEMHP----LGVKVSVVEPGNFIAATSLYS-PERI 207
Dhes2_Hum ----------GGGAPMERLASYGSSKAAVTMFSSVMRLELSK----WGIKVASIQPGGFLTNIAGTS-DKW- 276
Dhes3_Hum ----------IALFPWPLYSMYSASKAFVCAFSKALQEEYKA----KEVIIQVLTPY-AVSTAM--------- 234
Ap27_Mouse ----------VAHVTFPNLITYSSTKGAMTMLTKAMAMELGP----HKIRVNSVNPT-VVLTDMGKKVSADP 194
Ke6_Mouse ----------IGKVGNIGQTNYASSKAGVIGLTQTAARELGR----HGIRCNSVLPG-FIATPMTQKMPEKV 213
Dhes_Human ----------GGLMGLPFNDVYCASKFALEGLCESLAVLLLP--FGVHLSLIECGPVHTAFMEKVLGSPEEV 203
Spre_Human ----------CALQPFKGWALYCAGKAARDMLFQVLALEEPN------VRVLNYAPGPLDTDMQQLARETSV 214
Adh_Drome ----------TGFNAIYQVPVYSGTKAAVVNFTSSLAKLAP----ITGVTAYTVNPGITRTTLVHK------ 192
αG βG
200 210 220 230 240 250
2bhd_Strex ----------IRQGEGNYPN-TPMGRVGNEPGEIAGAVVKLLSDTSSY-VTGAELAVDGGWTTGPTVKYVMGQ 255
Dhpr_Rat ----------EADFSSWTPLEFLVETFHDWITGNKRP-NSGSLIQVVTTDGKTELTPAYF 241
Pgdh_Human M----------GQYIEYKDHIKD-MIKYYGILDPPLIANGLITLIEDDAL---NGAIMKITTSKGIHFQDYDTTPFQAKTQ 266
Dhca_Human P----------EEGA-ETPVYLALLPPDAEGPHGQFVSEKRVEQW 276
Dhii_Rat ILSQAA----------PKQECALE---------IK-GTVLRKDEVYYDKSSWTPLLLGNPGRRIMEFLSLRSYNRDLFVSN 287
Bdh_Rat QAIAKK----------MWDELPEV-VRKDYGKKYF-DEKIAKMETYCNSGSTDTSSVINAVTHALTAATPYTRYHPMDYYWWLRMQVMTHFPG 288
Dhes2_Hum EKLEKD----------ILDHLPAE-VQEDYGQDYI-LAQRNFLLLINSLASKDFSPVLRDIQHAILAKSPFAYYTPGKGA-YLWICLAHYLPI 356
Dhes3_Hum ----------TKYLNTNVIT-KTADEFVKESLNYVTIGGETCGCLAHEILAGFLSLIPAWAFYSGAFQRLLLTHYVAY 301
Ap27_Mouse E----------FARKLKE-RHPLRKFAEVEDVVNSILFLLSDRSAS-TSGGGILVDAGYLAS 244
Ke6_Mouse K----------DKVTA-MIPLGHMGDPEDVADVVAFLASEDSGY-ITGASVEVSGGLFM 260
Dhes_Human LDRTD----------IHTFHRFYQYLAHSK-QVFREAAQNPEEVAEVFLTALRAPKPTLRYFTTERFLPLLRMRLDDPSGSNYVTAMH 280
Spre_Human D----------PDMRKGLQE-LKAKGKL-VDCKVSAQKLLSLLEKDEF-KSGAHVDFYDK 261
Adh_Drome ----------FNSWLDVEPQVAE-KLLAHPTQPSLACAENFVKAIE----LNQNGAIWKLDLGTLEAIQWTKHWDSGI 255

FIGURE 1: Multiple sequence alignment of mammalian short-chain dehydrogenases/reductases (SDR). Left columns contain accession codes as listed in Table 1. A complete alignment of all 57 highly different SDR enzymes is available in the supplementary material. The abbreviated form given here includes, in addition to the mammalian forms, *Streptomyces* 3α/20β-hydroxysteroid dehydrogenase (top) and *Drosophila* alcohol dehydrogenase (bottom). Black indicates >90% conservation, white against dark blue 70−90%, white against medium blue 50−70%, and black against light blue 30−50%. Positions for the 3α/20β-hydroxysteroid dehydrogenase are given above the top sequence line; positional numbers for the other sequences are given to the right. Secondary structure elements are shown (top line for 3α/20β-hydroxysteroid dehydrogenase; bottom line for dihydropteridine reductase) above the alignment segments, with thick bars for α-helices and zigzag lines for β-sheets, as obtained with the program ICM (Abagyan et al., 1994) utilizing the outline function based upon crystallographic analyses. Gaps apparent in all 13 structures are explained by the complete alignment when all 57 structures are included. The figure was created using the program PRALIN.

primary structures of prokaryotic ribitol dehydrogenase (Moore et al., 1974; Dothie et al., 1985) and *Drosophila* alcohol dehydrogenase (Schwartz & Jörnvall, 1976; Thatcher, 1980). The proteins were then not recognized as a family, but the alcohol dehydrogenase was found to differ from the well-known alcohol dehydrogenases of liver and yeast, and the concept of a family of short-chain dehydrogenases was established when a distinctive pattern with the two alcohol dehydrogenase types was found to apply also to other dehydrogenases (Jörnvall et al., 1981). Early data suggested that the coenzyme-binding site was situated in the N-terminal part of the molecules (Thatcher & Sawyer, 1980), that mechanisms were unrelated to those of other dehydrogenases known at that time (Schwartz & Jörnvall, 1976; Thatcher 1980), and that a TyrXXXLys segment was of special functional importance, constituting the most conserved segment (Jörnvall et al., 1981). The short-chain dehydrogenase family then received little further attention for about a decade. Glucose dehydrogenase was added (Jörnvall et al., 1984), but the family was often considered special, consisting of prokaryotic dehydrogenases and insect alcohol dehydrogenase.

The field was greatly activated when mammalian enzymes of central interest, $NAD^+$-linked 15-hydroxyprostaglandin dehydrogenase (Krook et al., 1990) and 17β-hydroxysteroid dehydrogenase (Peltoketo et al., 1988), were shown also to be typical short-chain dehydrogenases (Krook et al., 1990). These two enzymes considerably extended the family in activity type and occurrence, again emphasizing the conservation and importance of the TyrXXXLys structure. One year later, 20 members were known (Persson et al., 1991) and utilized for functional conclusions and comparisons (Baker, 1990; Tannin et al., 1991; Baron et al., 1991; Krozowski, 1992; Neidle et al., 1992; Chen et al., 1993). Now, in 1994, we know 60-odd different enzymes of this type. In addition, species and allelic variants, especially of the *Drosophila* alcohol dehydrogenase (Villarroya et al., 1989; Atrian et al., 1992), increase the number further.

Tertiary structures have been determined by crystallographic analyses of 3α/20β-hydroxysteroid dehydrogenase (Ghosh et al., 1991, 1994) and dihydropteridine reductase (Varughese et al., 1992). Furthermore, prostaglandin dehydrogenases have been modeled (Krook et al., 1993a,b) into the steroid dehydrogenase structure. Knowledge of active sites and catalytic mechanisms has also increased. Comparisons constantly highlight the TyrXXXLys pattern, cloning and mutagenesis confirm the importance of this pattern (Ensor & Tai, 1991; Obeid & White, 1992; Albalat et al., 1992, 1994; Cols et al., 1993; Chen et al., 1993), and a catalytic mechanism based on these residues has been proposed (McKinley-McKee et al., 1991; Ghosh et al., 1994). In short, the family is well established, with a suggested role of specific residues, with a known one-domain tertiary structure, and with a widespread occurrence. On the basis of this knowledge, prospects of finding further characteristics and functional properties appear great. In particular, relationships to other proteins would lead to a new understanding of conformational relationships and family connections.

