Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes

YVONNE KALLBERG,^{1,2} UDO OPPERMANN,¹ HANS JÖRNVALL,¹ AND BENGT PERSSON^{1,2}

¹Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden

²Stockholm Bioinformatics Centre, Karolinska Institutet, S-171 77 Stockholm, Sweden

(RECEIVED July 3, 2001; FINAL REVISION November 30, 2001; ACCEPTED November 30, 2001)

Abstract

The progress in genome characterizations has opened new routes for studying enzyme families. The availability of the human genome enabled us to delineate the large family of short-chain dehydrogenase/reductase (SDR) members. Although the human genome releases are not yet final, we have already found 63 members. We have also compared these SDR forms with those of three model organisms: *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana*. We detect eight SDR ortholog clusters in a cross-genome comparison. Four of these clusters represent extended SDR forms, a subgroup found in all life forms. The other four are classical SDRs with activities involved in cellular differentiation and signalling. We also find 18 SDR genes that are present only in the human genome of the four genomes studied, reflecting enzyme forms specific to mammals. Close to half of these gene products represent steroid dehydrogenases, emphasizing the regulatory importance of these enzymes.

Keywords: Short-chain dehydrogenases/reductases; human genome; cross-genome comparisons; orthologs; bioinformatics; steroid dehydrogenases

Short-chain dehydrogenases/reductases (SDRs) are one-domain NAD(P)(H)-dependent enzymes of typically 250 amino acid residues (Jörnvall et al. 1995). They display a wide substrate spectrum, ranging from steroids, alcohols, sugars, and aromatic compounds to xenobiotics. SDRs are subdivided into classical and extended SDRs (Persson et al. 1995), differing in lengths and cofactor-binding motifs. SDRs constitutes a large family (Jörnvall et al. 1999) with about 2000 forms known, including species variants. The family is highly divergent, with typically 15%–30% residue identity in pairwise comparisons. For close to 20 members, the three-dimensional structures have been determined. In

spite of the low residue identities between the different members, the folding pattern is conserved with largely superimposable peptide backbones (Krook et al. 1993; Ghosh et al. 2001). The criterion for SDR membership is the occurrence of typical sequence motifs, arranged in a specific manner. These motifs comprise Rossmann-fold elements for nucleotide binding and specific residues for the active site including its highly conserved triad of Ser, Tyr, and Lys residues (Jörnvall et al. 1995; Persson et al. 1995; Oppermann et al. 1997).

Recent progress in genome characterizations has opened new routes for studies of this enzyme family. We have previously reported on the occurrence of SDR members in the bacterial and yeast genomes (Jörnvall et al. 1999). The availability of the human genome (International Human Genome Sequencing Consortium 2001; Venter et al. 2001) enabled us to find novel relationships. We have also extracted SDR members from the genomes of three model organisms (*Caenorhabditis elegans*, *Drosophila melanogaster* and

Reprint requests to: Hans Jörnvall, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden; e-mail: hans.jornvall@mbb.ki.se; fax: +46-8-337-462.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.26902>.

Arabidopsis thaliana) and clustered the SDR sequences to define ortholog groups within this widespread enzyme family.

Results and Discussion

SDR occurrence

We have searched for the occurrence of SDR members in the human genome and in the genomes of a worm, a fruit fly and a plant. In total, we found 352 forms, ranging from 63 in the human to 132 in *A. thaliana* (Table 1). The high number in *A. thaliana* is compatible with the tetraploidy and gene multiplicity in that genome (Bancroft 2000). After reduction of close homologs (>60% residue identity in pairwise comparisons), the numbers do not vary much, but range from 56 to 62 (Table 1).

Human genome

We detected 63 different human SDR enzymes, which are reduced to 58 after elimination of close homologs that correspond to possible isozymes. Because the human genome is not yet fully interpreted, we estimate the final number of human SDR enzymes to be still larger. Of the 63 SDR forms found, 46 are of the classical type whereas 17 are of the extended type. This relationship is also found in yeast and *E. coli* (Jörnval et al. 1999), with the majority of the enzymes being classical rather than extended. This is in contrast to the situation for organisms with smaller genomes (<~2000 open reading frames), in which only the extended forms are found (Jörnval et al. 1999).

SDR forms occurring in all genomes

We have defined the orthologous forms of the human SDRs in the three model genomes of *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana*. As a result of this search, we find eight SDR clusters present in all four genomes (Table 2, top). Of these, half are of the extended type, thus exceeding the ratio expected from the

general distribution of human SDRs (13 extended clusters versus 34 classical clusters; Table 2). This emphasizes the ubiquitous occurrence of the extended SDRs, compatible with their presence also in genomes from unicellular organisms (Jörnval et al. 1999).

