Cloning and sequencing of the cDNA species for mammalian dimeric dihydrodiol dehydrogenases

Eiji ARIMITSU, Shinya AOKI, Syuhei ISHIKURA, Kumiko NAKANISHI, Kazuya MATSUURA and Akira HARA[1]

Laboratory of Biochemistry, Gifu Pharmaceutical University, Mitahora-higashi, Gifu 502-8585, Japan

Cynomolgus and Japanese monkey kidneys, dog and pig livers and rabbit lens contain dimeric dihydrodiol dehydrogenase (EC 1.3.1.20) associated with high carbonyl reductase activity. Here we have isolated cDNA species for the dimeric enzymes by reverse transcriptase-PCR from human intestine in addition to the above five animal tissues. The amino acid sequences deduced from the monkey, pig and dog cDNA species perfectly matched the partial sequences of peptides digested from the respective enzymes of these animal tissues, and active recombinant proteins were expressed in a bacterial system from the monkey and human cDNA species. Northern blot analysis revealed the existence of a single 1.3 kb mRNA species for the enzyme in these animal tissues. The human enzyme shared 94%, 85%, 84% and 82% amino acid identity with the enzymes of the two monkey strains (their sequences were identical), the dog, the pig and the rabbit respectively. The sequences of the primate enzymes consisted of 335 amino acid residues and lacked one amino acid compared with the other animal enzymes. In contrast with previous reports that other types of dihydrodiol dehydrogenase, carbonyl reductases and enzymes with either activity belong to the aldo–keto reductase family or the short-chain dehydrogenase/reductase family, dimeric dihydrodiol dehydrogenase showed no sequence similarity with the members of the two protein families. The dimeric enzyme aligned with low degrees of identity (14–25%) with several prokaryotic proteins, in which 47 residues are strictly or highly conserved. Thus dimeric dihydrodiol dehydrogenase has a primary structure distinct from the previously known mammalian enzymes and is suggested to constitute a novel protein family with the prokaryotic proteins.

Key words: cDNA cloning, 3-deoxyglucosone, medium-chain dehydrogenase/reductase family.

INTRODUCTION

trans-Dihydrodiol dehydrogenase (DD, EC 1.3.1.20) catalyses the $NADP^+$-linked oxidation of *trans*-dihydrodiols of aromatic hydrocarbons to the corresponding catechols. The enzyme has been shown to suppress the formation of carcinogenic dihydrodiol epoxides from polycyclic aromatic hydrocarbons [1,2] and conversely to be involved in the metabolic activation of the hydrocarbons through auto-oxidation of the catechol products [3,4]. The enzyme exists in multiple forms in mammalian tissues. Most of the enzymes purified from various mammalian tissues are monomeric and have been shown to be identical with 3α-, 17β- and 20α-hydroxysteroid dehydrogenases, aldehyde reductase and/or aldose reductase [5–10], which are members of the aldo–keto reductase superfamily [11]. In addition to the monomeric enzymes, dimeric DDs composed of 39 kDa subunits have been isolated from rabbit lens [10], monkey kidney [12,13], pig tissues [14] and dog liver [15]. The properties of dimeric DDs in the tissues are similar, but are different from those of the monomeric enzymes. The dimeric DD does not accept hydroxysteroids as substrates, and oxidizes (−)-(1*R*,2*R*)-*trans*-dihydrodiols of benzene and naphthalene, in contrast with the preference of the monomeric enzymes for the (+)-(1*S*,2*S*) isomers [16]. It exhibits high reductase activity for dicarbonyl compounds, aldehydes and ketones [10–15]. Thus dimeric DD is also functionally related to carbonyl reductase, which is a member of the short-chain dehydrogenase/reductase family [17], but the primary structure of the dimeric enzyme has not been determined.

The biotransformation of *trans*-dihydrodiols of aromatic hydrocarbons and carbonyl compounds depends on the level of expression of the multiple forms of DD, carbonyl reductases and other enzymes with the two enzyme activities. Dimeric DD has been reported to be expressed specifically in rabbit lens [10], monkey kidney [18], and dog liver and kidney [15], although the enzyme is present in various tissues in the pig [14]. The dimeric enzyme has been suggested to have a role in the detoxification of cytotoxic dicarbonyl compounds such as 3-deoxyglucosone and methylglyoxal in the tissues [13,14], whereas it is thought to be responsible for the pathogenesis of naphthalene cataract in the rabbit [3,10]. The marked species difference in the tissue distributions of dimeric DD provides a possibility that in other mammalian species including humans the dimeric enzyme is expressed in a tissue that has not been examined.

To compare mammalian dimeric DD with monomeric DDs and functionally related oxidoreductases more precisely, and to facilitate future work on its structure–function relationship, we isolated the cDNA species for dimeric DDs of *Cynomolgus* and Japanese monkey kidneys, dog liver, pig liver and rabbit lens. In addition, we found the expression of mRNA for dimeric DD in human intestine, and determined its cDNA sequence. The sequences obtained show high similarity to one another and low similarity to several prokaryotic proteins, and are distinct from

Abbreviations used: DD, *trans*-dihydrodiol dehydrogenase; GFO, glucose:fructose oxidoreductase; RACE, rapid amplification of cDNA ends; RT, reverse-transcriptase.

[1] To whom correspondence should be addressed (e-mail hara@gifu-pu.ac.jp).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers AB021928 (rabbit), AB021929 (pig), AB021930 (dog), AB021931 (Japanese monkey), AB021932 (*Cynomolgus* monkey) and AB021933 (human).

those of members of the aldo–keto reductase and short-chain dehydrogenase/reductase families.

EXPERIMENTAL

Materials

Isogen (solution for RNA extraction) and enzymes for the digestion of proteins were obtained from Wako Pure Chemicals (Osaka, Japan); pBluescript II SK(−), pCR2.1 and pKK223-3 plasmids were purchased from Stratagene, Invitrogen and Pharmacia Fine Chemicals respectively. Restriction and DNA-modifying enzymes, *Escherichia coli* cells, Moloney murine leukaemia virus reverse transcriptase, oligo(dT)$_{12–18}$ and ribonuclease inhibitor were purchased from Gibco BRL, Nippon Gene (Tokyo, Japan), Toyobo (Osaka, Japan) and Takara Shuzou (Shiga, Japan). *trans*-1,2-Dihydro-1,2-dihydroxynaphthalene (naphthalene dihydrodiol), *trans*-1,2-dihydroxycyclohexa-3,5-diene (benzene dihydrodiol) and 3-deoxyglucosone were synthesized by the methods of Platt and Oesch [19,20] and Khadem et al. [21] respectively. Total RNA samples of human tissues were purchased from Sawady Technology (Tokyo, Japan). The antibodies against the purified dimeric DD of Japanese monkey kidney were prepared as described previously [17].