## THE SDR FAMILY

Data searches reveal that 57 characterized, highly different members exist within the short-chain dehydrogenase/reductase family (abbreviated form, Figure 1; complete form in the supplementary material, Figure 7; Table 1). They have a residue identity level of 15−30%, indicating distant duplications and early divergence (Figure 2). Only one residue is strictly conserved, the above-mentioned Tyr residue (Tyr 152 in the numbering system of the 3α/20β-hydroxysteroid dehydrogenase used throughout this treatise unless otherwise stated, corresponding to Tyr 151 of human $NAD^+$-linked prostaglandin dehydrogenase), confirming the fundamental importance of this residue. Lys 156 is almost as conserved. Also, a GlyXXXGlyXGly segment characteristic of the coenzyme binding fold in dehydrogenases in general (Rossmann et al., 1975; Wierenga et al., 1985) is largely constant, as is a coenzyme-interacting Asp residue that regulates the specificity for $NAD^+$ or $NADP^+$ by giving constraints on coenzyme binding (Scrutton et al., 1990; Bocanegra et al., 1993) in this (Chen et al., 1991) and other families of dehydrogenases (Rossmann et al., 1975).

Conservation of the critical residues suggests that the short-chain dehydrogenases are functionally related, while remaining overall residue identities suggest that they are also structurally related. Nevertheless, the spread in enzyme activity, turnover, direction of reaction, structure, subunit size, tissue distribution, and species origin should be noted. The recognition of the family assignment helps identify functional roles. Several of the structures have not been isolated at the protein level but have been detected just as open reading frames from nucleotide sequence analyses, where the products can now be ascribed oxidoreductase functions.

## CONFORMATION

Two of the structures in Figure 1 have been crystallographically analyzed, 3α/20β-hydroxysteroid dehydrogenase (Ghosh et al., 1991, 1994) and dihydropteridine reductase (Varughese et al., 1992). Both constitute one-domain structures (Figure 3) with a seven (the steroid dehydrogenase) or eight (the pteridine reductase) strand pattern in α/β arrangements, typical of the coenzyme-binding fold of dehydrogenases in general (Rossmann et al., 1975). Such a one-domain structure constitutes a clear difference from the

Table 1: Characterized Proteins Aligned into the Family of Short-Chain Dehydrogenases/Reductases in Figure 1[a]

| Code | Enzyme | Source | Code | Enzyme | Source |
|---|---|---|---|---|---|
| 2bhd_Strex | 20-β-Hydroxysteroid DH | *Streptomyces* | Enta_Ecoli | 2,3-Dihydro-2,3-DB DH | *Escherichia* |
| Dhpr_Rat | Dihydropteridine red | Rat | Fixr_Braja | FixR protein | *Bradyrhizobium* |
| Pgdh_Human | 15-Hydroxyprostaglandin DH ($NAD^+$) | Human | Ap27_Mouse | Adipocyte p27 protein | Mouse |
| Dhca_Human | Carbonyl red (NADPH) | Human | Dhma_Flasl | *N*-Acylmannosamine 1-DH | *Flavobacterium* |
| Pcr_Horvu | Protochlorophyllide red | *Hordeum* | 3adh_Psesp | 3-α-Hydroxysteroid DH | *Pseudomonas* |
| Hmtx_Leima | Aldoketo red | *Leishmania* | Ba74_Eubsp | 7-α-Hydroxysteroid DH | *Eubacterium* |
| 3bhd_Comte | 3-β-Hydroxysteroid DH | *Comamonas* | Hdha_Ecoli | 7-α-Hydroxysteroid DH | *Escherichia* |
| Ts2_Maize | Sex determination gene product TS2 | Maize | Dhgb_Bacme | Glucose 1-DH | *Bacillus* |
| Mas1_Agrra | Agropine synthesis red | *Agrobacterium* | Ba72_Eubsp | 7-α-Hydroxysteroid DH | *Eubacterium* |
| Dhii_Rat | Corticosteroid 11-β-DH | Rat | Act3_Strco | Putative ketoacyl red | *Streptomyces* |
| Bdh_Rat | D-β-Hydroxybutyrate DH | Rat | Tr1_Dastr | Troponine red I | *Datura* |
| Dhes2_Hum | Estradiol 17 β-DH type 2 | Human | Tr2_Dastr | Troponine red II | *Datura* |
| 25kd_Sarpe | Development-specific 25 kd protein | *Sarcophaga* | P29x_Dastr | Troponine red homologue | *Datura* |
| P6_Drome | Fat body protein P6 | *Drosophila* | Ver1_Asppa | ver-1 gene product | *Aspergillus* |
| Adh_Drome | Alcohol DH | *Drosophila* | Nodg_Azobr | Nodulation protein G. | *Azospirillum* |
| Bnze_Psepu | *cis*-1,2-Dihydrobenzene-1,2-diol DH | *Pseudomonas* | Phbb_Alceu | Acetoacetyl-CoA red | *Alcaligenes* |
| Bphb_Pseps | Biphenyl-*cis*-diol DH | *Pseudomonas* | Phbb_Zoora | Acetoacetyl-CoA red | *Zoogloea* |
| Budc_Klete | Acetoin DH | *Klebsiella* | Nodg_Rhime | Nodulation protein G | *Rhizobium* |
| Gutd_Ecoli | Sorbitol-6-phosphate 2-DH | *Escherichia* | Ardh_Cantr | Arabinitol DH | *Candida* |
| Bend_Acica | *cis*-1,2-Dihydroxy-3,4-C-1-C DH | *Acinetobacter* | Scdh_Picab | Short-chain DH | *Picea* |
| Xyll_Psepu | *cis*-1,2-Dihydroxy-3,4-C-1-C DH | *Pseudomonas* | Fabg_Cupla | 3-Oxoacyl-ACP red | *Cuphea* |
| Fox2_Yeast | Hydratase-DH-epimerase | *Saccharomyces* | Fabg_Ecoli | 3-Oxoacyl-ACP red | *Escherichia* |
| Hde_Cantr | Hydratase-DH-epimerase | *Candida* | Ke6_Mouse | Ke 6 protein | Mouse |
| Yinl_Lismo | Hyp 26.8 kD prot in inlA 5'reg | *Listeria* | Dhk2_Strvn | GPK putative ketoacyl red 2 | *Streptomyces* |
| Yrtp_Bacsu | Hyp 25.3 kD prot in rtp 5'reg | *Bacillus* | Dhes_Human | Estradiol 17 β-DH | Human |
| Dhes3_Hum | Estradiol 17 β-DH type 3 | Human | Spre_Human | Sepiapterin red | Human |
| Ligd_Psepa | Cα-DH, gene ligD | *Pseudomonas* | Yura_Myxxa | Hyp prot in uraA 5'region | *Myxococcus* |
| Ridh_Kleae | Ribitol 2-DH | *Klebsiella* | | | |

[a] Identification codes refer to the Swissprot database (Bairoch & Boeckmann, 1993) and its original references, except for the structures not yet in the database, for which references are given in the supplementary material, where EC numbers and data on taxonomic classification, quaternary structure, and coenzyme utilization are also listed. Abbreviations: ACP, acyl-carrier protein; C-1-C, cyclohexadiene-1-carboxylate; DB, dihydroxybenzoate; DH, dehydrogenase; GPK, granaticin polyketide synthase; hyp, hypothetical; prot, protein; red, reductase; reg, region.

two-domain structures of the well-studied mammalian alcohol, lactate, malate, and glyceraldehyde 3-phosphate dehydrogenases (Rossmann et al., 1975), which also have the coenzyme-binding structure ("the Rossmann fold"), but in one of two domains in different relative order. The coenzymes were initially considered to be differently bound between the characterized short-chain dehydrogenases and the two-domain dehydrogenases but are now known to be similarly bound deep in the catalytic cavity of both protein types [cf. Ghosh et al. (1994)].