All four clusters with extended SDRs have defined roles in carbohydrate metabolism and display similarities to sugar-metabolizing enzymes from bacterial organisms. This is the case with clusters 2 and 3, UDP-glucose 4-epimerase and GDP-D-mannose dehydratase, which are present also in Archaea. Cluster 4 has GDP-fucose synthase activity; cluster 1 contains enzymes that have undefined functions but bear sequence similarities to bacterial dDTP-D-glucose dehydratases. The clusters with extended SDRs display a higher proportion of strictly conserved residues (24%–46%) than the clusters with classical SDRs (7%–30%).

The four generally occurring clusters with classical SDRs are represented by the human Hep27, FVT1, 17 β -HSD3, and WWOX structures. To date, the molecular function only of 17 β -HSD3 has been defined (Geissler et al. 1994). This enzyme is involved in tissue-specific testosterone synthesis and defects in its gene cause male pseudo-hermaphroditism. The corresponding SDR cluster (cluster 7 in Table 2) is the second largest of all eight clusters, containing additional sequences: two from human, five from *D. melanogaster*, three from *C. elegans*, and two from *A. thaliana*. The defined function of 17 β -HSD3 indicates that these gene products are involved in tissue- and species-specific steroid metabolism.

The other three clusters of the classical type contain human enzymes with hitherto undefined functions. Hep27 is a nuclear protein with a restricted expression pattern and the corresponding cDNA was cloned from a hepatocarcinoma cell line (Gabrielli et al. 1995). This cluster also contains the human gene SRL (O95162 in Table 2), which codes for a peroxisomal SDR and displays 64% identity at the amino acid level towards Hep27 (Fransen et al. 1999). Similarly, the remaining two clusters, represented by FVT1 and WWOX, show involvement in malignant transformations. FVT1 was identified as a target in a subset of chromosomal

Table 1. Number of SDR forms present in all four versus only three, two, or one of the genomes investigated

Genome	Number of genomes				Total	After reduction of close homologs ^a (total)	Genome size (genes)
	4	3	2	1			
Human	8 (15)	11 (14)	10 (16)	16 (18)	63	58	39 256
<i>D. melanogaster</i>	8 (16)	10 (27)	6 (18)	8 (8)	69	62	14 335
<i>C. elegans</i>	8 (19)	11 (17)	5 (21)	14 (27)	84	56	19 727
<i>A. thaliana</i>	8 (31)	1 (4)	3 (9)	44 (88)	132	62	27 423
Total	8 (81)	11 (62)	12 (64)				

Numbers without parentheses correspond to the clusters, while numbers within parentheses give the total number. The last column gives the total estimated number of genes within each genome.

^a>60% residue identity.

translocations in non-Hodgkin lymphomas, leading to a juxtaposition of immunoglobulin κ -light chains to the FVT1-gene on chromosome 18, coupled with overexpression in certain T-cell malignancies. It is also located near the BCL-2 gene. Combined, these facts indicate involvement of FVT1 in tumorigenic processes (Rimokh et al 1993).

Another cytogenetic study identified WWOX, a human representative of the largest SDR cluster found in this study (cluster 8), as a gene product located on chromosome 16q23.3–24.1, a region affected frequently by allelic loss in breast cancer (Bednarek et al. 2000). Over-expression of WWOX in breast cancer compared to normal tissue was noted. Expression of this protein in other steroidogenic tissues such as prostate, ovary, and testis indicates that the intrinsic enzymatic WWOX activity is related to sex-steroid metabolism. This assumption is supported further by the tissue distribution and structural features of two other human forms found in this cluster, CGI-82 and PAN2. CGI-82 is expressed abundantly in the prostate and exhibits a low expression level in other tissues (Lin et al. 2001); PAN2 displays sequence similarities to other hydroxysteroid dehydrogenases.

Genome comparisons

Apart from the 8 SDR clusters present in all genomes investigated, orthologs to the human forms were detected in two model genomes for 11 SDR clusters and in one model genome for a further 10 SDR clusters (Table 2). The SDR forms missing in one or two of the model genomes probably have developed after the respective species divergence or have resulted from gene loss. Of all 31 clusters with SDR members in at least two genomes (Table 2), only two do not have a human representative.