Protein sequencing

Homogeneous dimeric DDs were purified from kidneys of *Cynomolgus* and Japanese monkeys [12], livers of adult beagle dogs [15] and pigs [14]. The enzymes were digested with lysyl endopeptidase, tosylphenylalanylchloromethane ('TPCK')-treated trypsin or pepsin after the reductive S-alkylation [22]. The peptide fragments were separated by reverse-phase HPLC and the isolated peptides were sequenced by automated Edman degradation with a 473A protein sequencer (Applied Biosystems), as described previously [13].

Library screening and DNA sequencing

cDNA clones for dimeric DD were screened from the λgt 11 dog liver cDNA library (Clontech) with antibodies against Japanese monkey kidney DD as the probe, because the antibodies cross-reacted with dog liver dimeric DD [15]. Recombinant phages coding for DD antigen were identified by the plaque-screening method [23] with *E. coli* Y1090 as the host; immunopositive pharge plaques were detected as described previously [24]. The cDNA inserts (approx. 750 bp) from the positive clones were subcloned into *Eco*RI site of pBluescript II plasmids; the nucleotide sequences of the cDNA species were determined as described [24].

Isolation of dimeric DD cDNA by reverse transcriptase (RT)-PCR and rapid amplification of cDNA ends (RACE)-PCR

Total RNA was isolated from kidneys of the *Cynomolgus* and Japanese monkeys, livers of the dogs and pigs, and lenses of adult albino rabbits by the use of Isogen. First-strand cDNA synthesis was performed with reverse transcriptase and oligo(dT)$_{12–18}$ primer. The cDNA was subjected to PCR in 50 μl of 10 mM Tris/HCl, pH 8.3, containing 1.5 mM $MgCl_2$, 50 mM KCl, 50 μM deoxyribonucleotide triphosphates, 0.5 unit of *Taq* DNA polymerase and 50 nM forward and reverse primers (Table 1). PCR amplification consisted of an initial denaturation step at 94 °C for 10 min followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at a temperature that depended on the primer pairs for 1 min, and extension at 72 °C for 2 min. A final extension step at 72 °C for 10 min was also performed in a Biometra thermocycler. For isolation of cDNA for *Cynomolgus* monkey kidney DD, a 577 bp fragment was amplified with a pair of primers, d2f and d3r (Figure 1), that were designed on the basis of the sequence of the cDNA clone for dog DD and identical peptide sequences of the dog and monkey enzymes (Table 1). The 3′ and 5′ ends of the cDNA were generated by using 3′- and 5′-RACE kits (Gibco-BRL) and the gene-specific primers. The fragments were subcloned into pCR2.1 plasmids and sequenced as described above. Similarly, cDNA species for DDs of Japanese monkey kidney, dog liver, pig liver, rabbit lens and the human intestine were generated by RT-PCR and RACE-PCR (Figure 1). The sequence of the coding region was re-examined by sequencing at least three cDNA species, which had been amplified from the first-strand cDNA sample of a tissue

Table 1 Nucleotide sequences of the gene-specific primers used in this study

The restriction enzyme cutting site in each sequence is underlined, and the initiation and stop codons are shown in bold. The nucleotide positions correspond to those of the cDNA species shown in Figure 1. rExF contains an additional 5′ sequence corresponding to the region (positions 1–15) of the monkey cDNA.

Primer	Sequence	Position
d2f and d2r*	5′-GCCTACGGGTCCTATGAG-3′	160–177
d3f and d3r*	5′-CCTTGCTGGTGCCCAACGGAGCTGGTG-3′	754–780
p17f and p17r*	5′-CAGCACCCCCAGCACAAG-3′	223–240
p37f and p37r*	5′-GACCCGCTTCTTTCCTGC-3′	375–392
h1f and h1r*	5′-GCAGAAGCCAGAGAAGATTTC-3′	573–593
h2f and h2r*	5′-AGGGGAGAGCATAAGGAGTTC-3′	785–804
mExF	5′-CCGAATTC**ATG**GCGCTTCGCTGGGG-3′	1–16
mExR	5′-CCAAGC**TTA**GTGTTTGTCTTGGGGG-3′	990–1008
dExF	5′-CCGAATTCCCTAGGTTTTTGGCG**ATG**G-3′	−16 to 3
dExR	5′-CCAAGCTTTATTGACTCCGGGCAT-3′	1012–1029
pExF	5′-TTGAATTCTCGGGGTTGGGGCG**AT**-3′	−15 to 2
pExR	5′-TGAAGCTTGCACATCTTTATTTATTCAG-3′	1018–1036
rExF	5′-CCGAATTC**ATG**GCGCTTCGCTGGGGTGTCGTGTCC-3′	1–15 (monkey) and 16–27 (rabbit)
rExR	5′-CCAAGCTTATTTTTTCAGGGCAGGCG**TCA**-3′	1006–1028
hExF	5′-CCGAATTCCCTAGGTTTTTGGCG**ATG**-3′	−15 to 3
hExR	5′-CCAAGCTTATTCGGGGATACA**TTA**GCG-3′	1002–1023

* The forward (f) and reverse (r) primers were prepared but only the sequence of the forward primer is shown.

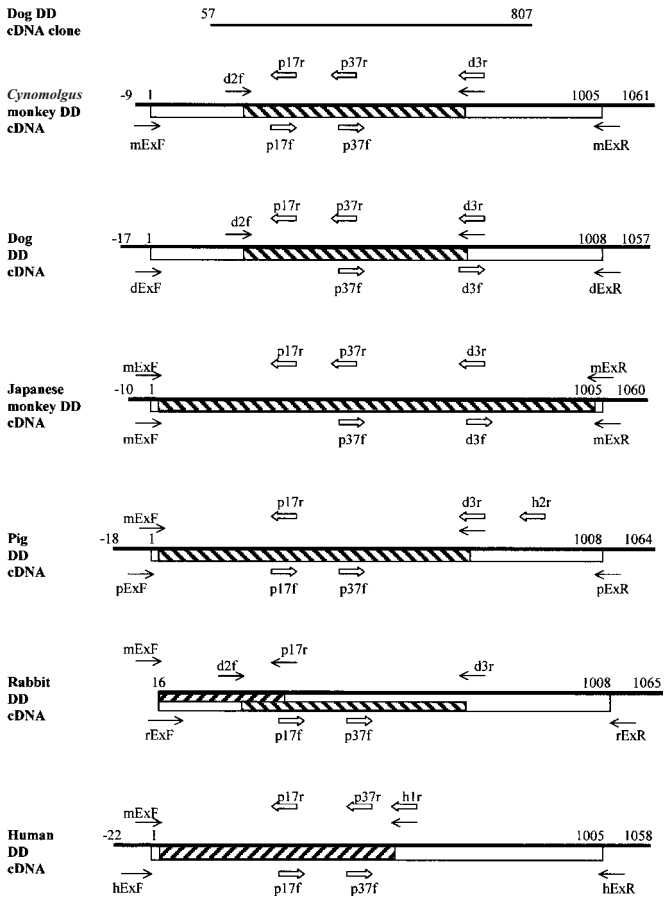

Figure 1 Isolation of the cDNA for six mammalian dimeric DDs by RT-PCR and RACE-PCR