The large size of the domain in the short-chain dehydrogenases is in part explained by three long strands ($\beta$D, $\beta$E, and $\beta$F) and two long helices (αE and αF). In the structures determined, helices αE and αF are conserved, constituting a subunit interacting area, giving strong dimer interactions in both the steroid dehydrogenase (tetramer) and dihydropteridine reductase (dimer). This surface corresponds to the interactions involving the areas symmetry-related by the crystallographic *Q* axis. Similarly, these helices are conserved in the two additional structures known from model building, $NAD^+$-linked prostaglandin dehydrogenase (Krook et al., 1993a) and $NADP^+$-linked prostaglandin dehydrogenase, or carbonyl reductase (Krook, et al., 1993b), but only the former is a dimer, while the latter is a monomer.

The two available crystal structures of typical short-chain dehydrogenases have similar folds for most regions but initially gave some different assignments, especially in the C-terminal parts and in coenzyme binding. On the basis of further data of the hydroxysteroid dehydrogenase (Ghosh et al., 1994), the overall subunit conformations and properties are now known to be quite similar, with localized differences

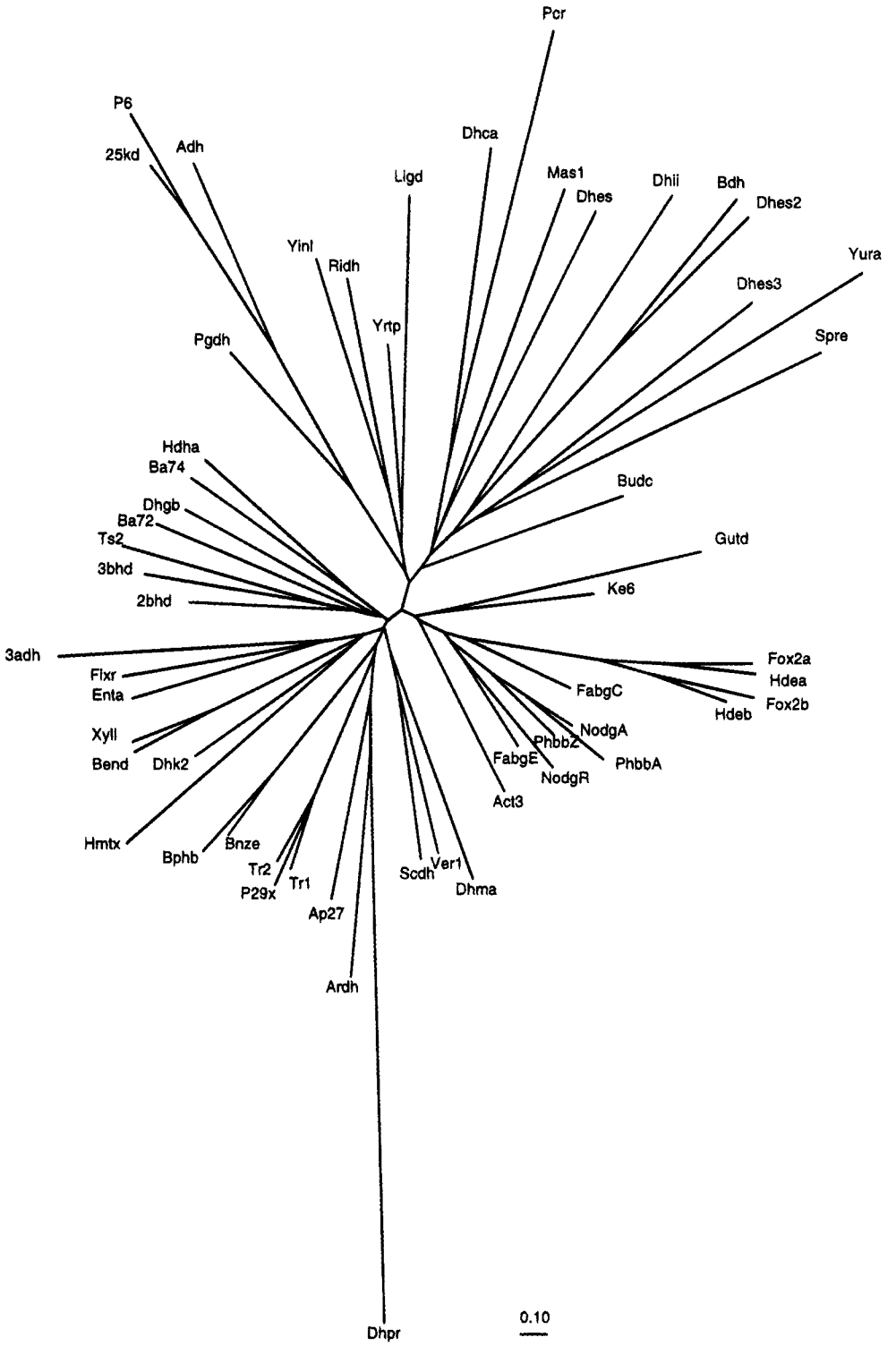


FIGURE 2: Phylogenetic tree drawn in the form of a star to emphasize the early and fairly uniformly distant duplications giving rise to the forms corresponding to the structures in Figure 7 (protein designations as in Table 1). The tree is calculated with the program ClustalW (Thompson et al., 1994), with branch lengths proportional to mutational distances (horizontal bar gives the length representing mutations of 10% of the positions).

largely in the C-terminal parts (Figure 4). The similarity between these two three-dimensional structures is impressive in view of the fact that the residue identity amounts to only 15%.

## CONSERVED RESIDUES

In the alignment of 57 short-chain dehydrogenases/reductases, only eight residues are conserved in more than 90% of the structures (black in Figure 1). They are the glycine residues involved in coenzyme binding (Gly 13, 17, and 19), the above-mentioned functionally important residues (Tyr 152 and Lys 156), and three further residues (Gly 132, Ser 139, and Pro 182). Significantly, half of the conserved residues are glycines, which is typical of distantly related proteins with a conserved fold [cf. Borrás et al. (1989)].

