Molecular analysis of two closely related mouse aldehyde dehydrogenase genes: identification of a role for *Aldh1*, but not *Aldh-pb*, in the biosynthesis of retinoic acid

Lily C. HSU*, Wen-Chung CHANG*, Ines HOFFMANN† and Gregg DUESTER†[1]

*Department of Biochemical Genetics, Beckman Research Institute of the City of Hope, Duarte, CA 91010, U.S.A., and †Gene Regulation Program, Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

Mammalian class I aldehyde dehydrogenase (ALDH1) has been implicated as a retinal dehydrogenase in the biosynthesis of retinoic acid, a modulator of gene expression and cell differentiation. As the first step towards studying the regulation of ALDH1 and its physiological role in the biosynthesis of retinoic acid, mouse ALDH1 cDNA and genomic clones have been characterized. During the cloning process, an additional closely related gene was also isolated and named *Aldh-pb*, owing to its high amino acid sequence identity (92 %) with the rat phenobarbitol-inducible ALDH protein (ALDH-PB). *Aldh1* spans about 45 kb in length, whereas *Aldh-pb* spans about 35 kb. Both genes are composed of 13 exons, and the positions of all the exon/intron boundaries are conserved with those of human *ALDH1*. The promoter regions of *Aldh1* and *Aldh-pb* demonstrate high sequence similarity with those of human *ALDH1* and rat *ALDH-PB*. Expression of *Aldh1* and *Aldh-pb* is tissue-specific, with mRNAs for both genes being found in the liver, lung and testis, but not in the heart, spleen or muscle. Expression of *Aldh-pb*, but not *Aldh1*, was also detected at high levels in the kidney. *Aldh1* and *Aldh-pb* encode proteins of 501 amino acids with 90 % positional identity. To examine the relative roles of these two enzymes in retinoic acid synthesis *in vivo*, *Xenopus* embryos were injected with mRNAs encoding these enzymes to assay the effect on conversion of endogenous retinal into retinoic acid. Injection of ALDH1, but not ALDH-PB, mRNA stimulated retinoic acid synthesis in *Xenopus* embryos at the blastula stage. Thus our results indicate that *Aldh1* can function in retinoic acid synthesis under physiological conditions, but that the closely related *Aldh-pb* does not share this property.

Key words: promoter, retinal, retinoid signalling, vitamin A, *Xenopus*.

INTRODUCTION

Retinoic acid and its derivatives have been the focus of considerable attention due to their role as potent modulators of gene expression, particularly regulation of vertebrate homeobox genes, and their role in cell differentiation and development [1,2]. Their action is mediated by nuclear receptors (e.g. retinoic acid receptors and retinoid X receptors), which are DNA-binding transcriptional regulators [3]. Retinoic acid deficiency or excess has been linked with developmental disorder symptoms [4]. Maintenance of homoeostatic retinoic acid levels in tissues is thus essential to secure normal embryogenesis, organogenesis and adult life.

Retinoic acid homoeostasis is achieved by multiple-step regulation of retinoid metabolism in retinoic acid-responsive cells, including the uptake of plasma retinol (the precursor of retinoic acid), and the biosynthesis and degradation of retinoic acid. The major biosynthetic pathway of retinoic acid from retinol (vitamin A) involves two sequential steps: (a) a reversible dehydrogenation into retinal catalysed by either cytosolic retinol dehydrogenases, which are members of the alcohol dehydrogenase family, or by microsomal retinol dehydrogenases, which are members of the short-chain dehydrogenase/reductase family; and (b) an irreversible oxidation of retinal to retinoic acid catalysed by cytosolic retinal dehydrogenases, which are members of the aldehyde dehydrogenase (ALDH) family, as reviewed previously [5]. Activity assays *in vitro* have identified several ALDHs able to oxidize retinal, including human class I aldehyde dehydrogenase (ALDH1) [6,7], mouse ALDH1 (previously known as Ahd-2) [8,9], rat ALDH1 ([10] also called RALDH [11] or RalDH-I [12]), mouse RALDH2 (previously called V2 [13]) and its rat homologue RalDH-II [14] and perhaps also its human homologue (called ALDH11) [15], and, finally, mouse V1 enzyme, localized in the embryonic ventral retina, which is distinct from ALDH1 and RALDH2 [16]. Members of the ALDH family are involved in the oxidation of a wide variety of exogenous and endogenous aldehydes in addition to retinal [6,17,18].