Mammal-specific SDRs

We investigated the SDR enzymes present in only one of the genomes (Table 1). After reduction of close homologs within each species, each of the four genomes was found to have between 8 and 44 forms. The majority of these forms belong to the classical SDRs and have defined enzymatic functions such as steroid dehydrogenase or retinol dehydrogenase activities.

Sixteen forms were found in the human genome only, corresponding to 18 different proteins (Table 2, bottom). Homologs of all but three were found in other mammals, such as mouse, rat, cow, and pig. Close to half of these proteins are active on steroids, reflecting the importance of steroids in regulation of physiological functions and metabolic conversions in mammals. As deduced from their functions, these enzymes appear to be suitable targets for development of novel drugs directed at influencing hormone metabolism.

Clustering technique

The advantage of the clustering technique used in these genome comparisons, with the clusters formed by reciprocal relationships, is its insensitivity to the extent of residue identity or evolutionary speed. This is reflected by the fact that we could detect eight SDR clusters in all four genomes (Table 2), although intra-cluster residue conservation ranged from 46% to 7%. In some cases, the clustering procedure can be ambiguous; an example is cluster 8, which would otherwise separate into two clusters, one with the human WWOX form and one with the human PAN2 form, disregarding the *A. thaliana* forms. However, the *A. thaliana* members show mixed relationship with both sub-clusters; therefore, all forms are joined into one large cluster. As a consequence, orthologs to human forms of cluster 8 can be assigned only for *D. melanogaster* and *C. elegans*. The widespread nature of this cluster is also reflected by the fact that only a minor part (7%) of the multiple sequence alignment is strictly conserved (Table 2, cluster 8).

Evolutionary speed

The eight clusters with members from all four genomes were investigated with respect to their speed of divergence, after correction for multiple mutations according to Kimura (1983). The values obtained were divided with the value corresponding to the evolutionary divergence times for the animal–plant, worm–insect/mammal, and insect–mammal splits, reported to be 1215, 1045, and 850 million years, respectively (Feng et al. 1997). The clusters with extended SDRs were found to have evolutionary speeds of about 5 changes per 100 residues and 100 million years, whereas the clusters with classical SDRs have evolutionary speeds more than double this value. The slow evolution of the extended SDR enzymes is compatible with the conclusions drawn above that the extended SDR enzyme activities are present in most organisms and represent an ancient metabolic solution. However, the slow speed of the extended versus the classical SDRs is also compatible with the fact that most classical SDR subunits are smaller than the extended SDR subunits. Another case with two types of intrafamily evolutionary speed was noticed early on for the medium-chain dehydrogenase/reductase (MDR) alcohol dehydrogenases, in which “constant” and “variable” forms differ in speed from 6 to 18 changes per 100 residues and 100 million years (Jörnvall et al. 1993). Subsequently, that enzyme variability has been extended further to cover a >4-fold rate difference between slowly and rapidly evolving dehydrogenase species (Hjelmqvist et al. 1995).

SDR enzyme activities

The substrate spectrum of SDR enzymes with characterized activities is widespread (Table 2). Assuming a function in

Table 2. Clusters of homologous SDR genes of the *H. sapiens*, *C. elegans*, *D. melanogaster* and *A. thaliana* genomes

