## REVIEW ARTICLE Comparative anatomy of the aldo-keto reductase superfamily

Joseph M. JEZ\*, Melanie J. BENNETT\*, Brian P. SCHLEGEL<sup>†</sup>, Mitchell LEWIS\* and Trevor M. PENNING<sup>†</sup><sup>‡</sup> Departments of \*Biochemistry & Biophysics and †Pharmacology, University of Pennsylvania Medical School, 3620 Hamilton Walk, Philadelphia, PA 19104, U.S.A.

The aldo-keto reductases metabolize a wide range of substrates and are potential drug targets. This protein superfamily includes aldose reductases, aldehyde reductases, hydroxysteroid dehydrogenases and dihydrodiol dehydrogenases. By combining multiple sequence alignments with known three-dimensional structures and the results of site-directed mutagenesis studies, we have developed a structure/function analysis of this superfamily. Our studies suggest that the  $(\alpha/\beta)_s$ -barrel fold provides a common scaffold for an NAD(P)(H)-dependent catalytic activity, with substrate specificity determined by variation of loops on the Cterminal side of the barrel. All the aldo-keto reductases are dependent on nicotinamide cofactors for catalysis and retain a similar cofactor binding site, even among proteins with less than 30 % amino acid sequence identity. Likewise, the aldo-keto reductase active site is highly conserved. However, our alignments indicate that variation of a single residue in the active site may alter the reaction mechanism from carbonyl oxidoreduction to

## carbon-carbon double-bond reduction, as in the 3-oxo-5βsteroid 4-dehydrogenases ( $\Delta^4$ -3-ketosteroid 5 $\beta$ -reductases) of the superfamily. Comparison of the proposed substrate binding pocket suggests residues 54 and 118, near the active site, as possible discriminators between sugar and steroid substrates. In addition, sequence alignment and subsequent homology modelling of mouse liver $17\beta$ -hydroxysteroid dehydrogenase and rat ovary 20a-hydroxysteroid dehydrogenase indicate that three loops on the C-terminal side of the barrel play potential roles in determining the positional and stereo-specificity of the hydroxysteroid dehydrogenases. Finally, we propose that the aldo-keto reductase superfamily may represent an example of divergent evolution from an ancestral multifunctional oxidoreductase and an example of convergent evolution to the same active-site constellation as the short-chain dehydrogenase/reductase superfamily.

## INTRODUCTION

Traditionally, oxidoreductases belong to either the long-chain alcohol dehydrogenase or the short-chain dehydrogenase/ reductase (SDR; formerly the short-chain alcohol dehydrogenase) superfamilies. Yeast and liver alcohol dehydrogenases typify the long-chain alcohol dehydrogenases, which use zinc for catalysis [1]. The SDRs are multimeric non-metallooxidoreductases with a conserved Tyr-Xaa-Xaa-Xaa-Lys catalytic motif and a Rossmann fold for NAD(P)(H) binding [2]. These two superfamilies have dissimilar three-dimensional structures and are reliant on either a zinc atom or a catalytic acid/base to perform their reactions [3,4]. An increasing number of oxidoreductases not related to either of these superfamilies have been identified by cDNA cloning; these enzymes are similar to each other, and belong to the aldo–keto reductase (AKR) superfamily [5,6].

The AKRs are monomeric proteins, about 320 residues in size, that bind nicotinamide cofactor without a Rossmann-fold motif [7–10]. Found in mammals, amphibians, plants, yeast, protozoa and bacteria, these proteins metabolize a diverse range of substrates, including aliphatic and aromatic aldehydes, mono-saccharides, steroids, prostaglandins, polycyclic aromatic hydrocarbons and isoflavinoids [6,11–13].

Many of the mammalian AKRs are potential therapeutic targets, and structure-based drug design may lead to compounds

with the desired specificity and clinical efficacy. Aldose reductase (ADR; EC 1.1.1.21) converts circulating glucose into sorbitol, a hyperosmotic sugar, and may play a role in diabetic retinopathy, neuropathy and nephropathy. As such, ADR inhibitors could prove useful as treatments for these complications of diabetes [14]. The clinically tested ADR inhibitors are from two main chemical classes: aryl-substituted acetic acids (ponalrestat and zopolrestat) and acidic cyclic imides, particulary spiro-hydantoins (sorbinil) and spiro-succinimides (alrestatin) [15]. The clinical potential of designed inhibitors has driven the investigation of the structure and function of ADR. Similarly, the mammalian aldehyde reductases (ALRs; EC 1.1.1.2) play a significant role in the metabolism of neurotransmitter aldehydes produced by monoamine oxidase [16], and ALR inhibitors may have anti-depressant properties.

Both  $3\alpha$ -hydroxysteroid dehydrogenase ( $3\alpha$ -HSD; EC 1.1.1.213; formerly EC 1.1.1.50, but renamed due to A-face specific hydride transfer) and prostaglandin F synthase (EC 1.1.1.188) are targets for non-steroidal anti-inflammatory drugs, and may regulate levels of inflammatory prostaglandins [17,18]. Other hydroxyprostaglandin dehydrogenases, involved in the metabolism of bronchoconstrictors, that may yet be assigned to this superfamily could be targets in the treatment of asthma. In addition, the HSDs of this superfamily have the potential to act as molecular switches, by converting potent steroid hormones into inactive metabolites, thereby regulating the amount of

‡ To whom correspondence should be addressed.

Abbreviations used: SDR, short-chain dehydrogenase/reductase (formerly short-chain alcohol dehydrogenase); AKR, aldo-keto reductase; ADR, aldose reductase (sorbitol-6-phosphate dehydrogenase/xylose reductase; EC 1.1.1.21); ALR, aldehyde reductase (EC 1.1.1.2); HSD, hydroxysteroid dehydrogenase; DD, dihydrodiol dehydrogenase (EC 1.3.1.20); FR-1, fibroblast growth factor-1-induced protein.

#### Table 1 Members of the AKR superfamily

The list of AKRs was compiled by screening the GenBank Nucleotide Sequence Database, the Protein Identification Resource Database and the SwissProt Database with the BLAST program [23,24] using the amino acid sequence of rat liver 3 $\alpha$ -HSD as a search query [25]. The AKR superfamily nomenclature indicated in parentheses has been described elsewhere [26].

Designation	Species	Enzyme	Reference
Hum-ALR	Homo sapiens (liver)	ALR (AKR1A1)	42
Por-ALR	Sus scrofa	ALR (AKR1A2)	43
Rat-ALR	Rattus norvegicus (liver)	ALR (AKR1A3)	44
Mou-LKR	Mus musculus (liver)	Ketoreductase (AKR1B10)	41
Hum-ADR	Homo sapiens (placenta)	ADR (AKR1B1)	5, 34
Rab-ADR	Oryctolagus cuniculus	ADR (AKR1B2)	35
Mou-ADR	Mus musculus (kidney)	ADR (AKR1B3)	36
Rat-ADR	Rattus norvegicus (lens)	ADR (AKR1B4)	37
Bov-ADR	Bos taurus (lens/testis)	ADR (AKR1B5)	38, 39
Por-ADR	Sus scrofa (lens)	ADR (AKR1B6)	40
Mou-VDP	Mus musculus	Androgen-dependent vas deferens protein (AKR1B7)	58
Mou-FR1	Mus musculus	FR-1 (AKR1B8)	65
CHO-AKR	Chinese hamster ovary cells	Tripeptide-inducible AKR from CHO cells (AKR1B9)	64
Hum-20aHSD	Homo sapiens (liver)	20x-HSD (AKR1C1) (also DD1)	29, 30
Hum-BABP	Homo sapiens (liver)	Bile acid binding protein (AKR1C2) (also DD2 and $3\alpha$ -HSD type III)	29, 31
Hum-3aHSDII	Homo sapiens (liver)	3α-HSD type II (AKR1C3)	32
Hum-ChlorRed	Homo sapiens (liver)	Chlordecone reductase (AKR1C4) (also DD4 and $3\alpha$ -HSD type I)	31–33
Rab-20aHSD	Oryctolagus cuniculus (ovary)	20x-HSD (AKR1C5)	20
Mou-17bHSD	Mus musculus (liver)	$17\beta$ -HSD (AKR1C6)	27
Bov-Pas	Bos taurus (lung)	Prostaglandin F synthase (AKR1C7)	18
Bov-DD3	Bos taurus (liver)	DD3 (AKR1C11)	56
Rat-20aHSD	Rattus norvegicus (ovary)	$20\alpha$ -HSD (AKR1C8)	28
Rat-3aHSD	Rattus norvegicus (liver)	3α-HSD (AKR1C9) (also DD)	25
Frog-Rho	Rana catesbeiana	ρ-Crystallin (AKR1C10)	57
Hum-3o5bred	Homo sapiens (liver)	3-Oxo-5 $\beta$ -steroid 4-dehydrogenase (AKR1D1)	48
Rat-3o5bred	Rattus norvegicus (liver)	3-Oxo-5 $\beta$ -steroid 4-dehydrogenase (AKR1D2)	49
Bar-ALR	Hordeum vulgare	ALR (AKR1E1)	45
Brgr-ALR	Bromus inermis	ALR (AKR1E2)	46
App-S6Pdh	Malus domestica	Sorbitol phosphate dehydrogenase (AKR2A)	59
Pic-XylRed	Pichia stipitis	Xylose reductase (AKR2B1)	60
Klu-XylRed	, Kluyveromyces lactis	Xylose reductase (AKR2B2)	61
Muc-TADH	Mucor mucedo	4-Dihydromethyltrisporate dehydrogenase (AKR2C)	62
Spor-ALR	Sporidiobolus salmonicolor	ALR (AKR3A)	47
Sac-GCY	, Saccharomyces cerevisiae	GCY protein (AKR3B)	63
Alf-ChalRed	Medicago sativa	Chalcone reductase (AKR4A1)	50
Soy-ChalRed	Glycine max	Chalcone reductase (AKR4A2)	12
Gly-PKR	Glycyrrhiza echinata	Polyketide reductase (AKR4A3)	51
Lei-putRed	Leishmania maior	Putative reductase (AKR5A)	55
Ps-Mordh	Pseudomonas putida	Morphine dehvdrogenase (AKR5B)	54
Cb-25dkg	Corynebacterium sp.	2,5-Diketo-p-gulonate reductase (AKR5C)	52
Cb2-25dka	Corvnebacterium sp.	2.5-Diketo-D-gulonate reductase (AKR5D)	53
Rat-AFBred	Rattus norvegicus (liver)	Aflatoxin B1 ALR (AKR7)	66

hormone that can bind and activate nuclear receptors [19,20]. This role has been well documented for type 2 11 $\beta$ -HSD (EC 1.1.1.146) and type 1 17 $\beta$ -HSD (EC 1.1.1.62), which regulate the occupancy of mineralocorticoid and oestrogen receptors respectively. Although these HSDs belong to the SDR superfamily, HSDs of the AKR superfamily could perform a similar function. Thus therapeutics directed against tissue-specific HSDs may modulate occupancy of steroid-hormone receptors.

Zopolrestat and ponalrestat bind to ADR with nanomolar affinity, but are weak inhibitors of  $3\alpha$ -HSD, a member of the AKR superfamily [21]. This suggests that the design of inhibitors that are specific for members of the superfamily can be achieved. However, there has been little further progress towards this goal. Before targeting individual AKRs for rational drug design, an understanding of the relationship between structure and function in the superfamily is required.

In this review, we demonstrate the high degree of amino acid sequence similarity among the AKRs, and address how studies of  $3\alpha$ -HSD, ADR and ALR provide insight into the structure/

function relationships across the superfamily. Amino acid comparisons suggest that proteins of the superfamily share a common  $(\alpha/\beta)_{s}$ -barrel three-dimensional fold and have a highly conserved nicotinamide-cofactor-binding pocket. While the AKR active site is also conserved both structurally and in sequence, our alignments suggest that a single amino acid change in this site may alter catalytic activity from carbonyl oxidoreduction to carbon-carbon double-bond reduction. In addition, our work indicates that the three loops on the C-terminal side of the structure are vital in discriminating between different substrates of the AKRs, especially for the HSDs of the superfamily. Finally, we raise the possibility of functional divergence within the superfamily from a common ancestral oxidoreductase, and of convergent evolution to the same catalytic mechanism as the SDR superfamily. The concept that a common ancestral oxidoreductase has undergone divergent evolution in the SDR superfamily with convergence to a common reaction mechanism (e.g. type I–IV 17 $\beta$ -HSD) has been raised previously [22]. Since both AKRs and SDRs contain alternating arrangements of *a*-helix

and  $\beta$ -sheet, albeit with different structural motifs, these proteins may have arisen from a now extinct common structural ancestor [2].