The hatched regions represent the cDNA fragments isolated by the first RT-PCR with the primer pairs (simple arrows above the cDNA species). The dark bars represent the full-length cDNA species determined by 5′-RACE-PCR and 3′-RACE-PCR with the three reverse primers (open arrows above the cDNA species) and two forward primers (open arrows below the cDNA species) respectively. The open boxes represent the coding regions of the cDNA species that were re-amplified with the primer pairs (simple arrows below the cDNA species) to confirm the nucleotide sequences. To determine the nucleotide sequence of the rabbit cDNA, the two fragments were prepared by RT-PCR and only 3′-RACE-PCR was performed. Nucleotide numbers are shown above the cDNA species, in which the 5′ region of the rabbit cDNA is postulated to have the same length as those of the other cDNA species.

with *Pfu* DNA polymerase and the primers corresponding to the regions near the 3′ and 5′ ends of the open reading frame (mExF and mExR, dExF and dExR, rExF and rExR, and hExF and hExR for the monkey, dog, rabbit and human cDNA species respectively).

Northern blot and RT-PCR analyses of mRNA in tissues

The size of mRNA for dimeric DD was analysed by Northern blot analysis. In addition to the animal tissues described above, total RNA was prepared from several tissues of the monkeys, pigs and rabbits. The total RNA samples (5–20 μg) were run on 1 % (w/v) agarose gels, transferred to non-charged nylon membrane (Millipore) and hybridized with digoxigenin-labelled RNA probe as described previously [25]. The probe was prepared from the cDNA for *Cynomolgus* monkey DD by using a digoxigenin luminescent detection kit (Boehringer Mannheim); detection of the probe was performed in accordance with the manufacturer's instructions.

The tissue distribution of mRNA for dimeric DD was examined by RT-PCR with gene-specific primers (Table 1). The total RNA was prepared from fresh tissues of the monkeys, pigs and rabbits, except that human RNA samples obtained from Sawady Technology were used directly. The cDNA synthesis from the RNA (2 μg) was performed in the presence of ribonuclease inhibitor (10 units) as described above. After the remaining RNA in the cDNA solutions (each 25 μl) had been removed by the addition of ribonuclease H (2 units), 1 μl portions of the solutions were subjected to PCR in the reaction mixture described for the isolation of the cDNA species. Identical PCRs were performed for different cycle numbers over the linear range of cycles for each template to examine the tissue-specific expression of the mRNA. The PCR product was amplified from the cDNA samples from many pig tissues by 40 cycles of the reaction, whereas it was detected only for the samples of some tissues of the other mammals. Therefore amplification was performed for 20, 25 and 30 cycles for the cDNA samples from the monkey, rabbit and human tissues respectively, in which the PCR was apparently in the linear range for the respective templates. For the analysis of mRNA in pig tissues, 35 cycles of the reaction were performed to detect the low expression level of the mRNA in some tissues, although the PCR products exceeded the linear range of amplification for the samples containing large amounts of the target cDNA.

Expression of recombinant dimeric DD

The cDNA species for mammalian DDs were digested with *Eco*RI and *Hin*dIII, then ligated into pKK223-3 plasmids. DNA sequencing confirmed that no unintended base substitution had been incorporated in the coding regions of the expression plasmids. Transformation of the plasmid into *E. coli* JM109, expression of the recombinant protein and preparation of *E. coli* cell extract were performed as described [9]. The recombinant enzyme was purified from the cell extract by the method described for the enzyme of Japanese monkey kidney [12].

Enzyme assay

The dehydrogenase activity of dimeric DD was assayed spectrophotometrically with 1.8 mM naphthalene dihydrodiol as substrate at pH 10.0 [10]; its reductase activity was determined at pH 7.5 [12]. One unit of the enzyme activity was defined as the amount catalysing the formation or oxidation of 1 μmol of NADPH in 1 min at 25 °C.

Other methods

Protein concentration was determined by Bradford's method with BSA as the standard [26]. SDS/PAGE was performed on a 12.5 % (w/v) slab gel [27]; protein was stained with a silver-stain kit (Daiichi Kagaku, Tokyo, Japan). The molecular mass of dimeric DD was estimated by gel filtration on a Sephadex G-100 column [10]. Protein for dimeric DD in the *E. coli* extracts was detected by Western blot analysis [28] with antibodies against Japanese monkey DD as the first antibody. Comparison between the nucleotide or deduced amino acid sequences of dimeric DD cDNA and those of cDNA species for other proteins was performed by FASTA and BLAST searches [29,30] of the GenBank, Protein Identification Research, Protein Research Foundation, EMBL and SwissProt databases. Multiple sequence alignment was performed with the program Clustal W [31] and

the phylogenetic tree was drawn with the program TreeView [32]. Predictions of secondary structures of dimeric DDs and related proteins were made with the PHD profile-network method [33].

RESULTS AND DISCUSSION

Peptide sequences of mammalian dimeric DDs and isolation of their cDNA species

The results of sequencing peptides derived from dimeric DDs of *Cynomolgus* and Japanese monkey kidneys, dog liver and pig liver are summarized in Table 2. Because the sequences of five peptides (K8, K16, K5, K14 and K6) obtained from the *Cynomolgus* and Japanese monkey enzymes were the same, the Japanese monkey enzyme was digested further with tosylphenylalanylchloromethane-treated trypsin and pepsin, and the sequences of the peptides isolated were analysed. For the Japanese monkey enzyme, its regions composed of a total of 141 residues were determined. The peptides derived from the dog and pig enzymes also showed highly similar sequences to those of the monkey enzymes, suggesting a high similarity of the entire primary structures of the mammalian dimeric DDs.

Approximately 2×10^5 plaques from the λgt11 dog liver cDNA library were screened with the antibody against monkey kidney DD, which cross-reacted with the dimeric enzyme of dog liver [15]. Three immunopositive clones were isolated; two of them could be subcloned into pBluescript plasmids, and the sequences of the 750 bp cDNA inserts were determined. The two cDNA species had an identical sequence; the deduced amino acid sequence perfectly matched the sequences of the six peptides (K6, K7, K5, K37, K31 and K21) derived from dog liver DD. However, the cDNA did not contain the regions corresponding to the other two peptide sequences, which implies that this cDNA lacks part of the open reading frame.