Although ALDH1 was the first human ALDH to be purified and characterized [19], its physiological role has long been unclear, owing to its broad substrate specificity (e.g. acetaldehyde, retinal, aldophosphamide and 11-hydroxythromboxane B_2), as well as its ability to bind various molecules (e.g. androgens and daunorubicin) [6,20–26]. It has been shown that purified human liver ALDH1 has extremely high activity for oxidation of free all-*trans*-retinal to all-*trans*-retinoic acid, with a K_m value of 0.06 μM at pH 7.5 [7]. Evidence supporting retinoic acid synthesis as a major role for ALDH1 includes studies showing that the catalytic efficiency (V_{max}/K_m) of human ALDH1 for retinal oxidation is about 700 times higher than that for either acetaldehyde or aldophosphamide oxidation, and studies showing that ALDH1 is the sole isoenzyme among six purified liver ALDHs that can mediate NAD^+-dependent retinal oxidation [6].

A mouse ALDH1 homologue (previously called Ahd-2) has been identified upon the basis of its substrate preferences and its coding sequence identity with human ALDH1 [27,28]. Six cytosolic ALDH isoenzymes were identified in adult mouse liver, and of these it was determined that ALDH1 mediates 95 % of NAD^+-dependent retinal oxidation in liver, and has a K_m value

Abbreviations used: ALDH, aldehyde dehydrogenase; ALDH-PB, phenobarbitol-inducible ALDH; ALDH1, mammalian class I ALDH; *ALDH1*, *Aldh1* and *Aldh-pb,* genes encoding human class I ALDH, mouse class I ALDH and the mouse homologue of rat ALDH-PB respectively; MMR, Marc's modified Ringer's solution; poly(A)$^+$, polyadenylated; UTR, untranslated region.

[1] To whom correspondence should be addressed (e-mail duester@burnham-inst.org).

© 1999 Biochemical Society

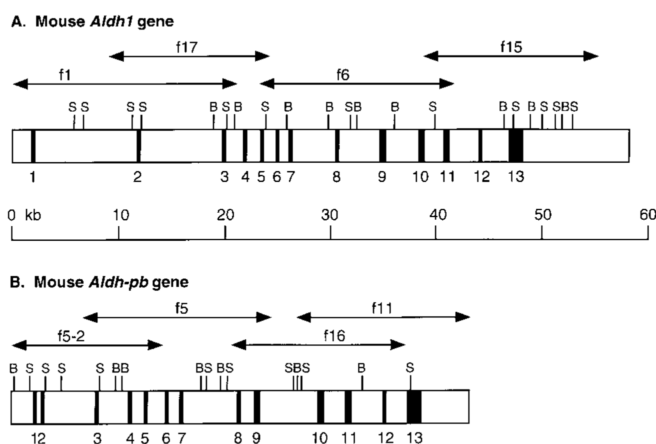

Figure 1 Organization of mouse *Aldh1* and *Aldh-pb* genes

Restriction maps of the *Aldh1* (**A**) and *Aldh-pb* (**B**) genes are shown as rectangular bars. A scale in kb is shown between the maps for each gene. The corresponding exons of the two genes are assigned with the same numbers and represented by vertical solid bars, and introns are represented by the open areas between the solid bars. Above each gene are indicated the overlapping phage genomic clones containing the genomic regions (see the Materials and methods section for details) defined by arrows. Restriction maps for each of the indicated enzymes are shown above the open bars: S, *Sst*I; B, *Bam*HI.

for free retinal of 0.7 μM [8]. In addition, cells transfected with mouse ALDH1 cDNA can convert retinal into retinoic acid, supporting its role as a retinal dehydrogenase [29]. Mouse ALDH1 was also found in the embryonic and adult dorsal retina, an organ rich in retinoic acid [30]. In addition, a rat ALDH1 homologue isolated from either kidney (also called RALDH) [11,31] or liver (also called RalDH-I) [10,12,32] demonstrates high amino acid sequence identity with mouse ALDH1, and has high activity for retinal oxidation.