## THE AKR SUPERFAMILY

A total of 42 oxidoreductases have been identified as members of the AKR superfamily (Table 1), including seven HSDs [five of which are also dihydrodiol dehydrogenases (DDs; EC 1.3.1.20)] [20,25,27–33], seven ADRs [5,34–41], six ALRs [5,42–47], two 3-oxo-5 $\beta$ -steroid 4-dehydrogenases ( $\Delta^4$ -3-ketosteroid 5 $\beta$ reductases) (ED 1.3.99.6) [48,49], three plant chalcone reductases [12,50,51], two 2,5-diketo-gulonate reductases from *Corynebacterium* [52,53], *Pseudomonas* morphine dehydrogenase (EC 1.1.1.218) [54], a putative reductase from *Leishmania* [55], human liver bile acid binding protein [29,31], bovine lung prostaglandin F synthase [17], bovine DD3 [56], frog lens  $\rho$ -crystallin [57], a mouse vas deferens protein [58], apple sorbitol-6-phosphate reductase [59], two xylose reductases from yeast [60,61], 4-dihydromethyltrisporate dehydrogenase from *Mucor mucedo* [62], yeast GCY protein [63], an inducible protein from Chinese hamster ovary cells [64], a murine protein induced by fibroblast growth factor-1 (FR-1) [65], and an ethoxyquin-inducible aflatoxin B1 aldehyde reductase [66]. It should be noted that the  $\beta$ -subunits of rat and bovine *Shaker* K<sup>+</sup> channels may be distantly related to the AKRs, but were not included in this analysis, since they do not exhibit any known catalytic activity [67].

Cluster analysis of the AKRs indicates that these proteins form six distinct groups within the superfamily (Figure 1). The most studied, and largest, group consists of the plant ALRs and the mammalian HSDs, ALRs and ADRs. Other families include proteins from yeast, prokaryotes and protozoa that catalyse reactions with substrates similar to those of ADR and ALR, but these proteins differ significantly from the latter enzymes in amino acid sequence. Three groups of the superfamily include proteins produced as a response to exogenous factors. For example, among the ADR family, FR-1 is induced by fibroblast growth factor [65], the mouse vas deferens protein displays



## Figure 1 Cluster analysis of the AKR superfamily

The dendrogram is based on pairwise sequence alignments, and indicates the relative percentage amino acid identity among proteins of the AKR superfamily. Abbreviation: BABP, bile acid binding protein.

androgen-dependent expression [58], and an inducible AKR from Chinese hamster ovary cells has recently been cloned [64]. In addition, the plant chalcone reductases synthesize isoflavinoid phytoalexins in conjunction with chalcone synthase as an induced response to pathogen attack [12], while metabolism of the dialdehyde of aflatoxin B1 occurs through an ethoxyquin-inducible ALR [66].

## A COMMON THREE-DIMENSIONAL STRUCTURE

Since members of the AKR superfamily show amino acid sequence identity, it is likely that they retain a common threedimensional fold. The structures of human placental and bovine muscle ADRs, rat liver  $3\alpha$ -HSD, pig muscle ALR and murine FR-1 adopt an identical  $(\alpha/\beta)_8$ -barrel fold, as determined by X-ray crystallography [7–10,69]. These structures have eight parallel  $\beta$ -strands, and each  $\beta$ -strand alternates with an  $\alpha$ -helix running anti-parallel to the strand, thereby forming the staves of the barrel (Figure 2).

Strict conservation of amino acid sequence occurs at 11 positions in the primary structure of all the AKRs: Gly-22, Gly-45, Asp-50, Lys-84, Asp-112, Pro-119, Gly-164, Asn-167, Pro-186, Gln-190 and Ser-271 (Figure 3). (The numbering of amino acid residues throughout this review corresponds to that of rat liver  $3\alpha$ -HSD.) Also, an additional eight positions are conserved in 41 of the AKRs: Gly-20, Tyr-55, Gly-62, Leu-113, Trp-148, Gly-158, Glu-192 and Arg-276. Five of these residues function in cofactor binding (Asp-50, Asn-167, Gln-190, Ser-271 and Arg-276), and three form parts of the active site (Asp-50, Tyr-55 and Lys-84). The remaining residues may play structural roles in forming the barrel core, since they are found within the  $\beta$ strands,  $\alpha$ -helices and short loops of the barrel. This pattern of conservation is not unusual. In other  $(\alpha/\beta)_{s}$ -barrel proteins, recurrent short loop structures between the  $\alpha$ -helices and the  $\beta$ strands of the barrel are thought to maintain the general size and ellipticity of the barrel [70-72].

Since the AKRs consist of approximately the same number of residues, we weighted the multiple alignments against gap insertion. The resulting multiple sequence alignment contained no major gaps in regions that correspond to the components of the barrel structure, further supporting a conserved threedimensional fold for the superfamily. However, major gaps in the alignments appear in the sequences of three large loops on the C-terminal side of the barrel (Figures 2 and 3). The AKRs seemingly maintain the architecture of the barrel scaffold while tailoring substrate specificity through modification of the loops near the active site, as discussed below.

#### **KINETIC AND CATALYTIC MECHANISM**

The AKRs catalyse oxidation and reduction reactions on a range of substrates using NAD(P)(H) as cofactor. An ordered Bi Bi kinetic mechanism, in which cofactor binds first and leaves last, has been demonstrated for ALR, ADR,  $3\alpha$ -HSD and 3-oxo- $5\beta$ steroid 4-dehydrogenase, and may be a hallmark feature of other AKRs [73–76]. The reduction reaction involves 4-*pro-R* hydride transfer from NAD(P)H to the substrate carbonyl and protonation of the oxygen by a residue of the enzyme acting as



Figure 2 A common three-dimensional fold for the AKRs

The ribbon drawing of the three-dimensional structure of rat liver  $3\alpha$ -HSD shows the common  $(\alpha/\beta)_8$ -barrel motif of the AKRs, with the C-terminus at the top. The  $\alpha$ -helices and  $\beta$ -sheets (in pink) of the barrel are indicated, as are two helices (H1 and H2) not in the barrel. In addition, the A-loop, B-loop and C-terminal loop are indicated. Reproduced with permission from Hoog, Pawlowski, Alzari, Penning and Lewis (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 2517–2521 [9]; copyright (1994) National Academy of Sciences, U.S.A.

a general acid. The reverse occurs in the oxidation reaction. Whether this reaction has a concerted mechanism or two distinct steps, with either hydride or proton transfer occurring first, remains unclear. As discussed below, the  $5\beta$ -reductases of the superfamily represent a variation on this general mechanism, and reduce a carbon–carbon double bond instead of a carbonyl group.

The structures of ADR, ALR,  $3\alpha$ -HSD and FR-1 indicate spatial conservation of the active-site residues Asp-50, Tyr-55, Lys-84 and His-117 (Figure 4) [7-10,69]. In addition, a water molecule is consistently bound between Tyr-55, His-117 and the C-4 of bound NADP+, an arrangement thought to mimic the location of the substrate carbonyl. The spatial arrangement of Tyr-55, His-117 and the nicotinamide ring has been proposed to create an oxyanion hole. Based on the original ADR · NADPH binary complex structure, Wilson and co-workers [7] suggested that either Tyr-55 or His-117 is the proton donor, based on the proximity required between the C-4 position of the nicotinamide ring and the anticipated position of the substrate carbonyl. Moreover, Asp-50 and Lys-84 were proposed to form a hydrogenbond/salt-bridge network that could decrease the  $pK_a$  of the tyrosine and facilitate proton donation. These authors also noted that the hydrophobic environment of His-117 would depress the  $pK_{a}$  of the imidazole side chain, making it less likely to function as the proton donor at physiological pH.

These structures laid the basis for site-directed-mutagenesis studies of the catalytic mechanism of ADR, ALR and  $3\alpha$ -HSD

#### Figure 3 Multiple sequence alignment of the AKR superfamily

The multiple sequence alignment used PILEUP from the GCG program suite [68] with the gap and the gap length penalties set to 12 and 4 respectively. The secondary structure of rat liver  $3\alpha$ -HSD is noted above the alignment [9]. B1, B2, H1 and H2 are  $\beta$ -sheets and  $\alpha$ -helices not forming the core  $(\alpha/\beta)_8$ -barrel structure. Also, residues corresponding to three large loops on the C-terminal side of the barrel are noted. Invariant residues in all AKRs are highlighted in pink (see the text). Abbreviations for the proteins are given in Table 1. Modified from Jez et al. [26]; reproduced with the permission of Plenum Press.