To establish the full-length cDNA sequence for dimeric DD, we next performed RACE-PCR for the total RNA species obtained from both dog liver and *Cynomolgus* monkey kidney. Based on the sequence of the partial dog DD cDNA clone, the primers were designed to anneal the regions that corresponded to the identical amino acid sequences of the enzymes from the two animal species (Table 2). RT-PCR with primers d2f and d3r amplified the 580 bp cDNA fragments from the total RNA species, and the sequences of full-length cDNA species for DDs of monkey kidney (1070 bp) and dog liver (1074 bp) were determined by 3′- and 5′-RACE-PCR with the respective cDNA fragments (Figure 1). Each of the cDNA species contained a stop codon (TGA or TAA), a polyadenylation signal sequence, the poly(A) sequence, and the sequence around the first ATG codon, which fits well with the rule of Kozak [34]. The cDNA species for the two animal DDs showed a high sequence identity (87%) within their open reading frames, but the cDNA for *Cynomolgus*

Table 2 Sequence data of the peptides derived from the digestion of dimeric DDs with lysylendopeptidase (K), tosylphenylalanylchloromethane-treated trypsin (T) and pepsin (P)

The unidentifiable phenylthiohydantoin amino acid derivative is denoted by X. The yields shown are the average repetitive yields (%); the values in parentheses are pmol of the first and last phenylthiohydantoin amino acid derivatives. The positions correspond to those in Figure 2. The sources of monkey DDs are identified by c (*Cynomolgus* monkey) and j (Japanese monkey).

DD source	Peptide	Sequence	Yield	Position
Monkey	jP21	EFAXKHDIPKA-	85 (51–0.2)	44–53
	cK8	AYGXYEELAK	77 (8.7–0.2)	54–63
	jK8	AYGSYEELAK	90 (172–33)	54–63
	cK16	DPNVEVAYVG-	82 (10–2.0)	64–73
	jK16	DPNVEVAYVGTQHPQHK	86 (101–12)	64–80
	cK5	AVLCEK	72 (7.8–0.2)	92–97
	jK5	AVLCEK	70 (15–14)	92–97
	jP27	MGVNAXE-	81 (5.6–2.4)	99–105
	jT22	SVLAQGTXGDLR	89 (24–11)	137–148
	jP26	LLDLGIY-	88 (539–77)	174–180
	cK14	GMAQLLNPC-	92 (16–2.0)	245–253
	jK14	GMAQLLNPCWCPTELVVK	86 (98–4.8)	245–262
	jK13	EFLLPPVPK	86 (42–10)	267–275
	cK6	NCNFD-	78 (3.2–1.2)	276–280
	jK6	NCNFDNGAGMSYEAK	92 (11–7.5)	276–290
	cK3	HVRECLRK	80 (9.2–0.2)	291–298
	jT28	ESPVIPLVESELLADILEEVRR	93 (19–5.6)	302–323
	jK17	AIGVTFPQDK	68 (101–3.6)	324–333
Dog liver	K6	AYGSYEELAK	91 (177–16)	54–63
	K7	DPNVEVAYIGTQHPQHK	93 (21–5.7)	64–80
	K5	AVLCEK	70 (80–5.3)	92–97
	K37	PMGVNAAEVREMVXEAR-	84 (48–3.4)	98–114
	K31	NFTHIPRSVDWAQAGGGLLDLGIYCIQFISMVFGG-	94 (40–10)	157–191
	K21	GIGQILDPCWCPTELVLK	84 (161–2.1)	245–262
	K14	EFNFTXGAGMAYEAK	90 (54–0.3)	276–290
	K27	ESPVIPLAESELLADILEEIR-	85 (250–1.9)	302–322
	K11	AIGVTFPQD-	69 (201–6.6)	324–332
Pig liver	K12	AYGSYEELAK	72 (57–0.4)	54–63
	K8	DPNVEVAYIGTQHP-	80 (65–5.5)	64–77
	K24	SIANVPRSVDWAQAGGSLLDLGIYXLQF-	87 (112–2.2)	157–184
	K19	GMAQILDPCXCPTEL-	73 (49–0.2)	245–259
	K23	ESPMITLAEXELLADI-	78 (120–1.2)	302–317
	K10	AIGVTFPQDK	68 (53–0.6)	324–333

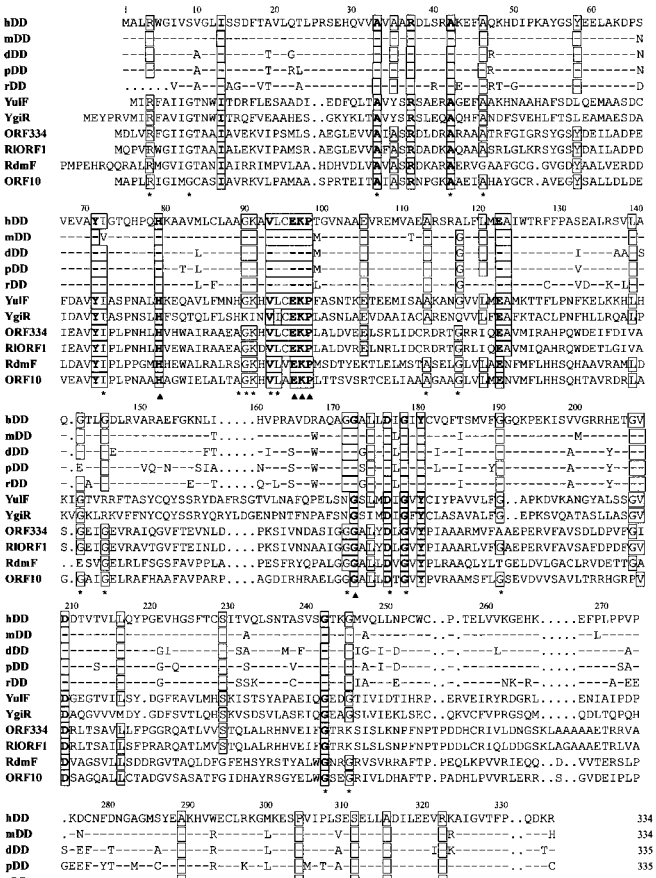

Figure 2 Multiple sequence alignment of mammalian dimeric DDs and hypothetical proteins of prokaryotic genes

Identical residues are indicated by dashes in the alignments of the five dimeric DDs: hDD, human DD; mDD, monkey DD; dDD, dog DD; pDD, pig DD; rDD, rabbit DD. The six prokaryotic proteins, which are defined in Table 3, are also aligned with the dimeric DDs. The strictly conserved residues are highlighted in bold; those conserved in nine or more proteins are boxed. The residues conserved strictly (▲) and highly (more than 75% conservation, shown by asterisks) in the multiple sequence alignment of 20 proteins in Figure 4 are indicated below the ORF10 sequence.

monkey DD lacked a codon corresponding to the AGC codon of the dog DD cDNA at positions 823–825, which corresponds to Ser-275 of the deduced amino acid sequence in Figure 2. Therefore the cDNA species for dimeric DDs of Japanese monkey kidney, pig liver and rabbit lens were isolated by RT-PCR and RACE-PCR to compare the sequences with those of the cDNA species for dog and *Cynomolgus* monkey DDs. Although the 5′-coding region of the cDNA for the rabbit enzyme (corresponding to the N-terminal six residues of the other animal enzymes) could not be determined because of an inability to perform 5′-RACE-PCR, the cDNA species for the enzymes of Japanese monkey kidney and pig liver contained full-length open reading frames (Figure 1). The codon at positions 823–825 was present in the cDNA species for pig and rabbit DDs, unlike that for Japanese monkey DD. The sequences of the peptides derived from DDs of monkey kidneys, pig liver and dog liver matched exactly the sequences deduced from the cDNA species for the respective animal enzymes (Table 2 and Figure 2). In particular, peptide jK13 of Japanese monkey kidney DD verified the deletion of this codon from the cDNA.