The kinetic studies provide *in vitro* support for human ALDH1, as well as its mouse and rat homologues, in the pathway of retinoic acid biosynthesis. In the present study we have begun a genetic examination of ALDH1 function and regulation by cloning and characterizing the mouse *Aldh1* gene. A closely related gene, *Aldh-pb* [a gene encoding the mouse homologue of rat phenobarbitol-inducible ALDH], was also isolated and characterized. Both *Aldh1* and *Aldh-pb* were examined for their ability to function in retinoic acid synthesis *in vivo* using a *Xenopus* embryo assay. Our results provide firm support for ALDH1 as a retinoic acid-synthesizing enzyme under physiological conditions, and show that the closely related ALDH-PB cannot perform this function.

MATERIALS AND METHODS

Genomic cloning

A mouse genomic library in FIXII prepared from DNA of mouse strain 129 (Stratagene Cloning Systems, La Jolla, CA, U.S.A) was screened with a probe containing the human *ALDH1* coding region in order to clone the mouse *Aldh1* gene. Genomic walking with intron probes derived from an isolated *Aldh1* clone (f15) was subsequently carried out to obtain overlapping *Aldh1* clones (f6, f17 and f1). During the process of cloning *Aldh1*, we also identified two overlapping clones (f5 and f16) of a novel mouse gene, *Aldh-pb*. A 5′-end overlapping clone for *Aldh-pb*, f5-2, was isolated from the library with a human *ALDH1* exon 1 probe, and a 3′-end overlapping clone, f11, was identified by genomic walking using as a probe a *Bam*HI fragment from f16 containing exon 12. All hybridizations were carried out essentially as described previously [33] in the presence of 50% (v/v) formamide at 42 °C for 18 h, followed by 2 × 15 min washes in 0.1 × SSC [15 mM NaCl/1.5 mM sodium citrate (pH 7.0)] and 0.1% (w/v) SDS at 55 °C.

PCR cloning of mouse ALDH1 and ALDH-PB cDNAs

DNA purified from a mouse testis cDNA library (Clontech Laboratories, Palo Alto, CA, U.S.A.) and RNA from mouse liver were used as templates to clone the ALDH1 and ALDH-PB cDNAs. PCR cycles were performed at 94 °C for 45 s, 56 °C for 45 s and 72 °C for 1 min. The 5′- and 3′-portions of the cDNAs were amplified with coding-region-specific and λgt11 vector-end-specific probes. The PCR products were separated on a gel, Southern-blotted and hybridized with coding-region probes. The longest positive bands were eluted from the gel, subcloned into pBluescript vector and sequenced. GenBank accession numbers are M36069 for mouse *Aldh1* and U96401 for mouse *Aldh-pb*.

Restriction maps and exon/intron junctions of *Aldh1* and *Aldh-pb* genes

DNA was prepared from phage clones using Qiagen Lambda DNA Prep columns (Qiagen, Valencia, CA, U.S.A.). The restriction maps of each genomic clone were determined by single- and double-restriction digestion, and by hybridization with exon-specific probes and vector-end probes (T3- and T7-map primers). The restriction endonuclease fragments containing exonic sequences were isolated by electroelution and subcloned into pBluescript+ vector. The exonic restriction fragments of suitable sequencing size were further subcloned and/or sequenced. The exon–intron junctions were identified by comparing the genomic sequence with the cDNA sequence. PCR between exons was performed to estimate or confirm intron sizes. Introns of less than 5.5 kb were amplified in a DNA thermal cycler (PTC-100; MJ Research, Watertown, MA, U.S.A.) using the following programme: 94 °C for 3 min, and then 30 cycles of 94 °C for 45 to 60 s, 52 °C to 60 °C for 1 min, and 72 °C for 1 to 3 min. The exon-specific primers were derived from either the cDNA or exon sequence. The primers were also used to sequence intron/exon boundaries of subcloned and PCR-amplified DNA fragments. DNA-sequence determination of double-stranded DNA was performed by thermocycle sequencing with *Taq* polymerase.

Tissue distribution analysis

Northern blot analysis was performed with a mouse Multiple Tissue Northern blot (Clontech). The 3′-untranslated region (UTR) sequences of either *Aldh1* or *Aldh-pb* were used as specific hybridization probes. Probe labelling and hybridization were performed by standard methods [33].