6	20
υ	23

		19	P0 81		<b>61</b>	e	10	<b>a</b> 2		83		<b>a</b> 3	84	10	0.0. 3
Hum_ALR Por ALR		-MAASCVLLH	TGQKMPLIGL	GTWKSEP	. GOVKAAVKY	AL.SVGYRHI	DCAAIYGNEP DCAAIYGNEL	EIGEALKEDV	GPGKAVPREE GPGKAVPREE	LFVTSKLW	.NTKHHPEDV	EPALRKTLAD	LQLEYLDLYL LOLEYLDLYL	MHWPYAFE	RGDNPFPK
Rat_ALR Mou_LKR		-MTASSVLLH	TGQKMPLIGL ~MENIPTVGL	GTWKSEP	.GOVKAAIKY .GEVTDAVKL	AL.SVGYRHI AI.NLGYRHF	DCASVYGNET DCAYLYHNES	EIGEALKESV EVGMGISEKI	GAGKAVPREE .KEGVVKRED	LFVTSKLW LFVVSKLW	.NTKHHPEDV .CTCHKKSLV	EPAVRKTLAD KTACTNTLEA	LQLEYLDLYL LNLDYLDLYL	MHWPYAFE IHWPIGFK	RGDNPFPK PGEKDIPL
Hum_ADR Rab_ADR		~~MASRLLLN ~~MATHLVLY	NGAKMPILGL NGAKMPILGL	GTWKSPP GTWKSPP	. GQVTEAVKV . GQVTEAVKT	AI.DV <mark>G</mark> YRHI AI.DL <mark>G</mark> YRHI	DCAHVYQNEN DCAHVYQNEN	EVGVAIQEKL EVGVALQEKL	. REQVVKREE . KEQVVKREE	LFIVSKLW LFIVSKLW	.CTYHEKGLV .CTSHDKSLV	KGACQKTLSD KGACQKTLND	LKLDYL <b>D</b> LYL LKLDYL <b>D</b> LYL	IHWPTGFK IHWPTGFK	PGKEFFPL HGSEYFPL
Mou_ADR Rat_ADR	~~~~~~	~-MASHLELN ~-MASHLELN	NGTKMPTLGL	GTWKSPP	. GQVTEAVKV . GQVTEAVKV	AI.DLGYRHI AI.DMGYRHI	DCSQVYQNEK DCAQVYQNEK	EVGVALQEKL EVGVALQEKL	. KEQVVKRQD . KEQVVKRQD	LFIVSKLW LFIVSKLW	. CTFHDKSMV . CTFHDQSMV	KGAFQKTLSD KGACQKTLSD	LQLDYLDLYL LQLDYLDLYL	IHWPTGFK IHWPTGFK	PGPDYFPL PGPDYFPL
Por_ADR Mou_VDP		~-MASHLVLY	TGAKMPILGL TKAKMPILGL	GTWKSPP GTWKSPP	. GKVTEAVKV . GKVTEAVKV	AI.DLGYRHI AI.DLGYRHI AI.DLGYRHI	DCAHVYQNEN DCAHVYQNEN	EVGLALQAKL	.QGQVVKRED KENAVKRED	LFIVSKLW LFIVSKLW	. CTTHEKNLV . CTDHEKNLV	KGACQATLSD	LKLDYLDLYL	IHWPTGFK	PGKDPFPL AGNALLPK
Mou_FR1 CHO_AKR		~-MATFVELS ~-MSTFVELS	TKAKMPIVGL	GTWKSPP	NOVKEAVKA	AI.DAGYRHI AI.DAGYRHI	DCAYAYCNEN DCAYAYYNEH	EVGEAIQEKI	. KEKAVQRED . KEKAVRRED	LFIVSKLW	. PTCFEKKLL . PTCFERKLL	KEAFQKTLTD KEAFQKTLTD	LKLDYLDLYL	IHWPQGLQ IHWPQGLQ	PGKELFPK
Hum_20aHSD Hum_BABP	M	DSKYQCVKLN DSKYQCVKLN	DGHFMPVLGF DGHFMPVLGF	GTYA.PAEVP GTYA.PAEVP	KSKALEAVKL KSKALEATKL	AI.EA <mark>G</mark> FHHI AI.EA <mark>G</mark> FRHI	DSAHLYNNEE DSAHVYNNEE	QVGLAIRSKI QVGLAIRSKI	A.DGSVKRED A.DGSVKRED	IFYTS <mark>K</mark> LW IFYTS <mark>K</mark> LW	. SNSHRPELV . CNSHRPELV	RPALERSLKN RPALERSLKN	LQLDYVDLYL LQLDYVDLYL	IHFPVSVK IHFPVSVK	PGEEVIPK PGEEVIPK
Hum_3aHSD2 Hum_ChlorRed	~~~~M	DSKQQCVKLN ~PKYQRVELN	DGHFMPVLGF DGHFMPVLGF	GTYA.PPEVP GTYA.PPEVP	RSKALEVTKL RNRAVEVTKL	AI.EAGFRHI AI.EAGFRHI	DSAHLYNNEE DSAYLYNNEE	QVGLAIRSKI QVGLAIRSKI	A.DGSVERED A.DGSVKRED	IFYTSKLW IFYTSKLW	.STFHRPELV .CTFFQPQMV	RPALENSLKK QPALESSLKK	AQLDYVDLYL LQLDYVDLYL	IHSPMSLK LHFPMALK	PGEELSPT PGETPLPK
Rab_20aHSD Mou_17bHSD	M	DPKFQRVALS DSKQQTVRLS	DGHFIPVLGF	GTYA.PEEVP GTYA.PQEVP	KSKAMEATKI KSKATEATKI	AI.DAGFRHI AI.DAGFRHI	DSAYFYKNEK DSASMYQNEK	EVGLAIRSKI EVGLAIRSKI	A. DGTVKRED A. DGTVKRED	IFYTSKLW IFYTSKVW	. CTFHRPELV . CTFHRPELV	RVCLEQSLKQ	LOLDYVDLYL	IHFPMAMK	PGVEIIPT PGENYLPK
Bov_DD3 Rat 20aHSD	M	DPKGQRVKLN NSKIOKMELN	DGHFIPVLGF DGHSIPVLGF	GTFA.PREVP GTYA.TEENL	KSEALEVTKF	AI.EAGFRHI AI.DVGFRHI	DSAHLYQNEE DCSHLYONEE	QVGQAIRSKI EIGOAIVSKI	A. DGTVKRED E. DGTVKRED	IFYTSKLW IFYTSKLW	. STSLRPELV	RPALEKSLNN RPSLENSLRK	LQLDYVDLYI	IHFPVALK	PGETLFPT PGDELLPQ
Rat_3aHSD Frog_Rho	M MT	DSISLRVALN LTKETRVTLN	DGNFIPVLGF DGNMMPILGL	GTTV.PEKVA GTYA.APDVP	KDEVIKATKI KSLAEEAVKT	AI.DN <mark>G</mark> FRHF AI.DV <mark>G</mark> YRHI	DSAYLYEVEE DCAFITGNEM	EVGQAIRSKI HIGNGIRSKI	E.DGTVKRED S.DGTVERED	IFYTS <mark>K</mark> LW IFYTG <mark>K</mark> LW	.STFHRPELV .CTYFSPDMV	RTCLEKTLKS RKGLERSLRD	TQLDYV <mark>D</mark> LYI VGMDYL <b>D</b> LFL	IHF <mark>P</mark> MALQ MHW <b>P</b> VSLK	PGDIFFPR PSGASDPS
Hum_3o5bRed Rat_3o5bRed	MDL	SAASHRIPLS STANHHIPLN	DGNSIPIIGL	GTYSEPKSTP GTYSDPRPVP	KGACATSVKV .GKTFIAVKT	AI.DTGYRHI AI.DEGYRHI	DGAYIYQNEH DGAYVYRNEH	EVGEAIREKI EVGEAIREKV	A. EGKVRRED A. EGKVKREE	IFYCGKLW IFYCGKLW	. ATNHVPEMV . STDHDPEMV	RPTLERTLRV RPALERTLQT	LQLDYVDLYI	IEVPMAFK	PGDEIYPR PGEEFYPK
Bar_ALR Brgr_ALR App S6Pdb	~MASAKATMG ~MASAKAMMG	QERQDHFVLK QERQDHFVLK ~~MSTVTLS	SGHAMPAVGL SGHAIPAVGL	GTWRA GTWRA	GSDTAHSVRT GSDTAHSVQT KEVILNAIKI	AITEAGYRHV AITEAGYRHV	DTAAEYGVEK DTAAEYGVEK	EVGKGLKAAM	EA.G.IDRKD EA.G.IDRKD KTGLVKREE	LFVTSKLW LFVTSKLW	.CTDLVPDRV	RPALENTLKD RPALEKTLKD VEACKNSLEK	LOIDYLDLYL	IHWPFRLK VHYPMPTKHN	DGAHK . PPEA
Pic_XylRed Klu_XylRed	MT	~~MPSIKLN. YLAETVTLN.	SGYDMPAVGF	GCWKVDVDTC GCWKMPNDVC	SEQIYRAIKT	GYRLF	DGAEDYANEK	LVGAGVKKAI EVGOGVNRAI	. DEGIVKRED . KEGLVKRED	LFLTSKLW LVVVSKLW	. NNYHHPDNV . NSFHHPDNV	EKALNRTLSD	LQVDYVDLFL LQLDYVDIFY	IHFPVTFKFV IHFPLAFKPV	PLEEKYPPGF PFDEKYPPGF
Muc_TADH Sac_GCY	MPATL	MSTDYLTLNR HDSTKILSLN	TGDKMPIRGF TGAQIPQIGL	GCWKIDTKDC GTWQSKE	EETVYQAIKT .NDAYKAVLT	<mark>G</mark> YRLF AL.KD <mark>G</mark> YRHI	DGACDYGNEV DTAAIYRNED	EVGRGINKAI QVGQAI	. NEGLVKRED . KDSGVPREE	LFIVT <mark>K</mark> LW IFVTTKLW	.NTFHSKKHV .CTQHHEPEV	RALFDRQLKD ALDQSLKR	TGLEYF <b>D</b> LYL LGLDYV <b>D</b> LYL	IHF <b>P</b> VPLQYV MHW <b>P</b> ARLD	DPATVYPPGW PAYIKNEDIL
Sporo_ALR Alf_ChalRed	~~~~MGSVE	~-MVGTTTLN IPTKVLTNTS	TGASLELVGY SQLKMPVVGM	GTWQAAP GSAPDFTC	.GEVGQGVKV KKDTKDAIIE	AI.ETGYRHL AIKQ.GYRHF	DLAKVYSNQP DTAAAYGSEQ	EVGAAI ALGEALKEAI	. KEAGVKRED EL.GLVTRDE	LFITSKLW LFVTSKLW	. NNSHRPEQV . VTENHPHLV	EPALDDTLKE IPALQKSLKT	LGLEYLDLYL LQLDYLDLYL	IHWPVAFP IHWPLSSQ	PEGDITQNLF PGKFSFPIDV
Gly_PKR	MAAAAAAAIE	IPTIVEPNSS IPTKVLPNST GUDKAMUTLS	CELRVPVIGM	GS. APDFTC	KKDTKEAIIE	AIKQ.GYRHF	DTAAAYGSEQ DTAAAYGSET	ALGEALKEAR	DL.GLVSKQD DL.GLVTREE	LFVTSKLW LFVTSKLW	. VTENHPHLV . VTENHPHLV	IPALRKSLET LAAFFESROK	LOVEYLDLYL	IHWPLSSQ IHWPLSSQ IHWPRG	PGKFSFPIQV
Ps_Mordh Cb_25dkg	M	AGKSPLINLN ~MTVPSIVLN	NGVKMPALGL DGNSIPQLGY	GVF.AAS GVFKVPP	AEETASAIA	SAISSGYRLI EALEVGYRHI	DTARSYNNEA DTAAIYGNEE	QVGEGIRNS. GVGAAIAAS.	VDRAE	MFVTTKLF LFITTKLW	.NCDYGYERA .NDRHDGDEP	LRAFDESLGR AAAIAESLAK	LGLDYVDLYL LALDQVDLYL	LHWP	
Cb_25dkg2 Rat_AFBRed	~~~~~	MPNIPTISLN ~~~~~MSQAR	DGRPFAEPGL PATVLGAMEM	GTYNL.R GRRMDVTSSS	.GDEGVAAMV ASVRAFLQ	AAIDS <mark>G</mark> YRLL RGHTEI	DTAVNYENES DTAFVYANGQ	EVGRAVRAS.	SVDRDE GLGLGRSGCK	LIVAS <mark>K</mark> IP VKIAT <b>K</b> AAPM	. GRQHGRAEA FGKTLKPADV	VDSIRGSLDR RFQLETSLKR	LGLDVI <b>D</b> LQL LQCPRV <b>D</b> LFY	IHWP LHFPDHGT	PIEETLQ
	LCOP	<u>A</u>	<u>α4</u>	β5	α5		β6		α6	<b>β</b> 7		L	OOP B		<u>H1</u>
Hum_ALR Por_ALR		TICYDSTHYK TIRYDATHYK	ETWKALEALV DTWKALEALV	AKGLVQAL <mark>G</mark> L AKGLVRAL <b>G</b> L	SNFNSRQIDD SNFSSRQIDD	ILSVASV. VLSVASV.	. RPAVLQVEC . RPAVLQVEC	HPYLAQNELI HPYLAQNELI	AHCQARGL AHCQARGL	EVTAYSPLGS EVTAYSPLGS	S.DR S.DR			AWRDP	DEPVLLEEPV NEPVLLEEPV
MOU_LKR	DRNG	KVIPSHTSFL	DTWEAMEDLV	FEGLVKALGU	SNFSSRQ1DD SNFNHEQLER	L.LDKPGLR	VRPITNQIEC	HPYLNQKKLI HPYLNQKKLI	DF. CHKRNV	SVTAYSPLGS	S.GG P.DR			PWAKP	EDPSLLEDPR
Rab_ADR Mou_ADR	DAAG	NVIPSDTDFL NVIPSDTDFV	DTWEAMEGLV	DEGLVKSIGV	SNFNHLQIER SNFNPLOIER	ILNKPGLK	YKPAVNO IEC YKPAVNO IEC	HPYLTQEKLI HPYLTQEKLI	QYCHSKGI EYCHSKGI	VVTAYSPLGS VVTAYSPLGS	P.DR P.DR			PWAKP	EDPSLLEDPR
Rat_ADR Bov_ADR	DASG	NVIPSDTDFV NVIPSEKDFV	DTWTAMEQLV DTWTAMEELV	DEGLVKAIGV DEGLVKAIGV	SNFNPLQIER SNFNHLQVEK	ILNKPGLK ILNKPGLK	YKPAVNQIEC YKPAVNQIEC	HPYLTQEKLI HPYLTQEKLI	EYCHCKGI QYCNSKGI	VVTAYSPLGS VVTAYSPLGS	P.DR P.DR			PWAK P	EDPSLLEDPR EDPSILEDPR
Por_ADR Mou_VDP	DGDG	NVVPDESDFV KVLLSKSTFL	ETWEAMEELV DAWEAMEELV	DEGLVKAIGV DQGLVKALGI	SNFNHLQVEK SNFNHFQIER	ILNKPGLK LLNKPGLK	YKPAVNQIEV HKPVTNQIES	HPYLTQEKLI HPYLTQEKLI	EYCKSKGI QYCQSKGI	VVTAYSPLGS AVTAYSPLGS	P.DR P.DR			PWAKP	EDPSLLEDPR EDPVVMEIPK