Considering also the residues conserved down to the 70% level (dark blue in Figure 1), we get sequence motifs or

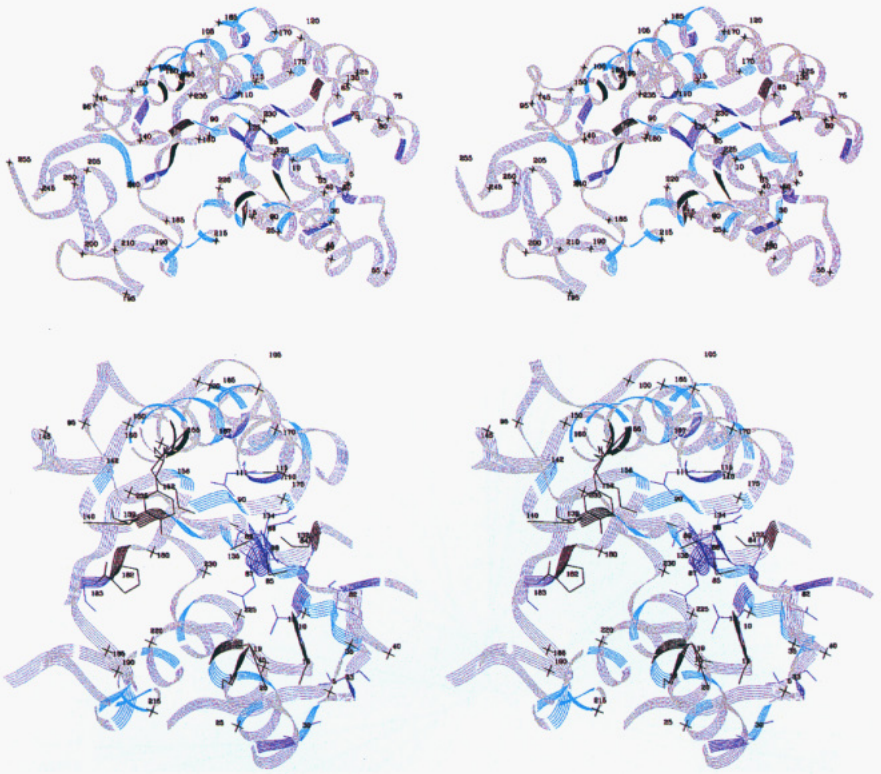

FIGURE 3: Stereoview of 3α/20β-hydroxysteroid dehydrogenase with conserved residues marked: (A, top) entire subunit; (B, bottom) close-up view of active site. Positions with a residue present in more than half of the structures of the multiple alignment (Figure 1) are marked with colors as in Figure 1. The figure was created using the program InsightII (Biosym). Most of the residues conserved (colored) are found around the active site, in the area of subunit contacts (top), and in the β-sheet. Coordinates from Ghosh et al. (1994).

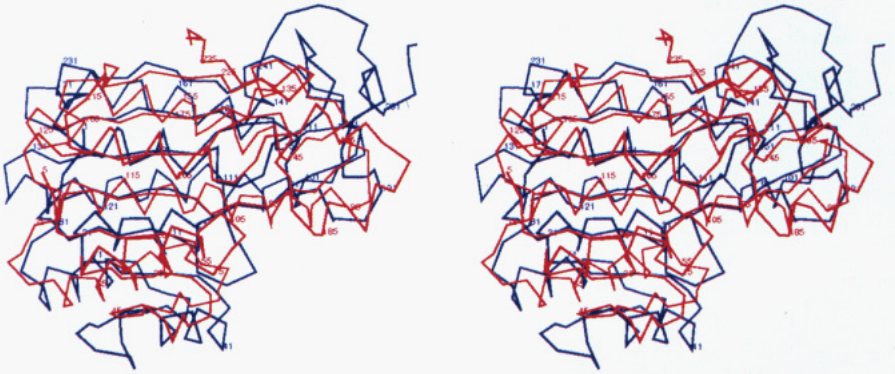

FIGURE 4: Superposition of the two presently available three-dimensional structures of short-chain dehydrogenases, 3α/20β-hydroxysteroid dehydrogenase [blue, from Ghosh et al. (1994)] and dihydropteridine reductase [red, from Varughese et al. (1992)], three-dimensionally aligned using the program MIDAS (Ferrin et al., 1988). The alignment is centered around segment 82−90 in the steroid dehydrogenase. A high similarity in the three-dimensional structure is clearly visible although the pairwise residue identity amounts to only 15%.

"fingerprints" typical of this enzyme family. They can be utilized in searching databases in order to find further family members. Still, glycine is the most common residue type conserved. Inclusion of residues conserved down to the 50% level (medium blue in Figure 1) increases the proportion of the hydrophobic residues Ile, Leu, Val, and Ala, indicating that a hydrophobic core of the fold is conserved throughout the family.

## THREE-DIMENSIONAL POSITIONS OF THE CONSERVED RESIDUES

The positions of the conserved residues in relation to the fold of 3α/20β-hydroxysteroid dehydrogenase are shown in Figure 3, color-coded in the same manner as in Figure 1. The conserved Tyr 152 and Lys 156 residues, as well as the largely conserved Gly pattern and the coenzyme-binding site deduced from the crystallographic analysis of the steroid dehydrogenase, are outlined. Exact positions of substrates are not known, but interpretations of the relationships suggest a substrate binding pocket and a catalytic site that affect Tyr 152. This is compatible with the sequence patterns and conservations. Also, Ser 139 and the segment Asn-Asn-Ala-Gly, positions 86−89, seem to be of interest (Figure 3B) regarding the active site and the coenzyme binding. Crystallographic analyses of the hydroxysteroid dehydrogenase suggest that Ser 139, as well as Thr 12, Arg 16, and Asn 87 (Ghosh et al., 1994), has enzymatic or binding functions (cf. below).

A cautious note should be added that active sites and mechanisms need not be identical for all short-chain dehydrogenases, considering the enormous spread of substrate specificity of the 57 enzymes characterized. The family may

Table 2: Mutants and Natural Variants Reported for Short-Chain Dehydrogenases[a]

| enzyme and position | mutated to | effect | enzyme and position | mutated to | effect |
|---|---|---|---|---|---|
| ADH | | | Tyr 152 | Gln | inactive |
| Asn 8 | Ala | active ($ADH_{UF}$) | | Cys | 0.25% |
| Ala 13 | Gly | 168% | Lys 156 | Ile | inactive |
| Gly 14 | Val | virtually inactive | | Arg | 2.2% |
| | | | Ser 165 | Phe | inactive |
| | Ala | 69% | Ala 167 | Val | inactive |
| | Asn | inactive | Pro 182 | Arg | inactive |
| Asp 38 | Asn | active | Gly 184 | Leu | inactive |
| | Leu | inactive | Thr 185 | Ile | inactive |
| | Arg | inactive | Lys 192 | Thr | active ($ADH_F$) |
| Ala 45 | Asp | active ($ADH_{UF}$) | Ser 195 | Phe | inactive |
| Ala 51 | Glu | active ($ADH_F'$) | Pro 214 | Ser | increased thermostability ($ADH_{ChD}$) |
| Asn 56 | Thr | 133% | | | |
| Thr 104 | Ile | inactive | Gly 232 | Glu | active ($ADH_D$) |
| Gly 130/Gly 132 | Cys/Ile | inactive | Trp 235 | Arg | inactive |
| Gly 132 | Val | inactive | PGDH | | |
| | Asp | inactive | Tyr 151 | SD | inactive |
| Cys 135 | Ala | 117% | HSDH | | |
| Cys 135/Cys 218 | Ala/Ala | 145% | Asp 110 | SD | decreased |
| Tyr 152 | Phe | inactive | Tyr 179 | Phe | inactive |
| | His | inactive | | Ser | inactive |
| | Glu | inactive | Lys 183 | Arg | inactive |

[a] Abbreviations: ADH, alcohol dehydrogenase from *D. melanogaster* and *D. lebanonensis*; PGDH, $NAD^+$-linked 15-hydroxyprostaglandin dehydrogenase; HSDH, 11$\beta$-hydroxysteroid dehydrogenase. In the column "effect", values relate to the activity versus the wild-type activity. Positional numbers refer to the actual number in the enzyme quoted and not to that in 3$\alpha$/20$\beta$-hydroxysteroid dehydrogenase as elsewhere in the text. Tyr 152 in *Drosophila* ADH corresponds to Tyr 151 in PGDH and Tyr 179 in HSDH (cf. Figure 1). The method of mutagenesis and references are given in the supplementary material.

have diverged into distinct sublines, having separate active site relationships and catalytic mechanisms. This would at this stage prevent further correlations. Nevertheless, such differences need not be postulated now, and if anything, the extreme conservation of a tyrosine residue suggests conserved functional influences.