Cluster	Enzyme annotation	Human sequence identifier	H	D	C	A
Four genomes						
1c	dTDP-D-glucose 4,6 dehydratase	hCP43221	1	1	1	6
2c	UDP-galactose-4-epimerase	GALE_HUMAN	1	1	1	9
3e	GDP-D-mannose dehydratase	GMDS_HUMAN	1	1	2	2
4e	GDP-fucose synthetase	hCP45624	1	1	1	2
5c	HEP27 protein	HE27_HUMAN, O95162	2	1	2	1
6c	follicular variant translocation protein	FVT1_HUMAN	1	1	1	2
7c	17 β -hydroxysteroid dehydrogenase 3	DHB3_HUMAN, hCP50075, Q9Y6G8	3	3	5	2
8c	WWOX protein	Q9NZC7, Q9Y391, hCP39409, ENSP201652, Q9HBH5	5	7	6	7
Three genomes						
9c	dihydropteridine reductase	DHPR_HUMAN	1	1	1	
10c	3-hydroxyacyl-CoA dehydrogenase type II	HCD2_HUMAN	1	1	1	
11c	17 β -hydroxydehydrogenase 4	DHB4_HUMAN	1	1	1	
12c	carbonyl reductase 2 (CBR2)	Q9UHY9	1	1	1	
13c	17 β -hydroxydehydrogenase 8	KE6_HUMAN	1	1	2	
14c	retinal short-chain dehydrogenase/reductase (retSDRI)	O75911, Q9UKU4, hCP34876	3	3	3	
15e	unknown (homologous to dTDP-D-glucose 4,6-dehydratase)	ENSP202124	1		2	4
16e	unknown (distantly related to male sterility like protein 2 from <i>A. thaliana</i>)	hCP48399, hCP40916	2	16	1	
17e	unknown (possibly homologous to L-threonine 3-dehydrogenase)	hCP37118	1	1	1	
18c	unknown	hCP44151	1	1	2	
19c	unknown	Q9Y3A1	1	1	2	
Two genomes						
20e	NAD(P)-dependent steroid dehydrogenase-like protein	NSDL_HUMAN	1			2
21c	mitochondrial 2,4-dienoyl-CoA reductase	DECR_HUMAN	1		3	
22c	carbonyl reductase 1 (CBR1)	DHC3_HUMAN, DHCA_HUMAN	2			6
23c	sepiapterin reductase	SPRE_HUMAN	1	1		
24c	15-OH-prostaglandin dehydrogenase	PGDH_HUMAN	1	1		
25c	retinol dehydrogenase	O43275, RDH1_HUMAN, Q9Y2P9, O75452, hCP51066	5		4	
26e	unknown	hCP1614864, ENSP204270	2			1
27c	unknown	hCP49980	1		3	
28c	unknown	hCP44781	1	10		
29c	unknown	—		4	1	
30c	unknown	Q9Y394	1	1		
31c	unknown	—		1	10	
Human genome only						
32c	dTDP-4-keto-6-deoxy-D-glucose 4-reductase	hCP47659	1			
33e	3- β -hydroxy-5-ene steroid dehydrogenase	3BH1_HUMAN, 3BH2_HUMAN, hCP47581	3			
34c	corticosteroid 11 β -dehydrogenase, isozyme 1	DHI1_HUMAN	1			
35c	corticosteroid 11 β -dehydrogenase, isozyme 2	DHI2_HUMAN	1			
36c	17 β -hydroxydehydrogenase 1	DHB1_HUMAN	1			
37c	17 β -hydroxydehydrogenase 2	DHB2_HUMAN	1			
38c	17 β -hydroxydehydrogenase 7	DHB7_HUMAN	1			
39c	3-hydroxybutyrate dehydrogenase	BDH_HUMAN	1			
40c	trans enoyl oxidoreductase	Q9NY60	1			
41c	retinal dehydrogenase 3	Q9UKU3	1			
42c	all-trans retinol dehydrogenase	hCP43960	1			
43c	putative peroxisomal 2,4-dienoyl-CoA reductase	ENSP168861	1			
44e	unknown	hCP38763	1			
45c	unknown	hCP34509	1			
46c	unknown	hCP50488	1			
47c	unknown	hCP52363	1			

Letter after cluster number denotes subfamily (c-classical SDR, e-extended SDR). Human sequence identifiers refer to the databases Swissprot (Bairoch and Apweiler 2000), TrEMBL (Bairoch and Apweiler 2000), Celera human genome (Venter et al. 2001) or ENSEMBL (<http://www.ensembl.org>). All accession numbers are available at <http://www.cbb.ki.se/sdr>. The numbers in the columns under H, D, C, and A refer to the number of sequences from each species belonging to the respective clusters.

sugar and nucleotide metabolism for extended SDRs, we anticipate a related function for the ancestral SDR progenitor. Along with the appearance of steroid molecules, requiring the presence of oxygen, adaptation and development of classical SDRs with novel substrate specificities might have occurred. In line with this conclusion, it has been noted that bacterial steroid-metabolising enzymes are not only of the SDR type, but also of the cytochrome P450 type (CYP; Nelson 1999), indicating a role for cholesterol derivatives even in prokaryotes. Interestingly, CYP and SDR members do not appear only to complement each other in several distinct pathways of hormones and mediators (New and White 1995), or in xenobiotic metabolism, but they also display broad substrate specificities. Both superfamilies have numerous members in the genomes investigated thus far, indicating their importance for multi-cellular life. The families are of similar sizes, with ~50–80 members in worm and mammals and >130 in *A. thaliana*. Several knockout models or genetic variants have proven the essential role of SDRs in development and homeostasis in humans (Geissler et al 1994; cf. Oppermann et al. 2001 and references therein), insects (Torroja et al. 1998), and plants (DeLong et al. 1993). Along with the determination of novel ligands for orphan receptors (Peet et al. 1998; Chawla et al. 2000), we assume that novel activities and functions for SDRs will be found (Nobel et al. 2001), a task for the next phase of functional genomics in SDR research.