CHO_AKR Hum 20aHSD	DDQG	NVLTSKITFL	DAWEVMEELV	DEGLVKALGV	SNFNHFQIER SNFNHFQIER	I.LNKPGLK	HKPVTNQVEC	HPYLTQEKLI HPYENORKLI	EY. CHSKGI	TVTAYSPLGS	P.NR			PWAKP	EDPSLLEDPK
Hum_BABP Hum_3aHSD2	DENG	KILFDTVDLC KVIFDIVDLC	ATWEAVEKCK TTWEAMEKCK	DAGLAKSIGV DAGLAKSIGV	SNFNRRQLEM SNFNRRQLEI	I.LNKPGLK	YKPVCNQVEC YKPVCNQVEC	HPYFNQRKLL HPYFNRSKLL	DFCKSKDI DFCKSKDI	VLVAYSALGS VLVAYSALGS	HREE QRDK			PWVDP	NS PVLLEDPV NS PVLLEDPV
Hum_ChlorRed Rab_20aHSD	DENG	KVIFDTVDLS KAIFDTVDIC	ATWEVMEKCK ATWEAMEKCK	DAGLAKSI <mark>G</mark> V DAGLAKSI <mark>G</mark> V	SNFNCRQLEM SNFNRRQLEM	ILNKPGLK ILNKPGLK	YKPVCNQVEC YKPVCNQVEC	HPYLNQSKLL HPYLNQGKLL	DFCKSKDI EFCKSKGI	VLVAHSALGT VLVAYSALGS	QRHK			LWVDP	NSPVLLEDPV SAPVLLEDPL
Mou_17bHSD Bov_PGS Bow_DD3	DENG	KLIYDAVDIC KLIFDSVDLC	DTWEAMEKCK HTWEALEKCK	DAGLAKSIGV DAGLTKSIGV	SNFNRRQLEK SNFNHKQLEK	ILKKPGLK ILNKPGLK	YKPVCNQVEC YKPVCNQVEC	HPYLNQGKLL HPYLNQSKLL	DF. CRSKDI EF. CKSHDI	VLVAYSALGS VLVAYAALGA	HREK			QWVDQ	SSPVLLDNPV NNPVLLEDPV
Rat_20aHSD Rat_3aHSD	DEHG	NLILDTVDLC KLLFETVDIC	DTWEAMEKCK	DAGLAKSIGV	SNFNRRQLEK SNFNRRQLEK	I.LNKPGLK	HRPVCNOVEC	HLYLNQSKLL HLYLNQSKLL	AY. CKMNDI	VLVAYGALGT TLVSYCTLGS	QRYK			YCINE	DTPVLLDDPI
Frog_Rho Hum_3o5bRed	DKDK	PFIYDNVDLC KWLYHKSNLC	ATWEALEACK ATWEAMEACK	DAGLVRSLGV DAGLVKSLGV	SNFNRRQLER SNFNRROLEL	ILNKPGLK	YKPVCNQVEC HKPVSNOVEC	HVYLNQNKLH HPYFTOPKLL	AYCKSKDI KFCOOHDI	VLVAYSVLGS VITAYSPLGT	HRDR SRNP			NWVDL	SLAVLLDDPI SSPPLLKDAL
Rat_3o5bRed Bar_ALR	GEVLEFDMEG	RVIYHKSNLC	ATWEALEACK .VWKEMENLV	DAGLVKSL <mark>G</mark> V KDGLVKDI <mark>G</mark> V	SNFNRRQLEV CNYTVTKLNR	ILNKPGLK LLRSAK	YKPVTNQVEC IPPAVCQMEM	HPYFTQTKLL HPGWKNDKIF	EV.SASSMTS EACKKHGI	FIVAYSPLGT HVTAYSPL	CRNP			LWVNV GS	SSPPLLKDEL SEKNLAHDPV
App_S6Pdh Ric YylRed	TASLLGEDKV	LDIDVTISLQ	QTWEGMERTV	KDGLVKDIGV SLGLVRSIGL	CNYTVTKLNR SNYELFLTRD	LLQSAK CLAYSK	IAPAVCOMEM IKPAVSOFET	HPGWKNDKIL HPYFQRDSLV	EACKKHGI KFCMKHGV	HATAYSPL LPTAHTPLGG	AA			CS	SEKNLAHDPV GSVSPLDDPV
Klu_XylRed Muc_TADH	YTGKEDEAKG YVGDAK	HIEEEQVPLL	DTWRALEKLV	DQGKIKSLGI	SNFPGALLLD SNFSGALIQD	L.LR.GAT L.LR.GAR	IKPSVLQVEH IKPVALQIEH	HPYLTQERLI HPYLTQERLI	KY. VKNAGI	QVVAYSSFGP OTTAYSSFGP	VSFL			ELENKKAL	NTPTLFEHDT YTS.LLEHAS
Sac_GCY Sporo_ALR	SVPTKKDGSR PKANDKEVKL	AVDITNWNFI DLEVSLV	KTWELMQELP DTWKAMVKLL	KTGKTKAVGV DTGKVKAIGV	SNFSINNLKD SNFDAKMVDA	LLASQGNK IIEATG	LTPAANQVEI VTPSVNQIER	HPLLPQDELI HPLLLQPELI	NFCKSKGI AHHKAKNI	VVEAYSPLGS HITAYSPLGN				T	DAPLL.KEPV GAPLLVQHPE
Alf_ChalRed Soy_ChalRed	ADLLPFDVKG EDLLPFDVKG		. VWESMEESL . VWESMEECQ	KLGLTKAI <b>G</b> V KLGLTKAI <b>G</b> V	S <mark>N</mark> FSVKKLEN S <mark>N</mark> FSVKKLQN	LLSVAT LLSVAT	VLPAVNQVEM IRPVVDQVEM	NLAWQQKKLR NLAWQQKKLR	EFCNAHGI EFCKENGI	VLTAFSPL IVTAFSPL				RKGASR	GPNEVMENDM GPNEVMENDV
Lei_putRed	KD	ILSKEGKKYL	. VWESMEECL DSWRAFEQLY	KECKEVRAIGV	SNFSVKKLQN SNFHIHHLED	LLSVAT VLAMCTVT	IRPAVNOVEM	NLAWQQKKLR HPLNNQADLR	EFCTANGI AFCDAKQI	VLATFSPL KVEAWSPLGQ	NORW			G	KLLSNPI
Cb_25dkg Cb_25dkg2		TPAADNYV NPSVGRWL	HAWEKMIELR	AAGLTRSIGV	SNHLVPHLER SNFTEPMLKT	I UAATGVV	PAVNQIEL	HPAYQQREIT	DW. AAAHDV AF. HDEHGI	KIESWGPLGQ RTESWSPLAR	ngrw				KYD.LFGAEP RSE.LLTEQL
Rat_AFBRed			.ACHHVH	QEGKFVEL <mark>G</mark> L	SNYVSWEVAE	ICTLCKKNGW	IMPTVYQGMY	NAITROVETE	LFPCLRHFGL	RFYAFNPLAG	GLLTGRYKYQ	DKDGKNPESR	FFGNPFSOLY	MDRYWKEEHF	NGIALVEKAL
	H1	<b>67</b>		ßa	<b>C</b> <sup>2</sup>		10			008.0					
Hum_ALR	VLALAEKYGR VOALAEKYNR	SPAQILLENQ	vo	RKVICIPKSI	TPSRILQNIK	VFD.FTFSPE	EMKQLNALNK	NWRYIVPMLT	VDGKRVPRDA	GHPLYPFNDP GHPLYPFNDP	Y				
Rat_ALR Mou_LKR	VLALAEKHGR IRKIAKKHGK	SPAQILLRWQ SPAQILIRFQ	VQ IQ	RKVICIPK <mark>S</mark> I RNLIVIPK <mark>S</mark> V	TPSRILQNIQ TPSRIRENIQ	VFD.FTFSPE VFD.FELTEK	EMKQLDALNK DMEELLSLDK	NWRYIVPMIT NLRFATFP	VDGKRVPRDA	GHPLYPFNDP NHQDYPFHIE	Ү Ү				
Hum_ADR Rab_ADR	IKAIAAKHNK IKAIADKHKK	TTAQVLIRFP TTAQVLIRFP	MQ	RNLVVIPK <b>S</b> V RNLVVIPK <b>S</b> V	TPERIAENFK TPARIAENFQ	VFD.FELSSQ VFD.FELSSE	DMTTLLSYNR DMTTLLSYNR	NWRVCALL NWRVCALV	SCT	SHKDYPFHEE SHKDYPFHAE	F	~~~~~~			
Rat_ADR Boy_ADR	IKEIAAKYNK IKEIAAKYNK	TTAQVLIRFP TTAQVLIRFP	IQ	RNLVVIPKSV RNLVVIPKSV RNLTVIPKSV	TPARIAENEK	VFD.FEVSSE VFD.FELSNE	DMATLLSYNR	NWRVCALM	SCA	KHKDYPFHAE KHKDYPFHAE SURDYPFHEF	V				
Por_ADR Mou_VDP	IKAIAAKYNK IKEIAAKHKK	TTAQVLIRFP	MQ	RNLIVIPKSV RNVVVIPKSV	TPERIAENFQ	VFD.FELSPE VFD.FOLSEE	DMNTLLSYNR	NWRVCALM.	SCA	SHKDYPFHEE TEEDYPFHEE	Y	~~~~~~			
Mou_FR1 CHO_AKR	IKEIAAKHEK IKEIAAKHKK	TSAQVLIRFH TSAQVLIRFH	IQ IQ	RNVVVIPK <b>S</b> V RNVVVIPK <b>S</b> V	TPSRIQENIQ TPARIHENFQ	VFD.FQLSDE VFD.FQLSDQ	EMATILSFNR	NWRACLLP	ETV	NMEEYPYDAE NMEEYPYDAE	Ŷ~~~~~~ Y~~~~~~~	~~~~~~			
Hum_20aHSD Hum_BABP	LCALAKKHKR LCALAKKHKR	TPALIALRYQ TPALIALRYQ	LQ	RGVVVLAK <mark>S</mark> Y RGVVVLAK <mark>S</mark> Y	NEQRIRQNVQ NEQRIRQNVQ	VFE.FQLTSE VFE.FQLTSE	EMKAIDGLNR EMKAIDGLNR	NVRYLTLD NVRYLTLD	IFA IFA	GPPNYPFSDE GPPNYPFSDE	Ү~~~~ Ү~~~~				
Hum_ChlorRed	LCALAKKHKR LCALAKKHKR	TPALIALRYQ TPALIALRYQ TPALIALRYQ	LQ	RGVVVLAKSY	NEQRIRENIQ	VFE.FQLTAE VFE.FQLTSE	DMKAIDGLDR DMKVLDGLNR	NLHYFNSD NYRYVVMD		DHPDYPFSDE	Y	~~~~~~			
Mou_17bHSD Bov_PGS	LGSMAKKYNR LCAIAKKHKO	TPALIALRYQ TPALVALRYO	LQ	RGVVVLAKSF RGVVVLAKSF	SEKRIKENMQ	VFE.FQLTSE VFE.FQLTSE	DMKVLDDLNK	NIRYISGS	SFK	DHPDFPFWDE GHPEYPFSEE	Y~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~				
Bov_DD3 Rat_20aHSD	LSAIAKKHRO LCTMAKKYKR	TPALVALRYQ TPALIALRYQ	IQ LE	RGVVVLAK <mark>S</mark> Y RGIVTLVK <b>S</b> F	NKKRIKENIQ NEERIRENLQ	VFD.FELTPE VFD.FQLASD	DMKAIDGLNS DMEILDNLDR	NMRYNELL . NLRYFPAN	LGV	GHPEYPFVEE AHPNFPFSDE	Ү Ү				
Rat_3aHSD Frog_Rho	LCAIAKKYKQ LNKIAAKYNR	TPALVALRYQ SSAEVAMRFI	LQ LQ	RGVVPLIR <b>S</b> F KGIVVLAKSF	NAKRIKELTQ TPARIKQNLG	VFE.FQLASE VFE.FELKPE	DMKALDGLNR DMKTLESLDR	NFRYNNAK NLHYGPFR	YFD EVK	DHPNHPFTDE QHPEYPFHDE	¥~~~~~	~~~~~~~~~			
Rat_305bRed Bar ALR	LTSLGKKYNK VEKVANKLNV	TQAQIVLERD TPGOVITENA	IQ	RGLVVIPKST RGLVVIPKST RGTSVIPKS	TPERIKENFQ	IFD.FSLTEE	EMKDIEALNK	NVRFVELL		DHPEYPFHDE UNKTHOPYPC	Y				
Brgr_ALR App_S6Pdh	VEKVANKLNK	TPGQVLIKWA SVAQICLRWG	LQ	RGTIVIPKSS	KDERIKENIQ KIORLKENIF	VFG.WEIPEE VLE.FOLSDF	DFQVLCSIKD	EKRVLTGE	ELF	VNKTHGPYKS GLDVYA~~~~	ASEVWDNEN~				
Pic_XylRed Klu_XylRed	IKAIAAKHGK IKSIASKHKV	SPAQVLLRWS TPQQVLLRWA	sQ TQ	RGIAIIPK <b>S</b> N NGIAIIPK <b>S</b> S	TVPRLLENKD	VN.SFDLDEQ INDALTLTDD	DFADIAKLDI ELKQISGLNO	NLRFNDP		DKIPIFV~ LDNEFPTFI~					
Muc_TADH Sac_GCY	VKSVADKHNV ILEIAKKNNV	STGQVLLRWA QPGHVVISWH	LD	REFAVIPK <mark>S</mark> V RGYVVLPKSV	NAGRMKANLE NPDRIKTNRK	ILD.IKLDAE IFTLSTEDFE	DNKTLDSLKT AINNISKEKG	NQRFNDP EKRVVHP		GFG.LPLFD~ PFEVFK~~~~	********	~~~~~~~			
Sporo_ALR Alf_ChalRed Sov_ChalRed	IKKIAEKNGC LKEIADAHGK	TPAQVLIAWA SVAQISLRWL STAOVGI PM	1V YE	GGHSVIPKSV QGVTFVPKSY	TPSRIGENFK DKERMNQNLR	UVSLSQEDVD IFD.WSLTKE	AVSKLGEGSG DHEKIAQIKQ	RRRYNIPC NRLIPGPT		PKWDINVFGE LNDLYDD	EDEKSCKNAV	KIK			
Gly_PKR Lei_putRed	LKGIAEAHGK	SIAQVSLRWL TAAQVILRWN	YE	QGVTFVAKSY KNLITIPKSV	DKERMNONLO	IFD.WELTTE	DHQKIDQIKQ DVMSIDALNT	NRLIPGPT		LNDLWDDEL~ DEAOF~~~~~	~~~~~~~~~~				
Ps_Mordh Cb_25dkg	ITTIAEAHGR VTAAAAAHGK	SAAQVILRWH TPAQAVLRWH	FQ LQ	NDVVAIPKSV KGFVVFPKSV	NPERIAKNID	VFD.FALSDA VFD.FDLTDT	EMAQLDELDT	GVRIG GDGSGRVS	PDP	RDVDTSSFAE DEVD~~~~~	FV				
Cb_25dkg2 Rat_AFBRed	LQELAVVYGV .KTTYGPTAP	TPTQVVLRWH SMISAAVRWM	VQ YHHSQLKGTQG	LGSTPIPK <mark>S</mark> A DAVILGMS <b>S</b> L	DPDRQRENAD EQLEQNLALV	VFG.FALTAD EEG.PLEPA	QVDAISGLER VVDAFDQA	GRLWDGDP	DTH CPN	EEM~~~~~~ YFR~~~~~~	~~~~~~~~~	~~~~~~~~~~~			