The Gly pattern, with conserved GlyXXXGlyXGly, is largely identical in the short-chain dehydrogenases and the medium-chain, two-domain, liver alcohol, glyceraldehyde, and lactate dehydrogenases, highlighting an importance of strict space relationships and little free space in the coenzyme/enzyme interactions. However, slightly different spacings do occur [cf. Jörnvall et al. (1984)]. Furthermore, the sequence pattern is present also in other nucleotide-binding dehydrogenases/kinases, allowing recognition of binding sites in many cases (Wierenga et al., 1985; Bocanegra et al., 1993).

## SITE-DIRECTED MUTAGENESIS

DNA and cDNA data are available for nearly all of the short-chain dehydrogenases in Table 1, and many are also available as recombinant forms in expression systems ready for study by site-directed mutagenesis. Special predictions from the tertiary structures, and from comparisons of the native molecules, have constituted primary goals for testing by mutagenesis (Table 2). It is therefore no surprise that early reports on site-directed mutagenesis have involved the Gly pattern (Chen et al., 1990; Albalat et al., 1994) or the critical Tyr and Lys residues in 11$\beta$-hydroxysteroid dehydrogenase (Obeid & White, 1992) and *Drosophila* alcohol dehydrogenase (Albalat et al., 1992; Cols et al., 1993; Chen et al., 1993). In addition, the two Cys residues of *Drosophila* alcohol dehydrogenase have been studied (Chen et al., 1990) by mutagenetic exchanges, since all the alcohol dehydrogenases of the liver family type are Cys-rich with crucial Cys functions. In all the short-chain dehydrogenases studied, results of the site-directed mutagenesis fully confirm and extend the conclusions from crystallography and comparisons of native forms. Thus, recombinant proteins with the Gly pattern removed have virtually no activity (Chen et al., 1990), those with the complete Gly pattern restored lead to an increase of activity (Albalat et al., 1994), and those with the Tyr or Lys residues removed are inactive (Cols et al., 1993; Chen et al., 1993) except for special replacements (Y152C and K156R) which show some fractional activity (Chen et al., 1993). Direct chemical modifications also support a role of the Tyr residue (Krook et al., 1992), while mutagenetic exchanges of the Cys residues do not give inactive enzymes (Chen et al., 1990). Chemical mutagenesis (Schwartz & Jörnvall, 1976; Thatcher, 1980; Fossett et al., 1990; Mahmoud et al., 1991) has also been used and gives conclusions on roles of additional residues (Table 2).

It is to be expected that further site-directed mutagenesis will allow rapid progress along two additional lines. One is to establish whether there is uniformity among the short-chain dehydrogenases, i.e., whether functionally important residues are the same in most or all of the enzymes, as presently appears to be the case, or whether sublines and further subdivisions occur. The other is to establish further distinctions of crucial segments. Since few residues are strictly conserved, the similar structures appear to be stabilized more by conservation of properties and approximate positions than by conservation of exact positions. Alternatively, the presence of just a few strictly conserved residues might indicate that sublines and differences occur to a larger extent than hitherto recognized. These questions can be studied by analyses of additional, native forms but will also be amenable to analysis by mutagenesis. Suitable candidates for exchanges are those marked as conserved at the 70% and 90% levels in Figures 1 and 3 but still not

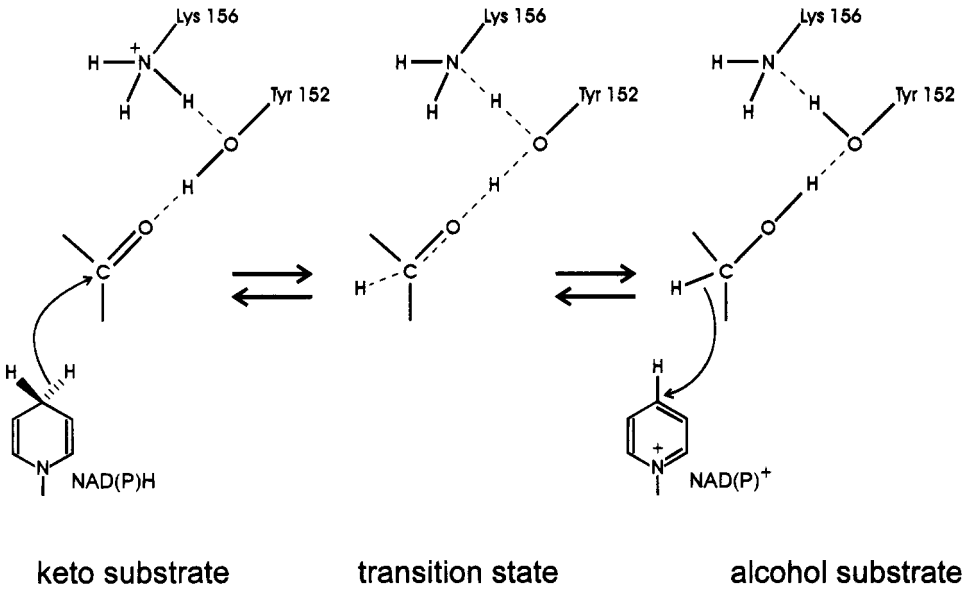


FIGURE 5: Reaction mechanism of short-chain dehydrogenases involving the conserved Tyr and Lys residues. The mechanism is as suggested by McKinley-McKee et al. (1991) and supported by crystallographic interpretations (Ghosh et al., 1991, 1994; Whiteley et al., 1993). Residue numbers refer to 3α/20β-hydroxysteroid dehydrogenase.

mutated (Table 2) and all those lining the active site (Figure 3B).

## CATALYTIC MECHANISM

A mechanism based on the observed conformations, native structures, and residue conservations would require a role for Tyr 152 and Lys 156. Such a role has been suggested (Jörnvall et al., 1981; Krook et al., 1990), and a mechanism with the Tyr in ionized form, stabilized via the side chain of Lys 156, giving a nucleophilic attack on the substrate, is likely (McKinley-McKee et al., 1991; Ghosh et al., 1991, 1994; Whiteley et al., 1993; Figure 5). However, other potentially important residues have also been pointed out, and the crystallographic study of the steroid dehydrogenase highlights the positions of Thr 12, Arg 16, and Asn 87 (Ghosh et al., 1991, 1994). Thus, the mechanism centers around a few residues which can now be studied further.

The Tyr/Lys mechanism would constitute an explanation of the strict conservation and appears compatible with many known data. However, considering that the proteins have widely different functions, dehydrogenases as well as reductases, the latter of both CO and CC double bonds, such a uniform mechanism might not have been anticipated. Notably, in the medium-chain alcohol dehydrogenases/reductases (MDR), relationships around the catalytic zinc atom, and even the catalytic zinc atom itself, are not at all conserved (Borrás et al., 1989; Persson et al., 1993). Thus, the possibility of a uniform SDR mechanism, as well as of a common Tyr function, should be further scrutinized.