When the existence of mRNA for dimeric DD in several human tissues was examined by RT-PCR with primers mExF and h1r, a DNA fragment with a sequence similar to that of the monkey DD cDNA was amplified from the total RNA of human intestine (Figure 1). The full-length cDNA for human DD generated by RACE-PCR had an open reading frame of 1005 bp, which also lacked nucleotides for a codon at the same position as in the monkey DD cDNA species. Therefore proteins predicted from the cDNA species for the primate DDs were shorter by one amino acid than those for the other animal DDs.

Properties of the recombinant protein

The active recombinant enzyme was expressed in *E. coli* cells from the cDNA for Japanese monkey DD and purified to electrophoretic homogeneity; it showed a single and immunoreactive protein band on SDS/PAGE and by Western analysis with antibodies against the monkey kidney DD (results not shown). The specific activity (18 units/mg), subunit molecular mass (39 kDa) and non-denatured molecular mass (76 kDa) of the purified recombinant enzyme were comparable to those of the enzyme purified from Japanese monkey kidney. The recombinant enzyme oxidized naphthalene dihydrodiol (K_m 2.6 mM, V_{max} 36 units/mg) and benzene dihydrodiol (K_m 0.9 mM, V_{max} 16 units/mg) at an optimal pH of 10.0, and reduced 3-deoxyglucosone (K_m 1.2 mM, V_{max} 17 units/mg), camphorquinone (K_m 0.12 mM, V_{max} 34 units/mg) and methylglyoxal (K_m 1.3 mM, V_{max} 10 units/mg) at an optimal pH of 7.5. About half of the enzyme activity was inhibited by the addition of 1 mM isoascorbic acid, 10 μM 4-hydroxyacetophenone or 5 μM 4-chloromercuriphenylsulphonate. These properties of the recombinant enzyme were essentially identical with those of Japanese monkey kidney DD [12,13,35], indicating that this cDNA encodes the subunit of homodimeric monkey DD. The extract of *E. coli* cells transfected with the expression plasmid for human DD cDNA also contained a 39 kDa protein that reacted with the antibody against the monkey dimeric DD and showed a DD activity of 0.2 m-unit/mg. Because of the poor expression of the recombinant human enzyme, it could not be characterized further. Thus the identities of the cDNA species for DDs of at least the monkey, dog, pig and human have been established by expressing the active recombinant enzymes and/or by locating peptides from the respective purified animal enzymes in the deduced amino acid sequences (Table 2).

Sequence comparison between mammalian dimeric DDs

The coding regions of the cDNA species for DDs of *Cynomolgus* and Japanese monkey kidneys showed a difference of only four nucleotides; their predicted amino acid sequences were identical. The sequence identities of the cDNA for *Cynomolgus* monkey DD with those for the human, dog, pig and rabbit enzymes were 94%, 87%, 85% and 83% respectively within their coding regions. Although no direct evidence has been established, the high sequence identity of the cDNA species suggests that the cDNA isolated from rabbit lens encodes dimeric DD. With regard to the deduced amino acid sequences (Figure 2), human DD differs by 19 residues from those of the monkey enzymes (94% identity), and shares 86%, 84% and 82% sequence identity with the dog, pig and rabbit enzymes respectively. There were more than 32 amino acid differences between the enzymes of the primates and the other animals, of which a most striking difference is the deletion of one residue at position 275 observed

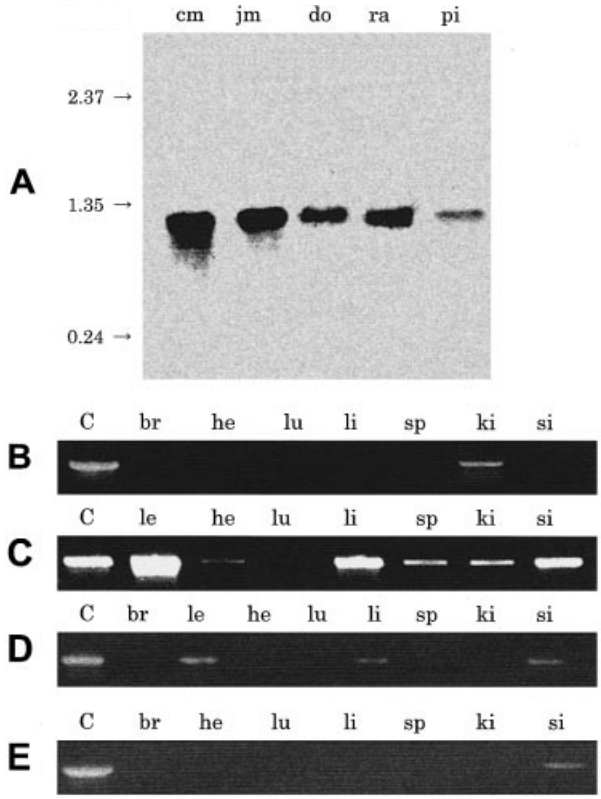

Figure 3 Tissue distribution of mammalian dimeric DDs

(**A**) Northern hybridization analysis of the total RNA species (approx. 20 μg) of *Cynomolgus* monkey kidney (cm), Japanese monkey kidney (jm), dog liver (do), rabbit lens (ra), and pig liver (pi). The mobilities of RNA size markers are indicated (in kb) at the left. (**B**–**D**) RT-PCR analyses of mRNA expression in Japanese monkey (**B**), pig (**C**), rabbit (**D**) and human (**E**) tissues. The primer pairs used were mExF and mExF (**B**), pExF and pExR (**C**), rExF and rExR (**D**) and hExF and hExR (**E**). The PCR products were subjected to electrophoresis on 2% (w/v) agarose gels and revealed by staining with ethidium bromide. As the positive controls (lanes C), the cDNA species for the enzymes of the respective species were included. Tissues: brain (br), lens (le), heart (he), lung (lu), liver (li), spleen (sp), kidney (ki) and small intestine (si).

only for the primate enzymes. The regions (at positions 272–280) around the missing residue are the most diverse between the primate and non-primate enzymes or between the non-primate enzymes (Figure 2). The deletion of the amino acid is probably an event during divergent evolution that might not affect the nature of the enzyme. In addition, the residues different between the animal DDs are not functionally or structurally important, because the properties of the enzymes have been reported to be similar [10,12–15,35].