#### Figure 4 AKR active site

The three-dimensional structures of all AKRs retain the same spatial relationship between the active-site residues (Asp-50, Tyr-55, Lys-84 and His-117), the nicotinamide cofactor (shown in red) and a water molecule, as shown in this stereo-view of the rat liver  $3\alpha$ -HSD·NADP<sup>+</sup> binary-complex active site. Reprinted, with permission, from Bennett et al. [86]; copyright 1996 American Chemical Society.

[77–80]. In these enzymes, Asp-50 is not crucial for catalysis, although mutations of this residue interfere with substrate turnover and cofactor binding. Likewise, His-117 mutants of each enzyme retain some degree of catalytic activity, suggesting that this residue is also not the proton donor. Mutagenesis and molecular-modelling studies support a role for this histidine residue in orienting the substrate carbonyl in the active site [78,81]. Mutations of both Tyr-55 and Lys-84 severely impair or completely inactivate catalytic activity. However, based on the crystal structures, Lys-84 is not in the correct position to act as the general acid/base. The results to date thus favour Tyr-55 as the catalytic acid in the reaction mechanism [13].

Conservation of the active-site residues argues for a similar mechanism of oxidoreduction in other superfamily members. In all of the AKRs, both Asp-50 and Lys-84 are invariant. Tyr-55 is replaced only in  $\rho$ -crystallin, but this protein has extremely low enzymic activity, suggesting the importance of this residue in catalysis [82]. Interestingly, all the AKRs that catalyse carbonyl oxidoreduction reactions retain His-117, but the 3-oxo-5 $\beta$ -steroid 4-dehydrogenases do not.

The 3-oxo-5 $\beta$ -steroid 4-dehydrogenases catalyse the transfer of a hydride ion from the 4-*pro*-R position of NADPH to the  $5\beta$ position on the steroid substrate, with the addition of an antifacial proton at the  $4\alpha$  position [83]. This change in reaction mechanism is accompanied by a change of His-117 to a glutamic acid residue. As noted by Akhtar et al. [84], carbon-carbon double-bond reduction in steroids can occur by mechanisms that involve either protonation or hydride transfer occurring first. Protonation would give rise to a carbonium-ion intermediate, so that hydride addition could occur by a Markovnikov addition. The presence of a glutamate in the 3-oxo-5 $\beta$ -steroid 4-dehydrogenase active site could provide a mechanism for stabilizing an intermediate carbonium ion. If hydride transfer occurs first, then a negatively charged enolate would form at C-3 of the 3-oxo-5 $\beta$ steroid substrate. Although this enolate could be stabilized by the oxyanion hole at the AKR active site, this hole may be compromised by the lack of His-117. Whether or not the equivalent of Tyr-55 would function as the general acid in double-bond reduction is unknown. Interestingly, an enolate is formed in 3-oxosteroid  $\Delta^5 - \Delta^4$ -isomerase ( $\Delta^5$ -3-ketosteroid isomerase) reaction, where a tyrosine does function as a general acid [85].

## NAD(P)(H) BINDING

The multiple sequence alignment, combined with the threedimensional structures, suggests a similar mode of cofactor binding in all superfamily members. The enzyme 3NADP(H)



Figure 5 Schematic of NADP<sup>+</sup> binding in the AKRs

The atomic interactions between NADP<sup>+</sup> (depicted in red) and an AKR are indicated based on the structure of the  $3\alpha$ -HSD·NADP<sup>+</sup> binary complex [86]. Reprinted by permission of the publisher from Penning, Pawlowski, Schlegel, Jez, Lin, Smith-Hoog, Bennett and Lewis (1996) Steroids **61**, 508–523; copyright 1996 by Elsevier Science Inc.

binary-complex structures of ADR, ALR, FR-1 and  $3\alpha$ -HSD show the cofactor bound in an extended conformation on the C-terminal side of the barrel, with the nicotinamide ring located at the core of the barrel and the pyrophosphate bridge straddling the lip [7,8,10,69,86]. A short portion of the B-loop undergoes a

#### Table 2 Residues of the AKR cofactor-binding pocket

A multiple sequence alignment of the amino acids that comprise the cofactor binding pocket based on the known three-dimensional structures of AKRs is shown. Designations for the AKRs are as in Table 1. Strictly conserved residues are in bold face; positions with conservative changes are in italics.

 Protein	Residue	24	50	166	167	190	216	219	221	270	271	272	276	279	280
Hum-ALR		W	D	S	N	Q	Ŷ	L	S	K	S	I	R	Q	N
Por-ALR		W	D	S	Ν	Q	Υ	L	S	Κ	S	V	R	Q	Ν
Rat-ALR		W	D	S	Ν	Q	γ	L	S	Κ	S	I	R	Q	N
Mou-LKR		W	D	S	Ν	Q	γ	L	G	Κ	S	V	R	Ε	N
Hum-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Rab-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Mou-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	Ν
Rat-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Bov-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Por-ADR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Mou-VDP		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	Ν
Mou-FR-1		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
CHO-AKR		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Ε	N
Hum-20aHSD		Y	D	S	Ν	Q	γ	L	S	Κ	S	Y	R	Q	N
Hum-BABP		Y	D	S	Ν	Q	Υ	L	S	Κ	S	Y	R	Q	Ν
Hum-3aHSDII		Y	D	S	Ν	Q	γ	L	S	Κ	S	Y	R	Q	N
Hum-ChlorRed		Y	D	S	Ν	Q	Н	L	Т	Κ	S	Y	R	Ε	N
Rab-20aHSD		Y	D	S	Ν	Q	γ	L	S	Κ	S	F	R	Ε	N
Mou-17bHSD		Y	D	S	Ν	Q	γ	L	S	Κ	S	F	R	Ε	N
Bov-Pgs		Y	D	S	Ν	Q	γ	L	А	Κ	S	F	R	Ε	N
Bov-DD3		F	D	S	Ν	Q	Υ	L	S	Κ	S	Y	R	Ε	Ν
Rat-20aHSD		Y	D	S	Ν	Q	γ	L	Т	Κ	S	F	R	Ε	N
Rat-3aHSD		Т	D	S	Ν	Q	γ	L	S	R	S	F	R	Ε	L
Frog-Rho		Y	D	S	Ν	Q	γ	L	S	Κ	S	F	R	Q	N
Hum-3o5bRed		Y	D	S	Ν	Q	γ	L	Т	Κ	S	F	R	Ε	N
Rat-3o5bRed		Y	D	S	Ν	Q	γ	L	Т	Κ	S	Т	R	Ε	N
Bar-ALR		W	D	С	Ν	Q	γ	L	_	Κ	S	S	R	Ε	N
Brgr-ALR		W	D	С	Ν	Q	γ	L	_	Κ	S	S	R	Ε	N
App-S6Pdh		W	D	S	Ν	Q	Н	L	G	Κ	S	S	R	Ε	N
Pic-XylRed		W	D	S	Ν	Q	γ	F	Р	Κ	S	Ν	R	Ε	N
Klu-XylRed		W	D	S	Ν	Q	γ	F	Р	Κ	S	S	R	D	N
Muc-TADH		W	D	Α	Ν	Q	γ	F	Р	Κ	S	V	R	А	N
Sac-GCY		W	D	S	Ν	Q	γ	L	S	Κ	S	V	R	Т	N
Sporo-ALR		W	D	S	Ν	Q	γ	L	Ν	Κ	S	V	R	Ε	N
Alf-ChalRed		_	D	S	Ν	Q	F	L	_	Κ	S	Y	R	Q	N
Soy-ChalRed		-	D	S	Ν	Q	F	L	_	Κ	S	Y	R	Q	N
Gly-PKR		-	D	S	Ν	Q	F	L	_	Κ	S	Y	R	Q	Ν
Lei-putRed		W	D	S	Ν	Q	W	L	Q	Κ	S	V	R	Ε	Ν
Ps-Morph		F	D	С	Ν	Q	W	1	С	Κ	S	V	R	Κ	Ν
Cb-25dkg		F	D	S	Ν	Q	W	L	Q	Κ	S	V	R	Ε	Ν
Cb2-25dkg		Y	D	S	Ν	Q	W	L	R	Κ	S	А	R	Ε	Ν
Rat-AFBRed		R	D	S	Ν	Q	F	L	G	S	S	L	Ε	L	Α

conformational change upon cofactor binding and locks the pyrophosphate portion of NADP(H) in place; this represents the rate-limiting step in the reaction pathway [74,87,88].