Materials and methods

Data sets

Our data sets consisted of the genome databases for *H. sapiens*, compiled by Celera (February 2001; Venter et al. 2001) and ENSEMBL (version 0.8.0; <http://www.ensembl.org>); *D. melanogaster* (January 2001; Adams et al. 2000); *C. elegans* (Wormpep 39; Wilson 1999); and *A. thaliana* (November 2000; Huala et al. 2001). The two human data sets were made non-redundant using FASTA3 (Pearson and Lipman 1988). Because these data sets are not complete, the KIND database (Kallberg and Persson 1999) was also utilized. To identify the members of the SDR family in the genomes analyzed, we used Hidden Markov Models (HMMs) as implemented in SAM (Karplus et al. 1998). The two groups of SDRs, classical and extended, are too different to be captured by a single HMM. Instead, two models were created, one for each group. Still, when scanning the data sets with these models, we found that it was not possible to define a definite threshold that captures every member and excludes every non-member. For this reason, an expect value $<10^{-10}$ was used, followed by a motif analysis step on the resulting sequence data set. The coenzyme-binding region in the classical SDRs typically contains a TGxxxGxG motif and an NNAG pattern approximately 70 residues downstream. The extended SDRs usually have TGxxGxxG and HxAS patterns instead. These motifs, together with the active site pattern YxxxK, were used to discriminate between SDR members and non-members.

The members identified in the different genomes were clustered using a procedure similar to the one described for genome analysis

(Tatusov et al. 1997). Automated ortholog servers exist, for example, TOGA (Quackenbush et al. 2000) and HomoloGene (Wheeler et al. 2001), clustering sequences at the nucleotide level. In this paper, we focus on the SDR family and cluster the members at the protein level. FASTA3 was used to compare every member against the genome databases. Sequences were clustered according to these comparisons, defining a cluster as a group, in which every member ranks every other member higher than the first non-member. The cluster assignments are reciprocal, that is, usage of either sequence as the query sequence results in the proper sequence detection. These clusters will in most cases correspond to orthologs.

Evolutionary distances

Clusters with protein sequences from all four genomes were aligned using ClustalW (Thompson et al. 1994). The observed distances between the homologs in an alignment were calculated over comparable positions, (i.e., excluding gaps). These distances were then corrected for multiple hits according to Kimura (1983). The evolutionary distances between the homologs were calculated using the divergence times of Doolittle (Feng et al. 1997), 1215 Mya for the plant–animal separation, 1045 Mya for the worm–insect/mammal separation, and 850 Mya for the insect–mammal separation.

Acknowledgments

We thank the Swedish Research Council, the Swedish Foundation for Strategic Research, BioNetWorks GmbH, Munich, and Karolinska Institutet for financial support.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**: 45–48.
- Bancroft, I. 2000. Insights into the structural and functional evolution of plant genomes afforded by the nucleotide sequences of chromosomes 2 and 4 of *Arabidopsis thaliana*. *Yeast* **17**: 1–5.
- Bednarek, A.K., Laflin, K.J., Daniel, R.L., Liao, Q., Hawkins, K.A., and Aldaz, C.M. 2000. WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3–24.1, a region frequently affected in breast cancer. *Cancer Res.* **60**: 2140–2145.
- Chawla, A., Saez, E., and Evans, R.M. 2000. Don't know much bile-ology. *Cell* **103**: 1–4.
- DeLong A., Calderon-Urrea A., and Dellaporta S.L. 1993. Sex determination gene TASSELSEED2 of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. *Cell* **74**: 757–768.
- Feng, D.F., Cho, G., and Doolittle, R.F. 1997. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl Acad. Sci.* **94**: 13028–13033.
- Fransen, M., Van Veldhoven, P.P., and Subramani, S. 1999. Identification of peroxisomal proteins by using M13 phage protein VI phage display: Molecular evidence that mammalian peroxisomes contain a 2,4-dienoyl-CoA reductase. *Biochem. J.* **340**: 561–568.
- Gabrielli, F., Donadel, G., Bensi, G., Heguy, A., and Melli, M. 1995. A nuclear protein, synthesized in growth-arrested human hepatoblastoma cells, is a novel member of the short-chain alcohol dehydrogenase family. *Eur. J. Biochem.* **232**: 473–477.