## CHEMICAL MODIFICATIONS

During the analyses of one of the short-chain dehydrogenases, a novel type of native, chemical modification in proteins was established. Several forms of the carbonyl reductase, identical to $NADP^+$-linked prostaglandin dehydrogenase, differ by modification of a single lysine residue, at position 238 of that enzyme (Krook et al., 1993b; Forrest et al., 1990), where a Lys residue can be reductively alkylated by biogenic ketones, ranging from pyruvate (Krook et al., 1993b) to oxaloacetate and α-ketoglutarate (Wermuth et al., 1993). The reductive step appears to be carried through by an autocatalytic mechanism, and it is probably significant that carbonyl reductase, the protein for which this new type of modification was detected, has carbonyl compounds as substrates and reduction as its chemical reaction. Hence, although of general interest because of the frequent presence of Lys residues in proteins and of α-keto derivatives in cells, occurrence of this type of chemical modification may be limited. Lys 238 has a position separated from but close to the active site (Figure 3A). The different forms appear to have similar enzyme activity (Krook et al., 1993b; Wermuth et al., 1993). The residue obtained with the reduced pyruvate adduct is *N*-(carboxyethyl)lysine.

## EXTENDED SDR FAMILY

Another family of steroid dehydrogenases, including 3β-hydroxy-5-ene steroid dehydrogenase and UDP-glucose 4-epimerase, has been described (Baker & Blasco, 1992; Labesse et al., 1994). It now consists of 22 proteins and is distantly related to the SDR family. This was detected already in a previous multiple alignment study (Persson et al., 1991) and received additional support from a three-dimensional structural alignment of the epimerase and two members of the SDR family (Holm et al., 1994). The α-helices and β-sheets of these three enzymes are similarly positioned for the N-terminal 180-odd residues (Figure 6). Beyond position 180 in the UDP-glucose 4-epimerase, however, there are no similarities between this enzyme and the steroid dehydrogenase or dihydropteridine reductase. This position is at the domain border of the UDP-glucose 4-epimerase (Bauer et al., 1992). Therefore, the SDR family and the epimerase family appear to have a building block of 180 amino acid residues in common. Tyr 149 and Lys 153 of the epimerase, corresponding to Tyr 152 and Lys 156 of the 3α/20β-hydroxysteroid dehydrogenase, are also located at the binding pocket (Bauer et al., 1992). It seems likely that the SDR and epimerase enzyme families have similar

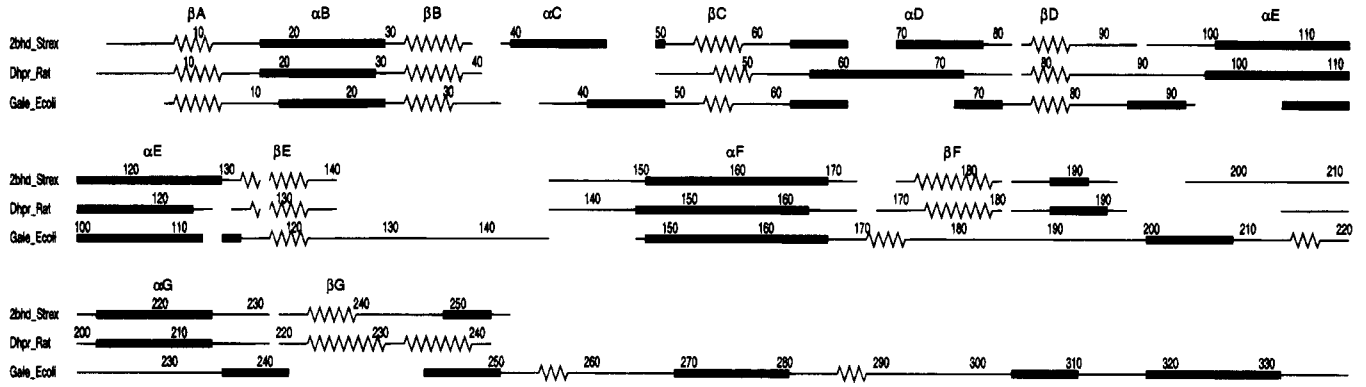


FIGURE 6: Schematic view of secondary structure elements of the three known tertiary structures. Top lines represent 20β-hydroxysteroid dehydrogenase (2bhd_Strex; Ghosh et al., 1991), middle lines dihydropteridine reductase (Dhpr_Rat; Varughese et al., 1992), and bottom lines UDP-glucose 4-epimerase (Gale_Ecoli; Bauer et al., 1992). α-Helices are indicated by thick bars and β-sheets by zigzag lines. The architectures of all three enzymes are similar for approximately the first 180 residues.

catalytic mechanisms. The extended SDR family then represents a remarkable spread of enzyme reactions, covering EC numbers from three different enzyme classes, i.e., oxidoreductases (dehydrogenases and reductases, EC 1), lyases (dehydratases, EC 4.2), and isomerases (epimerases, EC 5.1, and isomerases, EC 5.3).

## HYPOTHESIS: INTERFAMILY EXTENSIONS

The tertiary structure of enzymes is the structural level most conserved. It often implies family relationships even at stages when divergence has removed similarities in the primary structures, as demonstrated within the field of dehydrogenases by the nucleotide-binding fold [cf. Rossmann et al. (1975)] which exhibits few sequence similarities outside the Gly pattern mentioned above. In fact, very few examples are known of greatly altered tertiary structures that derive from similar origin, although influenza hemagglutinin (Bullough et al., 1994) and small peptides [cf. Jorgensen et al. (1991)] may constitute such examples, in revealing major conformational changes of even unaltered primary structures upon pH changes, membrane fusion, or binding interactions. Overall, it appears that conserved tertiary structures indicate common origins, while different tertiary structures often suggest separate origins.

However, it is known that single point mutations can destabilize protein structures, as when zinc ligands are removed in a dehydrogenase with structural zinc (Jeloková et al., 1994), critical residues are altered in the hemoglobinopathies well-known for a long time [cf. Perutz and Lehmann (1968)], or aggregational properties are altered with amyloid-forming proteins (McCutchen et al., 1993). If there has been a general divergence with successively arising new functions, many present-day structures now having different folds and functions should have related origins. In this regard, short-chain dehydrogenases/reductases are of special interest. They have a minimally part-chain relationship (Rossmann et al., 1975) to the medium-chain dehydrogenases [cf. Jörnvall et al. (1981)], the SDR enzyme dihydropteridine reductase appears to share structural features with enzymes of other families (Whiteley et al., 1993), and the extended SDR family involves enzyme types of several different EC classes, oxidoreductases, lyases, and isomerases. Furthermore, several SDR enzymes exhibit superficial coincidences with other reductases in metabolism. One such reductase with a completely different conformation is aldose reductase. It has an eight-folded β-pleated sheet barrel structure with coenzyme binding (Wilson et al., 1992; Borhani et al., 1992; Rondeau et al., 1992) different from that of both the short-chain dehydrogenases now outlined and the well-known medium-chain (MDR; Persson et al., 1994) enzymes of mammalian dehydrogenases (Rossmann et al., 1975). Nevertheless, aldose reductase has a size similar to that of the short-chain dehydrogenases, has eight β-strands, largely parallel arrangements of these, and belongs to another large family (Bruce et al., 1994) which catalyzes reactions similar to those of the short-chain dehydrogenases. It would be interesting to consider whether the multiple-strand nature, constituting an apparent coincidence between the SDR dihydropteridine reductase and the aldo−keto reductase family (both eight strands), could reflect a basic pattern in common, but now concealed by yielding two separate lines, with different assembly of the strands into a barrel (the aldo−keto reductases) or an extended Rossmann fold (SDR).