The contacts between protein and cofactor are nearly identical in all known AKR structures (Figure 5). The nicotinamide ring and Tyr-216 stack so that the cofactor faces the active site to maintain 4-*pro-R* hydrogen transfer. In this orientation, the nicotinamide ring is in the *anti* conformation with respect to the ribose. This conformation is also dependent upon the hydrogen bonds made by Ser-166, Asn-167 and Gln-190 with the carboxamide moiety of the cofactor. Asp-50 and Thr-24 interact with the nicotinamide ribose, while the pyrophosphate backbone hydrogen-bonds with Leu-219 and Ser-221 of the B-loop. Finally, the adenosine 2'-monophosphate has extensive hydrogen-bond and charge–charge interactions with the protein through Arg-270, Ser-271, Phe-272, Arg-276, Glu-279 and (generally) Asn-280.

Mutagenesis studies of ADR and ALR demonstrate the importance of Arg-270 and Arg-276 in discriminating between

the binding of NAD(H) and NADP(H) [89–91]. Additional mutagenesis has probed the contributions of other residues in cofactor binding. In ADR, a Y216F mutant had only slightly decreased affinity for the cofactor, consistent with the conservation of planar residues at this position that can stack with the nicotinamide ring [92]. Mutations of residues in contact with the nicotinamide ribose of the bound cofactor (Asp-50 and Thr-24) alter NADPH binding modestly [77,80,92]. However, Lys-84 mutants also exhibit changes in cofactor affinities, possibly reflecting alterations in the hydrogen-bond network with Asp-50 [77,80].

The identity of the cofactor-binding pocket across all the AKRs is striking, being evident even between proteins of less than 30 % overall amino acid sequence identity (Table 2). One-third of the residues of the cofactor pocket are strictly conserved (Asp-50, Asn-167, Gln-190 and Ser-271), and more are highly conserved (Ser-166, Leu-219, Arg-270, Arg-276, Glu-279 and Asn-280), implying a similar mode of cofactor binding across the superfamily. Also, a planar amino acid is always found at

position 216, and would maintain the stacking interaction with the nicotinamide ring. Since the residues involved in NAD(P)(H) binding encompass regions from the centre to the periphery of the  $(\alpha/\beta)_{s}$ -barrel structure, the conservation of the overall three-dimensional fold and of the cofactor-binding site may be related.

## SUBSTRATE SPECIFICITY

Based on the locations of the active site and the cofactor-binding pocket in known AKR structures, a putative substratebinding site has been proposed [9,86]. The topology of the substrate binding site resembles a cleft, and includes three components (Figure 6): (1) an oxyanion binding site (Tyr-55, His-117 and C-4 of the nicotinamide ring); (2) residues at the edge of the active site (Ala-52, Leu-54, Trp-86 and Phe-118); and (3) amino acids from three loops forming the sides of the cleft. The A-loop (Met-120, Phe-128 and Phe-129) contributes to one side of the apolar cleft, with the opposite side being formed by the B-loop (Trp-227) and the C-terminal loop (Asn-306, Ala-308 and Tyr-310).

Functional studies have confirmed the basic structural components of the substrate binding site in the AKRs. Of the residues near the active site, mutagenesis of Trp-86 in  $3\alpha$ -HSD modestly affects steroid binding; in the B-loop, mutation of Trp-227 severely impairs steroid binding [21]. Likewise, in the C-loop, point and deletion mutants of ADR and ALR hinder substrate binding, indicating the importance of this loop in forming part of the apolar cleft [80,93–95]. However, structure/function studies are only now addressing the following two questions: (1) how do the AKRs distinguish between sugar and steroid substrates? and (2) how do the HSDs attain their exquisite positional and stereospecificity for individual steroids?

A comparison of the residues in the putative substrate-binding pocket shows that some portions are variable, while others are conserved (Table 3). In general, positions near the catalytic residues tend to be conserved, with greater variation occurring in the loops. The active-site tyrosine and histidine are conserved for catalytic purposes, but could also lock the carbonyl group of the substrate into position for hydride transfer. Functional analysis supports the importance of these amino acids in substrate binding, since removal of the Tyr-55 hydroxy group (a Y55F mutant) produces a 30-fold increase in the  $K_a$  for testosterone in  $3\alpha$ -HSD [79]. Likewise, His-117 mutants of ADR result in an increase in the  $K_m$  for D-glyceraldehyde of almost 80000-fold [77] and can no longer discriminate between the D- and L-isomers of xylose [78].

Given the structures of aldehyde, monosaccharide and steroid substrates, it becomes apparent that residues at the edge of the active site may discriminate between these molecules. Near the active site, Ala-52 and Trp-86 are conserved. However, amino acids at positions 54 and 118 vary depending upon the substrate, and may have a role in determining sugar versus steroid specificity. Residue 54 is generally leucine or isoleucine in the HSDs, and is valine in the ADRs. Differences in this residue may alter the binding-site topology to accommodate different sized substrates. Also, in the ALRs and ADRs residue 118 is tryptophan, whereas it is nearly always phenylalanine in the HSDs, and is either value or methionine in the 5 $\beta$ -reductases. Molecular dynamics studies of ADR suggest that the indole nitrogen of Trp-118 hydrogen-bonds with the 2'-hydroxy group of substrates containing an R-CHOH-CHO moiety [81]. Also, ADR · NADPH · inhibitor ternary-complex structures exhibit this interaction [96,97]. Replacement of the tryptophan by phenylalanine or valine would remove this hydrogen bond and possibly provide a degree of discrimination between sugar and steroid substrates by controlling accessibility to or substrate orientation within the active site.

Beyond the immediate vicinity of the active site, differences in the composition and size of the three loops on the C-terminal side of the barrel suggest their role as additional determinants of substrate specificity. Since steroids are much larger than monosaccharides, these loops could determine the positional and stereo-specificity observed in the HSDs of the superfamily.

We have carried out homology modelling of the structures of mouse liver  $17\beta$ -HSD and rat ovary  $20\alpha$ -HSD (EC 1.1.1.149) using the three-dimensional co-ordinates of rat liver  $3\alpha$ -HSD as a template. The resulting models demonstrate the conservation of the barrel structure and highlight sequence differences within the three loops on the C-terminal side of the barrel (Figure 7). Docking of the steroid in the apolar cleft of rat liver  $3\alpha$ -HSD, based on mutagenesis results, suggests that the C-3 carbonyl of the steroid substrate is bound at the catalytic tetrad, with the  $\alpha$ face of the steroid oriented towards the A-loop, preserving the stereochemistry of hydride transfer [21]. In the case of  $17\beta$ -HSD, the steroid would have to bind backwards to place the substrate



#### Figure 6 Substrate binding site

This stereo-view displays the residues of the proposed substrate binding site in relation to the active-site residues (Asp-50, Tyr-55, Lys-84 and His-117) and bound NADP<sup>+</sup> (depicted in red) of  $3\alpha$ -HSD [86]. Apolar residues near the active site (Leu-54, Trp-86 and Phe-118) and on the A-loop (including Met-120, Phe-128, Phe-129, Leu-137 and Phe-139) are shown. The location of Trp-227 on the B-loop, and the position of Tyr-310 on the C-terminal loop are also indicated. Figure prepared using SETOR [100].

#### Table 3 Residues of the AKR substrate pocket

A multiple sequence alignment of the amino acids that comprise the putative substrate binding pocket based on the known three-dimensional structures for AKRs is shown. Designations of the AKRs are as in Table 1. Highly conserved residues with a proposed role in substrate binding are indicated in bold face.

Protein	Residue	52	54	55	86	117	118	120	128	129	227	306	308	310
Hum-ALR		A	I	Y	w	н	W	Y	Р	F	W	I	Р	R
Por-ALR		Α	I	Y	W	н	W	Y	Р	F	W	1	Р	R
Rat-ALR		Α	V	Y	W	н	W	Y	Р	F	W	1	Р	R
Mou-LKR		Α	L	Y	W	н	W	1	D	1	_	А	F	Т
Hum-ADR		Α	V	Y	W	н	W	Т	F	F	W	С	L	S
Rab-ADR		Α	V	Y	W	н	W	Т	Υ	F	W	С	L	S
Mou-ADR		S	V	Y	W	н	W	Т	Υ	F	W	С	L	S
Rat-ADR		Α	V	Y	W	н	W	Т	Υ	F	W	С	L	S
Bov-ADR		Α	V	Y	W	н	W	Т	F	F	W	С	L	S
Por-ADR		Α	V	Y	W	н	W	Т	Р	F	W	С	L	S
Mou-VDP		Α	V	Y	W	н	W	Q	L	L	Y	С	L	D
Mou-FR-1		Α	А	Y	W	н	W	Q	L	F	S	С	L	E
CHO-AKR		Α	А	Y	W	н	W	Q	L	F	W	С	L	E
Hum -20aHSD		Α	L	Y	W	н	F	V	V	1	W	L	L	I
Hum-BABP		Α	V	Y	W	н	F	V	V	1	W	L	L	I
Hum-3aHSDII		Α	L	Y	W	н	S	М	L	S	W	F	S	S
Hum-ChlorRed		Α	L	Y	W	н	F	М	Р	L	W	V	М	F
Rab-20aHSD		Α	F	Y	W	Н	F	Т	1	Ι	W	V	А	F
Mou-17bHSD		Α	Μ	Y	W	Н	F	М	Y	L	W	I	G	S
Bov-Pgs		Α	L	Y	W	Н	S	V	F	V	W	Y	F	К
Bov-DD3		Α	L	Y	W	н	F	V	L	F	W	Ν	L	L
Rat-20aHSD		S	L	Y	W	H	F	V	L	L	С	F	А	М
Rat-3aHSD		Α	L	Y	W	н	F	М	F	F	W	Ν	А	Y
Frog-Rho		Α	I	т	W	н	W	V	S	D	W	G	F	E
Hum-3o5bRed		Α	I	Y	W	E	V	М	1	Y	W	V	L	M
Rat-3o5bRed		Α	V	Y	W	Е	М	М	F	Y	W	V	М	M
Bar-ALR		Α	Е	Y	W	Н	W	F	М	-	-	L	G	E
Brgr-ALR		Α	E	Y	W	н	W	F	Κ	-	-	L	G	E
App-S6Pdh		Α	Н	Y	W	н	Y	М	_	-	-	S	Р	К
Pic-XylRed		Α	D	Y	W	н	F	V	Κ	Y	L	Ν	Р	W
Klu-XylRed		Α	D	Y	W	н	F	L	Κ	Y	L	Ν	Р	W
Muc-TADH		Α	D	Y	W	н	F	V	V	Y	L	Ν	Р	M
Sac-GCY		Α	I	Y	W	н	W	А	Κ	Ν	-	V	Р	N
Sporo-ALR		Α	V	Y	W	н	W	V	I	Т	-	Ν	Р	Т
Alf-ChalRed		Α	А	Y	W	н	W	-	S	F	-	Р	Р	К
Soy-ChalRed		Α	А	Y	W	н	W	-	S	F	-	S	Р	К
Gly-PKR		Α	A	Y	W	н	W	-	S	F	-	Р	Р	К
Lei-putRed		Α	Ι	Y	W	Н	W	R	-	-	-	S	Y	Р
Ps-Morph		Α	S	Y	F	Н	W	-	-	-	D	-	1	Р
Cb-25dkg		Α	I	Y	W	н	W	-	-	-	-	G	V	A
Cb2-25dkg		Α	Ν	Y	Р	н	W	-	-	-	-	D	D	D
Rat-AFBRed		Α	V	Y	А	н	F	D	Т	L	Y	V	Н	C

carbonyl at C-17 towards the catalytic residues, and upsidedown with the  $\beta$ -face towards the A-loop to maintain the observed stereochemistry. For 20 $\alpha$ -HSD, the steroid substrate need only bind backwards. These different binding modes may be permitted by alterations in the A-loop. In 3 $\alpha$ -HSD, Phe-128 and Phe-129 are in the A-loop. By contrast, in 17 $\beta$ -HSD, these residues are replaced by tyrosine and leucine respectively, and 20 $\alpha$ -HSD has two leucines in these positions. These differences would retain the apolar nature of the proposed binding site, but may generate steroid-specific van der Waals interactions that contribute to the positional and stereo-specificity of the reaction. In addition, the B- and C-loops on the opposite side of the apolar cleft could provide further specificity in the HSDs of the superfamily.