- 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., and Andersson, S. 1994. Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. *Nat. Genet.* **7**: 34–39.
- Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S., and Duax, W.L. 2001. Porcine carbonyl reductase: Structural basis for a functional monomer in short chain dehydrogenases/reductases. *J. Biol. Chem.* **276**: 18457–18463.
- Hjelmqvist, L., Estenius, M., and Jörnvall, H. 1995. The vertebrate alcohol dehydrogenase system: Variable class II type form elucidates separate stages of enzymogenesis. *Proc. Natl Acad. Sci.* **92**: 10905–10909.
- Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., and Rhee, S.Y. 2001. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **29**: 102–105.
- International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Jörnvall, H., Persson, B., and Jörnvall, H. 1993. Variability patterns of dehydrogenases versus peptide hormones and proteases/antiproteases. *FEBS Lett.* **335**: 69–72.
- Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D. 1995. Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**: 6003–6013.
- Jörnvall, H., Höög, J.-O., and Persson, B. 1999. SDR and MDR: Completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* **445**: 261–264.
- Kallberg, Y. and Persson, B. 1999. KIND — a nonredundant protein database. *Bioinformatics* **15**: 260–261.
- Karplus, K., Barrett, C., and Hughey, R. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14**: 846–856.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, UK.
- Krook, M., Ghosh, D., Strömberg, R., Carlquist, M., and Jörnvall, H. 1993. Carboxyethyllysine in a protein: Native carbonyl reductase/NADP⁺-dependent prostaglandin dehydrogenase. *Proc. Natl Acad. Sci.* **90**: 502–506.
- Lin, B., White, J.T., Ferguson, C., Wang, S., Vessella, R., Bumgarner, R., True, L.D., Hood, L., and Nelson, P.S. 2001. Prostate short-chain dehydrogenase reductase 1 (PSDR1): A new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium. *Cancer Res.* **61**: 1611–1618.
- Nelson, D.R. 1999. Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* **369**: 1–10.
- New, M.I. and White, P.C. 1995. Genetic disorders of steroid hormone synthesis and metabolism. *Baillieres Best Pract. Res. Clin. Endocrinol. Metab.* **9**: 525–554.
- Nobel, S., Abrahamsen, L., and Oppermann, U.H. 2001. Metabolic conversion as a pre-receptor control mechanism for lipophilic hormones. *Eur. J. Biochem.* **268**: 4113–4125.
- Oppermann, U.C., Persson, B., Filling, C., and Jörnvall, H. 1997. Structure–function relationships of SDR hydroxysteroid dehydrogenases. *Adv. Exp. Med. Biol.* **414**: 403–415.
- Oppermann, U.C., Filling, C., and Jörnvall, H. 2001. Forms and functions of human SDR enzymes. *Chem. Biol. Interact.* **130–132**: 699–705.
- Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci.* **85**: 2444–2448.
- Peet, D.J., Janowski, B.A., and Mangelsdorf, D.J. 1998. The LXRs: A new class of oxysterol receptors. *Curr. Opin. Genet. Dev.* **8**: 571–575.
- Persson, B., Krook, M., and Jörnvall, H. 1995. Short-chain dehydrogenases/reductases. *Adv. Exp. Med. Biol.* **372**: 383–395.
- Quackenbush, J., Liang, F., Holt, I., Pertea, G., and Upton, J. 2000. The TIGR gene indices: Reconstruction and representation of expressed gene sequences. *Nucleic Acids Res.* **28**: 141–145.
- Rimokh, R., Gadoux, M., Bertheas, M.F., Berger, F., Garoscio, M., Deleage, G., Germain, D., and Magaud, J.P. 1993. FVT-1, a novel human transcription unit affected by variant translocation t(2;18)(p11;q21) of follicular lymphoma. *Blood* **81**: 136–142.
- Russell, D.W. 2000. Oxysterol biosynthetic enzymes. *Biochim. Biophys. Acta* **1529**: 126–135.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278**: 631–637.
- Torroja, L., Ortuno-Sahagun, D., Ferrus, A., Hammerle, B., and Barbas, J.A. 1998. *scully*, an essential gene of *Drosophila*, is homologous to mammalian mitochondrial type II L-3-hydroxyacyl-CoA dehydrogenase/amyloid- β peptide-binding protein. *J. Cell Biol.* **141**: 1009–1017.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L., and Rapp, B.A. 2001. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **29**: 11–16.
- Wilson, R.K. 1999. How the worm was won. The *C. elegans* genome sequencing project. *Trends Genet.* **15**: 51–58.