The answer is unknown, and the example may be far-fetched or constitute a wrong case but raises further interest in both SDR/MDR enzymes and α/β proteins in general. In part, this almost brings us back in time to when patterns were looked for in dehydrogenases, by attempts at lumping Cys reactivities (Harris, 1964; Holbrook et al., 1967) or different domains of the two-domain dehydrogenases (Jörnvall, 1977) into common origins. In the case of the reactive residues, that lumping turned out to be wrong. Chemical reactivity of single residues is now known to arise by different mechanisms. Lack of significance still also seems true regarding attempts at lumping domains with unrelated folds, but the extreme divergence and the coincidences in sheet patterns merit further analyses of the MDR, SDR, and other reductase families. Supersecondary structures (Rao & Rossmann, 1973) and molecular building units [cf. Danielsson et al. (1994)] are of interest.

## PERSPECTIVES

Which lessons can be drawn from the rapid progress in short-chain dehydrogenase research and from the recognition of this family and its structures? Overall, three aspects appear especially interesting to biochemistry.

One concerns the value of early comparisons and basic research in general. Much of the present knowledge has been collected *en passant*. One of the initial SDR structures

(ribitol dehydrogenase) was obtained because of purification from a system feeding organisms with a suboptimal substrate (Hartley, 1973). The other initial structure (*Drosophila* alcohol dehydrogenase) was studied because this polymorphic enzyme (Johnson & Denniston, 1964) occupied a position in the evolutionary selectionist−neutralist debate and its function in catabolizing alcohols was similar to that of mammalian alcohol dehydrogenases (Schwartz & Jörnvall 1976; Thatcher 1980). Most of the cloned structures exist as "sequence reports", and several are not known in protein form but deduced as open reading frames. Although easily overlooked, they fit into the scheme (Figure 7) and give metabolic insight. Even limited conclusions, like those from the early comparisons that initially suggested the conserved Tyr/Lys pattern, are valuable in promoting recognitions.

The second aspect of general interest relates to the presence of multiply modified subforms of these enzymes. Many dehydrogenases occur in posttranslationally modified forms, giving different bands upon electrophoresis or chromatography. In several cases, these multiplicities have been related to adducts or suggested to be derived from deamidations, oxidations, or other modifications. We now see that modifying groups derived from intermediary metabolism also occur, giving rise to a new type of native protein modification. Thus, specific chemical modifications may be more widespread than anticipated. Considering the reaction established with carbonyl reductase and carboxyethylation of lysine residues, it may be important to study further dehydrogenases with reactive substrates, such as aldehyde dehydrogenases.

Finally, the distinction of the short-chain dehydrogenase family may possibly lead to recognition of novel family relationships between unrelated conformations as hypothesized above. If true, this would illustrate transitions in protein conformations that would perhaps promote discoveries of unexpected, further relationships among families, proteins, and their building elements.

## NOTE ADDED IN PROOF

After compilation of the present structures, novel Swissprot database entries with clear SDR patterns and corresponding annotations have appeared (Dhkr_Strcm, Dlte_Bacsu, Hetn_Anasp, Kdud_Erwch, Oxir_Strat, Sord_Klepn, Yim4_Yeast, Yiv5_Yeast, Yiv6_Yeast, Yjgi_Ecoli, Yjgu_Ecoli, Ykf5_Yeast, and Yohf_Ecoli; all at >70% residue differences toward previous forms). Two additional forms, Ywfh_Bacsu (annotated) and Sx19_Yeast (nonannotated), have parts of this pattern. A further, recently reported tertiary structure (Dessen et al., 1995) appears to have SDR similarities but was only assigned to dehydrogenases in general and to a special reductase, suggesting the possibility of still further SDR branches.

## SUPPLEMENTARY MATERIAL AVAILABLE

Two tables giving extended versions of Tables 1 and 2 in the text and one figure extending Figure 1 in the text by giving a complete alignment of the 57 enzyme structures listed in Table 1 (5 pages). Ordering information is given on any current masthead page. The supplementary material is also available via anonymous ftp from ftp.mbb.ki.se.

## REFERENCES


Abagyan, R., Totrov, M., & Kuznetsov, D. (1994) *J. Comput. Chem. 15*, 488−506.

Albalat, R., Gonzàlez-Duarte, R., & Atrian, S. (1992) *FEBS Lett. 308*, 235−239.

Albalat, R., Atrian, S., & Gonzàlez-Duarte, R. (1994) *FEBS Lett. 341*, 171−176.

Atrian, S., Marfany, G., Albalat, R., & Gonzàlez-Duarte, R. (1992) *Biochem. Genet. 11*, 19−29.

Aziz, N., Maxwell, M. M., St.-Jacques, B., & Brenner, B. M. (1993) *Mol. Cell. Biol. 13*, 1847−1853.

Bairoch, A., & Boeckmann, B. (1993) *Nucleic Acids Res. 21*, 3093−3096.

Baker, M. E. (1990) *Biochem. J. 267*, 839−841.

Baker, M. E., & Blasco, R. (1992) *FEBS Lett. 301*, 89−93.

Baron, S. F., Franklund, C. V., & Hylemon, P. B. (1991) *J. Bacteriol. 173*, 4558−4569.

Bauer, A. J., Rayment, I., Frey, P. A., & Holden, H. M. (1992) *Proteins 12*, 372−381.

Bauer, S., Galliano, H., Pfeiffer, F., Messner, B., Sandermann, H., Jr., & Ernst, D. (1993) *Plant Physiol. 103*, 1479−1480.

Bocanegra, J. A., Scrutton, N. S., & Perham, R. N. (1993) *Biochemistry 32*, 2737−2740.

Borhani, D. W., Harter, T. M., & Petrash, J. M. (1992) *J. Biol. Chem. 267*, 24841−24847.

Borrás, T., Persson, B., & Jörnvall, H. (1989) *Biochemistry 28*, 6133−6139.

Bruce, N. C., Willey, D. L., Coulson, A. F. W., & Jeffery, J. (1994) *Biochem. J. 229*, 805−811.

Bullough, P. A., Hughson, F. M., Skehel, J. J., & Wiley, D. C. (1994) *Nature 371*, 37−43.

Callahan, H. L., & Beverley, S. M. (1992) *J. Biol. Chem. 267*, 24165−24168.

Chambers, G. K., Laver, W. G., Campbell, S., & Gibson, J. B. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3103−3107.

Chen, Z., Lu, L., Shirley, M., Lee, W. R., & Chang, S. H. (1990) *Biochemistry 29*, 1112−1118.

Chen, Z., Lee, W. R., & Chang, S. H. (1991) *Eur. J. Biochem. 202*, 263−267.

Chen, Z., Jiang, J. C., Lin, Z.-G., Lee, W. R., Baker, M. E., & Chang, S. H. (1993) *Biochemistry 32*, 3342−3346.

Cols, N., Marfany, G., Atrian, S., & Gonzàlez-Duarte, R. (1993) *FEBS Lett. 319*, 90−94.

Danielsson, O., Atrian, S., Luque, T., Hjelmqvist, L., Gonzàlez-Duarte, R., & Jörnvall, H. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4980−4984.

Dessen et al. (1995) *Science 267*, 1638−1641.

David, J. (1978) *Reserche 9*, 482−483.

DeLong, A., Calderon-Urrea, A., & Dellaporta, S. L. (1993) *Cell 74*, 757−768.

Dothie, J. M., Giglio, J. R., Moore, C. B., Taylor, S. S., & Hartley, B. S. (1985) *Biochem. J. 230*, 569−578.

Ensor, C. M., & Tai, H.-H. (1991) *Biochem. Biophys. Res. Commun. 176*, 840−845.

Ferrin, T. E., Huang, C. C., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.