The atomic interactions that exist between the substrate and the B- and C-loops are difficult to model, because these loops are the least well defined features of the three-dimensional structure of rat liver  $3\alpha$ -HSD [9,86]. Whereas mutation of Trp-227 on the B-loop dramatically decreases the affinity for testosterone binding [21], this residue is highly conserved across the superfamily, implying that it does not play a role in substrate selectivity. It is possible that other amino acids on the B-loop interact with the substrate. Finally, the role of the C-loop in steroid binding remains unclear.

Residues of the C-terminal loop are found near the active site (position 306) and further along the apolar cleft (positions 308 and 310), suggesting a role for this loop in discriminating between sugars and steroids and also in defining steroid specificity. Among the HSDs this region displays a large degree of sequence variation, with no clear pattern relating to substrate specificity. However, in the ADRs residues 306, 308 and 310 are highly conserved, with a Cys-Leu-Ser sequence; in all the mammalian ALRs these residues become Ile-Pro-Leu/Ile, and in the plant ALRs they are Leu-Gly-Glu. Mutations of position 306 in ADR and ALR affect catalytic efficiency with a number of substrates [80,93]. Also, deletion of the C-terminal loop in ADR after position 295 selectively affects catalytic efficiency with uncharged substrates (e.g. D,L-glyceraldehyde and D-glucose),



# Figure 7 Homology modelling of 17 $\beta$ - and 20 $\alpha$ -HSDs using 3 $\alpha$ -HSD as a template: the three C-terminal loops

The crystallographic co-ordinates of the rat liver  $3\alpha$ -HSD apoenzyme structure (PDB entry 1RAL; [9]) were used to perform homology modelling of mouse liver  $17\beta$ -HSD [27] and rat ovary  $20\alpha$ -HSD [28] using QUANTA (Molecular Simulations Inc., Waltham, MA, U.S.A.). Since both  $17\beta$ - and  $20\alpha$ -HSD are of similar size and share high (> 70%) amino acid sequence identity with  $3\alpha$ -HSD, the polypeptide backbone and homologous residues were modelled directly from  $3\alpha$ -HSD, with dissimilar residues built into the model structures using standard Karplus rotomers. The initial models were energy-minimized using a conjugate gradient algorithm for 200 cycles with CHARMm (Molecular Simulations Inc.). The resulting structures were overlaid to determine the spatial relationship between the three proteins. Pink regions of the structure correspond to segments of high sequence and structural identity. And black regions correspond to areas of low sequence and structural identity. This view shows the ( $\alpha/\beta$ )<sub>8</sub>-barrel from the side and is rotated 180 ° about the barrel axis relative to the view shown in Figure 2. Figure prepared with MolScript [101].

but has little effect on the  $k_{eat}/K_m$  value for charged substrates (e.g. D-glucuronate) [94]. Recently, Barski et al. [95] have shown that the arginine corresponding to position 310 in the C-terminal loop of ALR determines the specificity of ALR for substrates with a negatively charged carboxylate group, such as glucuronate, succinic semialdehyde and *p*-carboxybenzaldehyde. Together, these results indicate that the C-terminal loop can play an important role in determining substrate specificity in the proteins of the AKR superfamily.

## EVOLUTIONARY RELATIONSHIPS IN THE AKR SUPERFAMILY

Since members of the superfamily show marked sequence identity and occur in species from protozoa to mammals, the AKR superfamily may represent a case of divergent evolution from a multifunctional ancestor protein. Too little is known about the gene structures of the AKRs to assess the mechanism of divergence, but the common constellation of active-site residues and the highly conserved NAD(P)(H)-binding pocket implies gene duplication and subsequent evolution of substrate specificity.

The substrate specificity of these proteins is also consistent with divergence from a common ancestor. Although common substrates, such as 4-nitrobenzaldehyde, are turned over by the AKRs, each AKR can also display unique substrate specificity. The  $3\alpha$ -HSD/DD isoforms of rat and human liver provide an example of such divergence. As the only form found in rat liver, the rat enzyme acts as a  $3\alpha$ -HSD/DD and a bile acid-binding protein [11]. This enzyme is 65 % identical to the human liver  $3\alpha$ -HSD/DD isoforms (DD2 and DD4), but each human isoform has unique functions. DD2 and DD4 both have  $3\alpha$ -HSD/DD activity and are 86 % identical with each other in amino acid sequence. However, DD4 has a high  $3\alpha$ -HSD activity and does not bind bile acids, while DD2 has low  $3\alpha$ -HSD activity and binds bile acids with high affinity [31]. Similar arguments for gene duplication and divergence of substrate specificity have been proposed for the HSDs of the SDR superfamily [22].

Based on the results of functional studies and our multiple sequence alignment, we suggest that divergence of function in the AKRs occurs through modification of the three large loops on the C-terminal side of the barrel, with conservation of the  $(\alpha/\beta)_{s}$ -barrel architecture to maintain the active site and the NAD(P)(H)- binding pocket. A precedent for this exists in glucose oxidase and cholesterol oxidase, where the evolution of similar enzymes to catalyse the same reaction on either sugars or steroids has been observed [98]. Reminiscent of the AKRs, these proteins have the same three-dimensional fold and similar cofactor-binding domains, as well as differences in the loops that alter substrate accessibility to the active site.

In addition to the divergent evolution of substrate specificity, the AKR superfamily may also be an example of convergent evolution to the same constellation of active-site residues as the SDRs. Although the AKRs and SDRs are not homologous and have completely different three-dimensional folds, the catalytic mechanisms of the two superfamilies are analogous. The AKRs catalyse NAD(P)(H)-dependent oxidoreduction reactions, in which a tyrosine and a lysine are key catalytic components. Likewise, the SDRs are nicotinamide-dependent enzymes defined by a Tyr-Xaa-Xaa-Xaa-Lys motif at the active site [2]. An overlay of the active sites of an SDR  $(3\alpha/20\beta$ -HSD from Streptomyces hydrogenans) and an AKR (rat liver  $3\alpha$ -HSD) showed that the tyrosine hydroxy group of each protein differs in space by less than 0.5 Å, while the lysines maintain approximately the same location relative to the hydroxy group of the tyrosine [86]. As in the classic example of the serine proteases, the evolution of similar active sites may be a consequence of a single energetically accessible pathway for a given chemical mechanism [99]. Although mechanistic studies of the AKRs and the SDRs have not progressed to the same level as studies of the serine proteases, these initial results are intriguing.

#### CONCLUSION

The AKR superfamily represents an interesting challenge to understanding structure/function relationships, for two main reasons. First, the mammalian AKRs represent potential drug targets, but the similarity of these proteins has made the design of specific inhibitors difficult. Only by understanding the structure/function relationship of the superfamily will rational drug design against individual AKRs be possible. Although the catalytic mechanism and mode of NAD(P)(H) binding in the AKRs is understood, how these proteins discriminate between sugars and steroids, and how positional and stereo-selectivity between different steroids is achieved, remain to be elucidated. Secondly, the architecture of these proteins may prove useful for studying the evolution of substrate specificity and of how a common scaffold can catalyse a similar reaction with a broad range of substrates. In the future, it may be possible to use the AKR scaffold as a basis for engineering new substrate specificity by altering the C-terminal loops of the barrel.

This work was supported by NIH grants DK47015 and CA39504 (to T. M. P.), and by a fellowship from the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation, DRG-1298 (to M. J. B. and M. L.). A preliminary account of this work was presented at the 2nd International Symposium on Molecular Steroidogenesis in Monterey, CA, U.S.A., between 8 and 11 June, 1996.

## REFERENCES

- 1 Jornvall, H., Persson, M. and Jeffrey, J. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4226-4230
- 2 Jornvall, H., Persson, M., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffrey, J. and Ghosh, D. (1995) Biochemistry 34, 6003–6013
- 3 Eklund, H., Horjales, E., Jornvall, H., Branden, C. I. and Jeffrey, J. (1985) Biochemistry 24, 8005–8012
- 4 Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L. and Orr, J. C. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 10064–10068
- 5 Bohren, K. M., Bullock, B., Wermuth, B. and Gabbay, K. H. (1989) J. Biol. Chem. **264**, 9547–9551
- 6 Bruce, N. C., Willey, D. L., Coulson, A. F. W. and Jeffrey, J. (1994) Biochem. J. 299, 805–811
- 7 Wilson, D. K., Bohren, K. M., Gabbay, K. H. and Quiocho, F. A. (1992) Science 257, 458–460
- 8 Wilson, D. K., Nakano, T., Petrash, J. M. and Quiocho, F. A. (1995) Biochemistry 34, 14323–14330
- 9 Hoog, S. S., Pawlowski, J. E., Alzari, P. M., Penning, T. M. and Lewis, M. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 2517–2521
- El-Kabbani, O., Judge, K., Ginell, S. L., Myles, D. A. A., DeLucas, L. J. and Flynn, T. G. (1995) Nature Struct. Biol. 2, 687–692
- Penning, T. M., Pawlowski, J. E., Schlegel, B. P., Jez, J. M., Lin, H.-K., Smith-Hoog, S., Bennett, M. J. and Lewis, M. (1996) Steroids 61, 508–523
- Welle, R., Schroder, G., Schiltz, E., Grisebach, H. and Schroder, J. (1991) Eur. J. Biochem. 196, 423–430
- 13 Petrash, J. M., Tarle, I., Wilson, D. K. and Quiocho, F. A. (1995) Diabetes 43, 955–959
- 14 Tomlinson, D. R., Stevens, E. J. and Diemel, L. T. (1994) Trends Pharmacol. Sci. 15, 293–297
- 15 Kador, P. F. and Sharpless, N. E. (1983) Mol. Pharmacol. 24, 521-531
- 16 Turner, A. J., Illingworth, J. A. and Tipton, K. F. (1974) Biochem. J. 144, 353-360
- 17 Penning, T. M. and Talalay, P. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 4504-4508
- 18 Watanabe, K., Fujii, Y., Nakayama, K., Ohkubo, H., Kuramitsu, S., Kagamiyama, H., Nakanishi, S. and Hayaishi, O. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 11–15
- 19 Dorfman, R. I. and Shipley, R. A. (1956) Androgens, pp. 122–126, John Wiley & Sons, New York
- 20 Lacy, W. R., Washenick, K. J., Cook, R. G. and Dunbar, B. S. (1993) Mol. Endocrinol. 7, 58–66
- 21 Jez, J. M., Schlegel, B. P. and Penning, T. M. (1996) J. Biol. Chem. 271, 30190–30198
- 22 Baker, M. E. (1996) Bioessays 18, 63-70
- 23 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
- 24 Gish, W. and States, D. J. (1993) Nature Genet. 3, 266-272
- 25 Pawlowski, J. E., Huizinga, M. and Penning, T. M. (1991) J. Biol. Chem. 266, 8820–8825
- 26 Jez, J. M., Flynn, T. G. and Penning, T. M. (1996) in Enzymology and Molecular Biology of Carbonyl Metabolism, vol. 6, (Weiner, H., Lindahl, R., Crabb, D. W. and Flynn, T. G., eds.), pp. 579–589, Plenum Press, New York
- 27 Deyashiki, Y., Ohshima, K., Nakanishi, M., Sato, K., Matsuura, K. and Hara, A. (1995) J. Biol. Chem. **270**, 10461–10467
- 28 Miura, R., Shiota, K., Noda, K., Yagi, S., Ogawa, T. and Takahashi, M. (1994) Biochem. J. **299**, 561–567
- 29 Hara, A., Matsuura, K., Tamada, Y., Sato, K., Miyabe, Y., Deyashiki, Y. and Ishida, N. (1996) Biochem. J. **313**, 373–376
- Stolz, A., Hammond, L., Lou, H., Takikawa, H., Ronk, M. and Shively, J. E. (1993) J. Biol. Chem. 268, 10448–10457
- 31 Deyashiki, Y., Ogasawara, A., Nakayama, T., Nakanishi, M., Miyabe, Y., Sato, K. and Hara, A. (1994) Biochem. J. 299, 545–552
- 32 Khanna, M., Qin, K. N., Wang, R. W. and Cheng, K. C. (1995) J. Biol. Chem. 270, 20162–20168
- 33 Winters, C. J., Molowa, D. T. and Guzelian, P. S. (1990) Biochemistry 29, 1080–1087
- 34 Chung, S. and LaMedola, J. (1989) J. Biol. Chem. 264, 14775-14777
- 35 Garcia-Perez, A., Martin, B., Murphy, H. R., Uchida, S., Murer, H., Cowley, B. D., Handler, J. S. and Burg, M. B. (1989) J. Biol. Chem. **264**, 16815–16821
- 36 Gui, T., Tanimoto, T., Kokai, Y. and Nishimura, C. (1995) Eur. J. Biochem. 227, 448–453