Tissue distribution

A single mRNA species (approx. 1.3 kb) was detected in *Cynomolgus* and Japanese monkey kidneys, pig liver, dog liver or rabbit lens by Northern blot analysis with the RNA probe corresponding to the cDNA for *Cynomolgus* monkey kidney DD (Figure 3A). No positive signal was detected in liver samples from *Cynomolgus* and Japanese monkeys (results not shown), because the enzyme protein is not detected in tissues other than kidney [18]. The tissue-specific expression of mRNA for the enzyme was examined by RT-PCR with specific primers for the human and animal cDNA species. The expected PCR products were amplified only for the total RNA of kidney in Japanese monkey (Figure 3B), and for those of almost all tissues of the pig (Figure 3C), in conformity with tissue distribution of the enzyme protein [14,18]. An RNA signal with the same size as the PCR product of the cDNA for the rabbit enzyme was detected in liver and small intestine (Figure 3D) as well as in lens. This is inconsistent with the specific distribution of the enzyme in rabbit lens observed by gel-filtration analysis of the enzyme activity [10], but faint immunopositive 39 kDa bands, compared with that of the lens, were observed in the extracts of the liver and intestine by Western analysis with antibodies against the dimeric DD (results not shown). Although the mechanism controlling the different tissue distribution of the enzyme between the four mammals remains to be determined, the expression of the enzyme might be regulated at transcriptional level rather than by tissue-specific effects on mRNA stability.

The present study has for the first time identified a cDNA for human dimeric DD mRNA, which was expressed specifically in small intestine (Figure 3E). Dimeric DD has been thought to be implicated in the pathogenesis of naphthalene-induced cataract in the rabbit. van Heyningen [3] first suggested that DD in the lens converts the *trans*-dihydrodiol of naphthalene to 1,2-dihydroxynaphthalene, which is rapidly auto-oxidized to cytotoxic 1,2-naphthoquinone in addition to H_2O_2. This was supported by the finding of large amounts of dimeric DDs in the rabbit and pig lens [10]. Although it is unknown whether dimeric DD oxidizes *trans*-dihydrodiols of polycyclic aromatic hydrocarbons, in addition to cytochrome P450 and epoxide hydrolase [36–38], the presence of the enzyme in human intestine would suggest that this organ is one of the susceptible human organs to the toxicity of ingested naphthalene and benzene. Dimeric DD also efficiently reduces dicarbonyl compounds such as methylglyoxal and 3-deoxyglucosone, which are formed *in vivo* and are implicated in the formation of advanced glycation end products [39]. In addition, the reactive dicarbonyl compounds have been suggested to be formed *in vivo* from cooked foods containing advanced glycation end products [40]. In this respect, dimeric DD in intestine might act as a detoxification enzyme against food-derived reactive dicarbonyl compounds.

Alignment of dimeric DDs with other proteins

The enzymes associated with DD and/or carbonyl reductase activities have been classified into two protein families, the aldo–keto reductases and short-chain dehydrogenase/reductases [41]; the latter includes prokaryotic *cis*-dihydrodiol dehydrogenases [42]. The sequences of the five dimeric DDs did not show significant similarity to the aldo–keto reductase family proteins including monomeric DDs [11], but Asp-49, Tyr-55 and Lys-91 of the dimeric enzymes (other than rabbit DD) almost correspond to the important residues in the catalysis of the aldo–keto reductase family enzymes [11]. The Tyr residue has been demonstrated by site-directed mutation into Phe to be the catalytic residue of the aldo–keto reductases [43,44]. However, the same mutation with the dimeric DD of Japanese monkey kidney yielded a recombinant enzyme that had a slightly lower activity than that of the wild-type enzyme (S. Aoki and S. Ishikura, unpublished work). Furthermore, the five dimeric DDs bear no sequence similarity with *cis*-dihydrodiol dehydrogenases and other members in the short-chain dehydrogenase/reductase family; no consensus active site sequence, Tyr-Xaa-Xaa-Xaa-Lys, of this family of proteins [18,42] is present in the sequences of dimeric DDs. Therefore our results represent the first complete sequence for a DD that does not belong to the aldo–keto reductase and short-chain dehydrogenase/reductase families. A previous chemical modification study showed the presence of an

Table 3 Proteins aligned with dimeric DDs in Figures 2 and 4

Databases are identified as follows: gp, GenBank; pir, Protein Identification Research; prf, Protein Research Foundation; emb, EMBL.

Abbreviation	Protein	Residue length	Species	Accession no.
YulF	Hypothetical protein yulF	328	*Bacillus subtilis*	pir: H70014
YgjR	ygiR protein	334	*E. coli* K-12	pir: D65097
ORF334	ORF334 gene product	334	*Rhizobium meliloti* L5-30	pir: S51570
RlORF1	ORF1 gene product	334	*Rhizobium leguminosarum*	gb: AF076240
RdmF	*rdmF* gene product	347	*Streptomyces purpurascens*	gb: U10405
ORF10	ORF10 gene product	330	*Amycolatopsis mediterranei*	gb: AF040570
YhhX	Hypothetical 38.8 kDa protein	345	*E. coli* K-12	pir: C65140
Spac	SPAC26H5.09 gene product	369	*Schizosaccharomyces pombe*	gb: Z99126
Hs11	Hypothetical protein s110816	371	*Synechocystis* sp.	pir: S75555
Pht4	PHT4 gene product	410	*Pseudomonas putida*	emb: D13229
CDD	*cis*-Diol dehydrogenase	397	*Alcaligenes* sp.	prf: U18133
YrbE	Opine catabolism homologue yrbE	341	*Bacillus subtilis*	pir: D69972
Hb13	Hypothetical protein b1315	351	*E. coli* K-12	pir: F64880
LapC	NADH-dependent dehydrogenase homologue	358	*Listeria monocytogenes*	gb: AF039207
GFO	Glucose:fructose oxidoreductase precursor	439	*Zymomonas mobilis*	gb: Z80356

essential His residue in the active site of pig liver dimeric DD [45]. In our continuing site-directed mutagenesis studies targeting the five His residues conserved in the mammalian dimeric DDs, the replacement of His-79 of the monkey DD with Gln or Arg almost abolished the enzyme activity (S. Aoki and S. Ishikura, unpublished work). His-79 of dimeric DDs might be a crucial residue for the catalytic process; this is also distinct from the catalytic residues in members of the aldo–keto reductase and short-chain dehydrogenase/reductase families.