Fletcher, T. S., Ayala, F. J., Thatcher, D. R., & Chambers, G. K. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5609−5612.

Forrest, G. L., Akman, S., Krutzik, S., Paxton, R. J., Sparekes, R. S., Doroshow, J., Felsted, R. L., Glover, C. J., Mohandas, T., & Bachur, N. R. (1990) *Biochim. Biophys. Acta 1048*, 149−155.

Fossett, N. G., Arbour-Reily, P., Kilroy, G., McDaniel, M., Mahmoud, J., Tucker, A. B., Chang, S. H., & Lee, W. R. (1990) *Mutat. Res. 231*, 73−85.

Geissler, W. M., Davis, D. L., Wu, L., Bradshaw, K. D., Patel, S., Mendonca, B. B., Elliston, K. O., Wilson, J. D., Russell, D. W., & Andersson, S. (1994) *Nature Genet. 7*, 34−39.

Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L., & Orr, J. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10064−10068.

Ghosh, D., Wawrzak, Z., Weeks, C. M., Duax, W. L., & Erman, M. (1994) *Structure 2*, 629−640.

Harris, I. (1964) *Nature 203*, 30−34.

Hartley, B. S. (1973) *Abstr. Int. Congr. Biochem. 9*, 7.

Holbrook, J. J., Pfleiderer, G., Mella, K., Volz, M., Leskowac, W., & Jeckel, R. (1967) *Eur. J. Biochem. 1*, 476−481.

Holm, L., Sander, C., & Murzin, A. (1994) *Nature Struct. Biol. 1*, 146−147.

Ichinose, H., Katoh, S., Sueoka, T., Titani, K., Fujita, K., & Nagatsu, T. (1991) *Biochem. Biophys. Res. Commun. 179*, 183−189.
Jeloková, J., Karlsson, C., Estonius, M., Jörnvall, H., & Höög, J.-O. (1994) *Eur. J. Biochem. 225*, 1015−1019.
Johnson, F. M., & Denniston, C. (1964) *Nature 204*, 906−907.
Jorgensen, W. L. (1991) *Science 254*, 954−955.
Jörnvall, H. (1977) *Eur. J. Biochem. 72*, 443−452.
Jörnvall, H., Persson, M., & Jeffery, J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4226−4230.
Jörnvall, H., von Bahr-Lindström, H., Jany, K.-D., Ulmer, W., & Fröschle, M. (1984) *FEBS Lett. 165*, 190−196.
Krook, M., Marekov, L., & Jörnvall, H. (1990) *Biochemistry 29*, 738−743.
Krook, M., Prozorovski, V., Atrian, S., Gonzàlez-Duarte, R., & Jörnvall, H. (1992) *Eur. J. Biochem. 209*, 233−239.
Krook, M., Ghosh, D., Duax, W., & Jörnvall, H. (1993a) *FEBS Lett. 322*, 139−142.
Krook, M., Ghosh, D., Strömberg, R., Carlquist, M., & Jörnvall, H. (1993b) *Proc. Natl. Acad. Sci. U.S.A. 90*, 502−506.
Krozowski, Z. (1992) *Mol. Cell. Endocrinol. 84*, C25−C31.
Labesse, G., Vidal-Cros, A., Chomilier, J., Gaudry, M., & Mornon, J.-P. (1994) *Biochem. J. 304*, 95−99.
Mahmoud, J., Fossett, N. G., Arbour-Reily, P., McDaniel, M., Tucker, A. B., Chang, S. H., & Lee, W. R. (1991) *Environ. Mol. Mutagen, 18*, 157−160.
McCutchen, S. L., Colon, W., & Kelly, J. W. (1993) *Biochemistry 32*, 12119−12127.
McKinley-McKee, J. S., Winberg, J.-O., & Pettersson, G. (1991) *Biochem. Int. 25*, 879−885.
Moore, C. H., Taylor, S. S., Smith, M. J., & Hartley, B. S. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, p 68, National Biomedical Research Foundation, Washington, DC.
Nakajima, K., Hashimoto, T., & Yamada, Y. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9591−9595.
Neidle, E., Hartnett, C., Ornston, N., Bairoch, A., Rekik, M., & Harayama, S. (1992) *Eur. J. Biochem. 204*, 113−120.
Obeid, J., & White, P. C. (1992) *Biochem. Biophys. Res. Commun. 188*, 222−227.
Peltoketo, H., Isomaa, V., Mäentausta, O., & Vihko, R. (1988) *FEBS Lett. 239*, 73−77.
Persson, B., Krook, M., & Jörnvall, H. (1991) *Eur. J. Biochem. 200*, 537−543.
Persson, B., Bergman, T., Keung. W. M., Waldenström, U., Holmquist, B., Vallee, B. L., & Jörnvall, H. (1993) *Eur. J. Biochem. 216*, 49−56.
Persson, B., Zigler, J. S., Jr., & Jörnvall, H. (1994) *Eur. J. Biochem. 226*, 15−22.
Perutz, M. F., & Lehmann, H. (1968) *Nature 219*, 902−909.
Rao, S. T., & Rossmann, M. G. (1973) *J. Mol. Biol. 76*, 241−256.
Rat, L., Veuille, M., & Lepesant, J.-A. (1991) *J. Mol. Evol. 33*, 194−203.
Rondeau, J.-M., Tête-Favier, F., Podjarny, A., Reymann, J.-M., Barth, P., Biellmann, J.-F., & Moras, D. (1992) *Nature 355*, 469−472.
Rossmann, M. G., Liljas, A., Brändén, C.-I., & Banaszak, L. J. (1975) in *The Enzymes, 3rd Edition* (Boyer, P. D., Ed.) Vol. 11, pp 61−102, Academic Press, New York.
Schwartz, M. F., & Jörnvall, H. (1976) *Eur. J. Biochem. 68*, 159−168.
Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Nature 343*, 38−43.
Skory, C. D., Chang, P.-K., Cary, J., & Linz, J. E. (1992) *Appl. Environ. Microbiol. 58*, 3527−3537.
Suzuki, K., Ueda, S., Sugiyama, M., & Imamura, S. (1993) *Gene 130*, 137−140.
Tannin, G. M., Agarwal, A. K., Monder, C., New, M. I., & White, P. C. (1991) *J. Biol. Chem. 266*, 16653−16658.
Thatcher, D. R. (1980) *Biochem. J. 187*, 875−883.
Thatcher, D. R., & Sawyer, L. (1980) *Biochem. J. 187*, 884−886.
Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) *Comput. Appl. Biosci. 10*, 19−29.
Varughese, K. I., Skinner, M. M., Whiteley, J. M., Matthews, D. A., & Xuong, N. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6080−6084.
Villarroya, A., Juan, E., Egestad, B., & Jörnvall, H. (1989) *Eur. J. Biochem. 180*, 191−197.
Wermuth, B., Bohren, K. M., & Ernst, E. (1993) *FEBS Lett. 335*, 151−154.
Whiteley, J. M., Xuong, N. H., & Varughese, K. I. (1993) *Adv. Exp. Med. Biol. 338*, 115−121.
Wierenga, R. K., Maeyer, M. C. H., & Hol, W. G. J. (1985) *Biochemistry 24*, 1346−1357.
Wilson, D. K., Bohren, K. M., Gabbay, K. H., & Quiocho, F. A. (1992) *Science 257*, 81−84.
Wu, L., Einstein, M., Geissler, W. M., Chan, H. K., Elliston, K. O., & Andersson, S. (1993) *J. Biol. Chem. 268*, 12964−12969.

BI942859X