Retinoic acid bioassay in *Xenopus* embryos

Full-length mouse ALDH1 and ALDH-PB cDNAs were subcloned into plasmid pSP65 (Promega, Madison, WI, U.S.A.) in the sense direction for *in vitro* transcription from the SP6 promoter. Plasmids were linearized and subjected to *in vitro* transcription with SP6 RNA polymerase, 5′-capping with 7-methylguanosine, and 3′-polyadenylation with *Escherichia coli* poly(A) polymerase to produce full-length mRNAs [33]. mRNAs were brought to a concentration of 0.2 μg/μl in water, and 0.4 μg was used for *in vitro* translation in a rabbit reticulocyte lysate (Stratagene) to verify the ability to produce full-length proteins.

Xenopus laevis embryos were produced by artificial fertilization and staged as described previously [34]. Embryos were placed in

© 1999 Biochemical Society

Table 1 Splice junction sites of the mouse *Aldh1* gene (a) and the mouse *Aldh-pb* gene (b)

Exon sequences are in capital letters; intron sequences are in lower case letters. Introns are positioned by applying the gt/ag rule. Exon 1 starts from the putative transcriptional initiation site. The codon phases of the intron boundaries are indicated: O, introns inserted between codons; I, introns inserted after the first nucleotide of a codon; II, introns inserted after the first nucleotide of a codon. Amino acids encoded at the splice sites are numbered from the Ser at the 2nd position C-terminal side from the initiator Met.

(a)

Exon	Exon size (bp)	Sequence at exon–intron junction						
		5′ Splice donor		Intron (kbp)	3′ Splice donor		Codon phase	aa interrupted
1.	(110)	–,CAT,ACC,AAG	gtgagtcact	1. (10.0)	ataatcacag	ATC,TTC,ATA	O	Lys21/Ile
2.	(105)	–,GGG,GAC,AAG	gtaaattttc	2. (7.5)	tgatttccag	GCT,GAT,GTT	O	Lys56/Ala
3.	(141)	–,CTG,CTA,GCT	gtgagtacta	3. (1.9)	gtttttcag	ACA,ATG,GAG,	O	Ala103/Thr
4.	(130)	ATA,CCA,AGT,G	gtaagtatac	4. (1.5)	gttttgccag	TGG,AAT,TTT	I	Asp147
5.	(62)	–,ATC,ATC,CCC	gtaagttctc	5. (1.4)	tgctttctag	TGG,AAT,TTT,	O	Pro167/Trp
6.	(129)	–,ATA,AAA,GAG	gtgagtttcc	6. (1.2)	ctgcttgtag	GCA,GGG,TTT,	O	Glu210/Ala
7.	(114)	–,TCA,ACA,CAG	gtagggccaa	7. (4.8)	gtgatttcag	GTT,GGC,AAG,	O	Gln248/Val
8.	(103)	GAT,GCC,GAC,T	gtgagtacag	8. (3.3)	ttctctgcag	TG,GAC,ATT,G	I	Leu283
9.	(185)	–,GGC,CCT,CAG	gcaagtatac	9. (3.6)	tttcaactag	ATT,GAC,AAG,	O	Gln344/Ile
10.	(165)	–,AAA,GAG,GAG	gtaatgtttc	10. (2.2)	ttctgtatag	ATA,TTT,GGA,	O	Glu399/Ile
11.	(158)	GG,GTG,GTG,TG	gtaagtgccg	11. (3.2)	tttgtttaag	G,GTT,AAC,TG	II	Trp452
12.	(75)	GA,AGA,GAA,CT	gtgagtaaag	12. (2.7)	cccccttag	G,GGT,GAA,CA	II	Leu477
13.	(> 570)							
		AG	gt(a/g)agt		(c/t)$_n$ncag	G		

(b)