A sequence comparison of the five dimeric DDs with proteins in the databases revealed low identities (12–25%) with more than 20 proteins, most of which are hypothetical proteins of prokaryotic, yeast and plant genes. Of the proteins, 15 proteins (Table 3) showed more than 11% residue identities between all pairs of the proteins including dimeric DDs, except that identity between YgjR and LapC was 5%. The multiple sequence alignment of the 15 proteins and the 5 dimeric DDs indicated five invariant residues (H79, E96, K97, P98 and G172) and 21 residues with more than 75% conservation; these are identified by triangles and asterisks respectively in Figure 2. As shown in the phylogenetic analysis (Figure 4) of the result of this multiple sequence alignment, the proteins cluster into three groups. The five dimeric DDs and six prokaryotic gene products (YulF, YgjR, ORF334, RlORF1, RdmF and ORF10) converge on the basis. In these 11 proteins, 17 invariant and 30 highly conserved residues are distributed in their entire sequences (Figure 2); similar secondary structures were predicted for the proteins except for two divergent regions covering residues 146–170 and 250–289 (results not shown). Despite the low similarity between the proteins, the numbers of conserved residues in the proteins are comparable to those in the short-chain dehydrogenase/reductase family. We suggest that at least the dimeric DD and the six hypothetical proteins of the prokaryotic genes constitute a protein family and that the prokaryotic proteins probably function as pyridine nucleotide-dependent dehydrogenases or reductases. Because the sizes (328–410 residues) of the proteins or subunits are longer than those (approx. 250 residues) of the short-chain dehydrogenase/reductase family proteins, the proposed novel protein family is tentatively named the medium-chain dehydrogenase/reductase family.

Of the proteins listed in Table 3, only glucose:fructose oxidoreductase (GFO), which catalyses the $NADP^+$-linked oxidation of glucose and the NADPH-linked reduction of fructose,

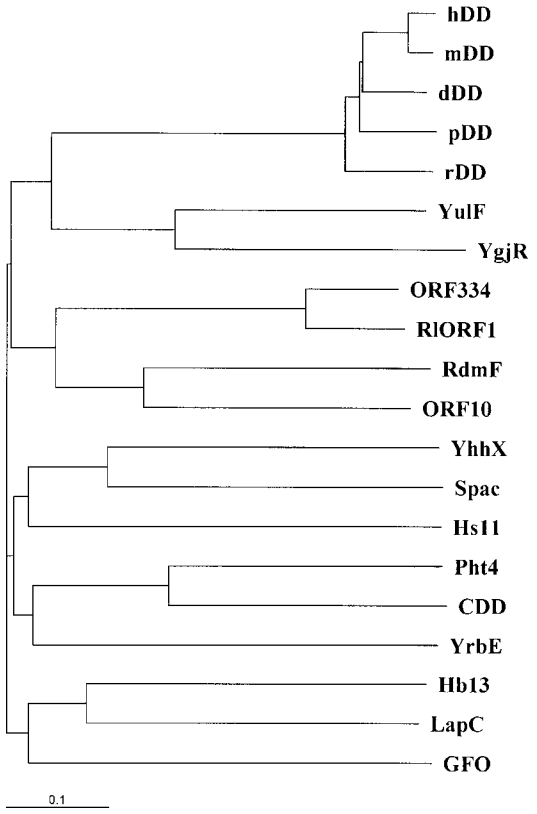

Figure 4 Phylogenetic tree of the medium-chain dehydrogenase/reductase family and related proteins

The tree was calculated with Clustal W [31] and drawn with TreeView [32]. The branch lengths are proportional to mutational distances (the scale bar represents the mutation of 10% of the positions). The designations and accession numbers of the proteins are listed in Table 3.

has been functionally and structurally characterized. Site-directed mutagenesis and X-ray crystallographic studies [46,47] have revealed that this enzyme has a typical dinucleotide-binding domain, i.e. a Rossmann (β-α-β) fold [48] with a fingerprint sequence, an N-terminal Gly-Xaa-Gly-Xaa-Xaa-Ala sequence

and a Glu-Lys-Pro motif for interaction with the carboxamide group of the nicotinamide ring and the nicotinamide ribose. Also reported in the studies were several proteins (some proteins overlap with those in Table 3) with sequence similarities to only the N-terminal half of GFO, in which only the Glu-Lys-Pro motif is almost completely conserved. Similarly, the Glu-Lys-Pro motif (at positions 96–98 in dimeric DD) is present in the dimeric DDs and proteins listed in Table 3 and there is no fingerprint matching the Gly-Xaa-Gly-Xaa-Xaa-Ala sequence and the other consensus sequences for coenzyme binding [48–50] at their N-termini. However, because the secondary structures of the proteins listed in Figure 2 predicted the β-α-β-fold around the N-terminal regions from Gly-6 to Gly-11 of dimeric DDs (results not shown), it is possible that dimeric DD also has a coenzyme-binding domain composed of the N-terminal β-α-β fold and the Glu-Lys-Pro motif. However, the substrate-binding domain of GFO has not been elucidated but two residues, Tyr-217 and Tyr-296, have been proposed as candidates for the catalytic residue on the basis of the three-dimensional structure of the enzyme [47]. Of these, only Tyr-217 is conserved at position 180 of dimeric DDs and the proteins in Figure 2, but is not present in the other proteins listed in Table 3. The other residues strictly conserved in all the dimeric DDs and proteins in Table 3 are His-79 and Gly-172, of which His is able to act as the acid–base catalyst. As described above, chemical modification [45] and our site-directed mutagenesis studies suggest that His-79 is important in catalysis or substrate binding in dimeric DD. Thus GFO and dimeric DD might be distinct with respect to their catalytic residues, because the subunit structures and catalytic properties of the two enzymes are different: GFO is tetrameric and catalyses a Ping Pong type of reaction in the oxidoreduction of glucose and fructose through the tightly bound coenzyme [46], whereas dimeric DD is a simple dehydrogenase that follows an ordered Bi Bi mechanism [51]. In the proposed medium-chain dehydrogenase/reductase family (Figure 2), several residues other than His-79, the Glu-Lys-Pro motif and Tyr-180 are strictly conserved; these are targets of a future site-directed mutagenesis study to understand further the structural and evolutionary relationship between the members of this family and related proteins, including GFO.

This work was supported by a Grant-in-Aid for Scientific Research (C) from the Ministry of Education, Science, Sports and Culture.