- 37 Carper, D., Nishimura, C., Shinohara, T., Dietzchold, B., Wistow, G., Craft, C., Kador, P. and Kinoshita, J. H. (1987) FEBS Lett. **220**, 209–213
- 38 Petrash, J. M. and Favello, A. D. (1989) Exp. Eye Res. 8, 1021-1027
- 39 Warren, J. C., Murdock, G. L., Ma, Y., Goodman, S. R. and Zimmer, W. E. (1993) Biochemistry **32**, 1401–1406
- 40 Jaquinod, M., Potier, R., Klarskov, K., Reymann, J. M., Sorokine, O., Kieffer, S., Barth, P., Andriantomanga, V., Biellman, J. F. and Van Dorsselaes, A. (1993) Eur. J. Biochem. **218**, 893–903
- 41 Bohren, K. M., Barski, O. A. and Gabbay, K. H. (1996) in Enzymology and Molecular Biology of Carbonyl Metabolism, vol. 6, (Weiner, H., Lindahl, R., Crabb, D. W. and Flynn, T. G., eds.), (Adv. Exp. Med. Biol. **414**), pp. 455–464, Plenum Press, New York
- 42 Wermuth, B., Omar, A., Forster, A., di Francesco, G., Wolf, M., von Wartbuth, J. P., Bullock, B. and Gabbay, K. H. (1987) Prog. Clin. Biol. Res. 232, 297–307
- 43 Flynn, T. G., Green, N. C., Bhatia, M. B. and El-Kabbani, O. (1995) Adv. Exp. Med. Biol. **372**, 193–201
- 44 Takahasi, M., Fujii, J., Teshima, T., Suzuki, K., Shiba, T. and Taniguchi, N. (1993) Gene 127, 249–253
- 45 Bartels, D., Engelhardt, K., Roncarati, R., Schneider, K., Rotter, M. and Salamini, F. (1991) EMBO J. 10, 1037–1043
- 46 Lee, S. P. and Chen, T. H. H. (1993) Plant Physiol. 101, 1089–1092
- 47 Kita, K., Matsuzaki, K., Hashimoto, T., Yanase, H., Kato, N., Chung, M. C. M., Kataoka, M. and Shimizu, S. (1996) Appl. Environ. Microbiol. **62**, 2303–2310
- 48 Kondo, K., Kai, M., Setoguchi, Y., Eggertsen, G., Sjoblom, P., Setoguchi, T., Okuda, K. and Bjorkhem, I. (1994) Eur. J. Biochem. 219, 357–363
- 49 Onishi, Y., Noshiro, M., Shimosato, T. and Okuda, K. (1994) FEBS Lett. 283, 215–218
- 50 Sallaud, C., El Turk, C., Bigerre, L., Sevin, H., Welle, R. and Esnault, R. (1995) Plant Physiol. **108**, 869–870
- 51 Akashi, T., Furuno, T., Futami, K., Honda, M., Takahashi, T., Welle, R. and Ayabe, S. (1996) Plant Physiol. **111**, 347
- 52 Anderson, S., Berman-Marks, C., Lazarus, R., Miller, J., Stafford, K., Seymour, J., Light, D., Rastetter, W. and Estell, D. (1985) Science 230, 144–149
- 53 Grindley, J. G., Payton, M. A., De Pol, H. and Hardy, K. G. (1988) Appl. Environ. Microbiol. 54, 1770–1775
- 54 Willey, D. L., Caswell, D. A., Lowe, C. R. and Bruce, N. C. (1993) Biochem. J. 290, 539–544
- 55 Samaras, N. and Spithill, T. W. (1989) J. Biol. Chem. 264, 4251-4254
- 56 Terada, T., Adachi, H., Nanjo, H., Fujita, N., Takagi, T., Nishikawa, J., Imagawa, M., Nishihara, T. and Maeda, M. (1996) in Enzymology and Molecular Biology of Carbonyl Metabolism, vol. 6, (Weiner, H., Lindahl, R., Crabb, D. W. and Flynn, T. G., eds.), (Adv. Exp. Med. Biol. **414**), pp. 545–553, Plenum Press, New York
- 57 Tomarev, S. I., Zinovieva, R. D., Dolgilevich, S. M., Luchin, S. V., Krayev, A. S., Skyryabin, K. G. and Gause, G. G. (1984) FEBS Lett. **171**, 297–301
- 58 Pailhoux, E. A., Martinez, A., Veyssiere, G. M. and Jean, C. G. (1990) J. Biol. Chem. 265, 19932–19936
- 59 Kanayama, Y., Moni, H., Imaseki, H. and Yamaki, S. (1992) Plant Physiol. 100, 1607–1608
- 60 Amore, R., Kotter, P., Kuster, C., Ciriacy, M. and Hollenberg, C. P. (1991) Gene 109, 89–97
- 61 Billard, P., Menart, S., Fleer, R. and Bolotin-Fukuhara, M. (1995) Gene 162, 93-97
- 62 Czempinski, K., Kruft, V., Wostermeyer, J. and Burmester, A. (1996) Microbiology 142, 2647–2654
- 63 Oechsner, U., Magodolen, V. and Bandlow, W. (1988) FEBS Lett. 238, 123-128
- 64 Hyndman, D. J., Takenoshita, R., Vera, N. L., Pang, S. G. and Flynn, T. G. (1997) J. Biol. Chem. 272, 13286–13291
- 65 Donohue, P. J., Alberts, G. F., Hampton, B. S. and Winkles, J. A. (1994) J. Biol. Chem. 269, 8604–8609
- 66 Ellis, E. M., Judah, D. J., Neal, G. E. and Hayes, J. D. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 10350–10354
- 67 McCormack, T. and McCormack, K. (1994) Cell 79, 1133–1135
- Devereux, J., Haeberil, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395
  Rondeau, J. M., Tete-Favier, F., Podjarny, A., Reymann, J. M., Barth, P., Biellman, J. F. and Moras, D. (1992) Nature (London) 355, 469–472
- 70 Scheerlink, J. P., Lasters, I., Claessens, M., De Maeyer, M., Pio, F., Delhaise, P. and Wodak, S. J. (1992) Proteins Struct. Funct. Genet. 12, 299–313
- 71 Janecek, S. (1996) Protein Sci. 5, 1136–1142
- 72 Farber, G. K. and Petsko, G. A. (1990) Trends Biochem. Sci. 15, 228-234
- 73 Davidson, W. S. and Flynn, T. G. (1979) Biochem. J. 177, 595-601
- 74 Grimshaw, C. E., Shahbaz, M. and Putney, C. G. (1990) Biochemistry 29, 9947-9955
- 75 Askonas, L. J., Ricigliano, J. W. and Penning, T. M. (1991) Biochem. J. 278, 835–841
- 76 Yoshida, M., Kuroki, Y., Kobayashi, E. and Tamaoki, B. (1992) J. Steroid Biochem. Mol. Biol. 41, 29–36

- 77 Tarle, I., Borhani, D. W., Wilson, D. K., Quiocho, F. A. and Petrash, J. M. (1993) J. Biol. Chem. **268**, 25687–25693
- 78 Bohren, K. M., Grimshaw, C. E., Lai, C. J., Harrison, D. H., Ringe, D., Petsko, G. A. and Gabbay, K. H. (1994) Biochemistry **33**, 2021–2032
- 79 Pawlowski, J. E. and Penning, T. M. (1994) J. Biol. Chem. 269, 13502-13510
- 80 Barski, O. A., Gabbay, K. H., Grimshaw, C. E. and Bohren, K. M. (1995) Biochemistry 34, 11264–11272
- 81 De Winter, H. L. and von Itzstein, M. (1995) Biochemistry 34, 8299-8308
- 82 Fujii, Y., Watanabe, K., Hayashi, H., Urade, Y., Kuramitsu, S., Kagamiyama, H. and Hayaishi, O. (1990) J. Biol. Chem. **265**, 9914–9923
- 83 Berseus, O. and Bjorkhem, I. (1967) Eur. J. Biochem. 2, 503-507
- 84 Akhtar, M., Wilton, D. C., Watkinson, I. A. and Rahimtula, A. D. (1972) Proc. R. Soc. London B 180, 157–177
- 85 Zhao, Q., Abeygunawardana, C., Talalay, P. and Mildvan, A. S. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 8220–8224
- 86 Bennett, M. J., Schlegel, B. P., Jez, J. M., Penning, T. M. and Lewis, M. (1996) Biochemistry 35, 10702–10711
- 87 Borhani, D. W., Harter, T. M. and Petrash, J. M. (1992) J. Biol. Chem. 267, 24841–24847
- 88 Kubiseski, T. J., Hyndman, D. J., Morjana, N. A. and Flynn, T. G. (1992) J. Biol. Chem. 267, 6510–6517

- 89 Bohren, K. M., Page, J. L., Shankar, R., Henry, S. P. and Gabbay, K. H. (1991) J. Biol. Chem. **266**, 24031–24037
- 90 Yamaoka, T., Matsuura, Y., Yamashita, K., Tanimoto, T. and Nishimura, C. (1992) Biochem. Biophys. Res. Commun. 183, 327–333
- 91 Kubiseski, T. J. and Flynn, T. G. (1995) J. Biol. Chem. 270, 16911–16917
- 92 Ehrig, T., Bohren, K. M., Prendergast, F. G. and Gabbay, K. H. (1994) Biochemistry 33, 7157–7165
- 93 Petrash, J. M., Harter, T. M., Devine, C. S., Olins, P. O., Bhatnagar, A., Liu, S. Q. and Srivastava, S. K. (1992) J. Biol. Chem. 267, 24833–24840
- 94 Bohren, K. M., Grimshaw, C. E. and Gabbay, K. H. (1992) J. Biol. Chem. 267, 20965–20970
- 95 Barski, O. A., Gabbay, K. H. and Bohren, K. M. (1996) Biochemistry 35, 14276–14280
- 96 Wilson, D. K., Tarle, I., Petrash, J. M. and Quiocho, F. A. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 9847–9851
- 97 Harrison, D. H., Bohren, K. M., Ringe, D., Petsko, G. A. and Gabbay, K. H. (1994) Biochemistry **33**, 2011–2020
- 98 Li, J., Vrielink, A., Brick, P. and Blow, D. M. (1993) Biochemistry 32, 11507–11515
- 99 Reardon, D. and Farber, G. K. (1995) FASEB J. 9, 497–503
- 100 Evans, S. V. (1990) J. Mol. Graph. 12, 134-138
- 101 Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950