REFERENCES

1 Oesch, F., Glatt, H. R., Vogel, K., Seidel, A., Petrovic, P. and Platt, K. L. (1984) in Biochemical Basis of Carcinogenesis (Greim, H., Jung, R., Kramer, M., Marquardt, H. and Oesch, F., eds.), pp. 23–31, Raven Press, New York
2 Smithgall, T. E., Harvey, R. G. and Penning, T. M. (1986) J. Biol. Chem. **261**, 6184–6191
3 van Heyningen, R. and Pirie, A. (1967) Biochem. J. **102**, 842–852
4 Flowers, L., Bleczinski, B., Burczynski, M. E., Harvey, R. G. and Penning, T. M. (1996) Biochemistry **35**, 13664–13672
5 Penning, T. M., Mukharji, I., Barrows, S. and Talalay, P. (1984) Biochem. J. **222**, 601–611
6 Wörner, W. and Oesch, F. (1984) FEBS Lett. **170**, 263–267
7 Hara, A., Hasebe, K., Hayashibara, M., Matsuura, K., Nakayama, T. and Sawada, H. (1986) Biochem. Pharmacol. **35**, 4005–4012
8 Hara, A., Taniguchi, H., Nakayama, T. and Sawada, H. (1990) J. Biochem. (Tokyo) **108**, 250–254
9 Deyashiki, Y., Oshima, K., Nakanishi, M., Sato, K., Matsuura, K. and Hara, A. (1995) J. Biol. Chem. **270**, 1041–1046
10 Hara, A., Nakayama, T., Harada, T., Kanazu, T., Shinoda, M., Deyashiki, Y. and Sawada, H. (1991) Biochem. J. **275**, 113–119
11 Jez, J. M., Bennett, M. J., Schlegel, B. P., Lewis, M. and Penning, T. M. (1997) Biochem. J. **326**, 625–636
12 Hara, A., Mouri, K. and Sawada, H. (1987) Biochem. Biophys. Res. Commun. **145**, 1260–1266
13 Sato, K., Inazu, A., Yamaguchi, S., Nakayama, T., Sawada, H. and Hara, A. (1993) Arch. Biochem. Biophys. **307**, 286–294
14 Nakayama, T., Sawada, H., Deyashiki, Y., Kanazu, T., Hara, A., Shinoda, M., Matsuura, K., Bunai, Y. and Ohya, I. (1990) in Enzymology and Molecular Biology of Carbonyl Metabolism 3 (Weiner, H., Wermuth, B. and Crabb, D. W., eds.), pp. 187–196, Plenum Press, New York
15 Sato, K., Nakanishi, M., Deyashiki, Y., Hara, A., Matsuura, K. and Ohya, I. (1994) J. Biochem. (Tokyo) **116**, 711–717
16 Yamaguchi, S., Inazu, A., Deyashiki, Y., Nakayama, T., Sato, K., Miyabe, Y. and Hara, A. (1994) J. Biochem. (Tokyo) **115**, 493–496
17 Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzàlez-Duarte, R., Jeffery, J. and Ghosh, D. (1995) Biochemistry **34**, 6003–6013
18 Nakagawa, M., Matsuura, K., Hara, A., Sawada, H., Bunai, Y. and Ohya, I. (1989) J. Biochem. (Tokyo) **106**, 1104–1109
19 Platt, K. L. and Oesch, F. (1977) Synthesis **7**, 449–450
20 Platt, K. L. and Oesch, F. (1983) J. Org. Chem. **48**, 265–268
21 Khadem, H. E., Horton, D., Meshreki, M. H. and Nashed, M. A. (1971) Carbohydr. Res. **17**, 183–192
22 Coligan, J. E. (1997) in Current Protocols in Protein Science (Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W. and Wingfield, P. T., eds.), pp. 11.1.1–11.1.19, John Wiley, New York
23 Hunyu, T. V., Young, R. A. and Davis, R. W. (1985) in DNA Cloning: A Practical Approach, vol. 1 (Glover, D. M., ed.), pp. 49–78, IRL Press, Oxford
24 Deyashiki, Y., Ogasawara, A., Nakayama, T., Nakanishi, M., Miyabe, Y. and Hara, A. (1994) Biochem. J. **299**, 545–552
25 Nakanishi, M., Deyashiki, Y., Ohshima, K. and Hara, A. (1995) Eur. J. Biochem. **228**, 381–387
26 Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
27 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
28 Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350–4354
29 Pearson, W. R. and Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2444–2448
30 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
31 Tompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Nucleic Acids Res. **22**, 4673–4680
32 Page, R. D. M. (1996) Comput. Appl. Biosci. **12**, 257–358
33 Rost, B. and Sander, C. (1993) J. Mol. Biol. **232**, 584–599
34 Kozak, M. (1986) Cell **44**, 283–292
35 Shinoda, M., Hara, A., Nakayama, T., Deyashiki, Y. and Yamaguchi, S. (1992) J. Biochem. (Tokyo) **112**, 840–844
36 Wollenberg, P. and Ullrich, V. (1980) in Extrahepatic Metabolism of Drugs and Other Foreign Compounds (Theodore, G., ed.), pp. 267–276, Spectrum Publications, New York
37 Kashfi, K., McDougall, C. J. and Dannenberg, A. J. (1995) Clin. Pharmacol. Ther. **58**, 625–630
38 Oesch, F., Glatt, H. and Schmassmann, H. (1977) Biochem. Pharmacol. **26**, 603–607
39 Degenhardt, T. P., Brinkmann-Frye, E., Thorpe, S. R. and Baynes, J. W. (1998) in The Maillard Reaction in Foods and Medicine (O'Brien, J., Nursten, H. E., Crabbe, M. J. C. and Ames, J. M., eds.), pp. 3–10, Royal Society of Chemistry, Cambridge
40 Koschinsky, T., He, C.-J., Mitsuhashi, T., Bucala, R., Liu, C., Buenting, C., Heitmann, K. and Vlassara, H. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 6474–6479
41 Maser, E. (1995) Biochem. Pharmacol. **49**, 421–440
42 Neidle, E., Hartnett, C., Ornston, N., Bairoch, A., Pekik, M. and Harayama, S. (1992) Eur. J. Biochem. **204**, 113–120
43 Penning, T. M. (1996) J. Endocrinol. **150**, S175–S187
44 Tarle, I., Borhani, D. W., Wilson, D. K., Quiocho, F. A. and Petrash, J. M. (1993) J. Biol. Chem. **268**, 25687–25693
45 Shinoda, M., Hara, A., Nakayama, T., Deyashiki, Y. and Sawada, H. (1992) J. Biochem. (Tokyo) **112**, 834–839
46 Wiegert, T., Sahm, H. and Sprenger, G. A. (1997) J. Biol. Chem. **272**, 13126–13133
47 Kingston, R. L., Scopes, R. K. and Baker, E. N. (1996) Structure **4**, 1413–1428
48 Wierenga, R. K., De Maeyer, M. C. H. and Hol, W. G. J. (1985) Biochemistry **24**, 1346–1357
49 Hanukoglu, I. and Gutfinger, T. (1989) Eur. J. Biochem. **180**, 479–484
50 Allaire, M., Li, Y., MacKenzie, R. E. and Cygler, M. (1998) Structure **6**, 173–182
51 Hara, A., Shinoda, M., Kanazu, T., Nakayama, T., Deyashiki, Y. and Sawada, H. (1991) Biochem. J. **275**, 121–126

Received 2 March 1999/14 June 1999; accepted 15 July 1999