Exon	Exon size (bp)	Sequence at exon–intron junction						
		5′ Splice donor		Intron (kbp)	3′ Splice donor		Codon phase	aa interrupted
1.	(110)	–,CAT,ACC,AAG	gtgagtcact	1. (0.38)	atctcaacag	ATC,TTT,ATA	O	Lys21/Ile
2.	(105)	–,GGG,GAC,AAG	gtatgttgcc	2. (4.8)	tgatttccag	GCT,GAT,GTT	O	Lys56/Ala
3.	(141)	–,CTG,CTG,GCT	gtgagtatta	3. (3.1)	tttccttcag	ACA,ATG,GAA,	O	Ala103/Thr
4.	(130)	ATA,CCC,AGT,G	gtaagtccat	4. (1.4)	tgtttttcag	AT,GGA,AAC,A	I	Asp147
5.	(62)	–,ATC,ATC,CCT	gtaagttttc	5. (1.7)	taattttcag	TGG,AAT,GGT,	O	Pro167/Trp
6.	(129)	–,ATA,AAA,GAG	gtgagtttcc	6. (1.1)	taacttgtag	GCA,GGG,TTT,	O	Glu210/Ala
7.	(114)	–,TCA,ACA,GAG	gtagggccat	7. (5.2)	gtgatttcag	GTT,GGC,AAA,	O	Gln248/Val
8.	(103)	GAT,GCC,GAC,T	gtgagtacag	8. (1.0)	tcctttgcag	TG,GAC,AGT,G	I	Leu283
9.	(185)	–,GGT,CCT,CAG	gtgagtaaac	9. (6.4)	tcccaactag	ATT,GAC,AAG,	O	Gln344/Ile
10.	(165)	–,AAA,GAG,GAG	gtaatatttc	10. (2.1)	cctttggtag	ATA,TTT,GGA,	O	Glu399/Ile
11.	(158)	GG,ATG,GTG,TG	gtaagtgccg	11. (3.3)	tgtgttttag	G,GTG,AAC,TG	II	Trp452
12.	(75)	GG,CGA,GAA,CT	gtaagttaaa	12. (2.7)	ccctctctag	G,GGC,GAA,CA	II	Leu477
13.	(> 583)							
		AG	gt(a/g)agt		(c/t)$_n$ncag	G		

1 × Marc's modified Ringer's solution (MMR)/5 % (v/v) Ficoll-400 for micro-injection of mRNA, as described previously [35]. Various amounts of mRNA (4.6–46.0 nl at a concentration of 0.2 ng/nl in water) were micro-injected into the vegetal pole of embryos at the 2–4 cell stages using a micropipette pre-filled with light mineral oil. After injection (4 h), embryos were transferred to 0.1 × MMR and incubated to stage 8 (blastula).

Retinoic acid was detected using a bioassay that employs the retinoic acid reporter cell line F9-RARE-*lacZ* [36]. This cell line was derived from the stable transfection of mouse F9 cells with a transgene containing a retinoic acid response element driving expression of *lacZ*. F9-RARE-*lacZ* detects the presence of retinoic acid in cultured embryo explants by detecting diffusion of this small lipid-soluble molecule from the embryo to the surrounding reporter cells. Bioassay of *Xenopus* retinoic acid was performed as described previously for mouse embryos [37]. Briefly, F9-RARE-*lacZ* cells were grown to 80–90 % confluence, at which time embryos were placed on top of the reporter cell monolayer, incubated for 18 h, and then fixed in 1 % (v/v) glutaraldehyde and assayed for β-galactosidase activity produced by *lacZ* expression. *Xenopus* embryos placed upon the reporter cells and incubated under standard mammalian cell culture conditions did not develop further, but remained intact for the duration of the assay. In our hands, the F9-RARE-*lacZ* cells are sensitive to as little as 0.1 nM all-*trans*-retinoic acid added to the media during an 18 h incubation, and there is virtually no background detection of β-galactosidase in cells incubated without added retinoic acid.

RESULTS

Organization of the mouse *Aldh1* and *Aldh-pb* genes

Four overlapping mouse genomic clones (f1, f17, f6 and f15) that together contain the entire coding sequence of *Aldh1* were cloned and characterized (Figure 1A). *Aldh1* spans 45 kb and contains 13 exons. During the process of cloning *Aldh1*, we also identified

© 1999 Biochemical Society

```
                                                                     M   S   S   P   A   Q   P   A   V   P   A   P   L   A   N    14
 at ttt gct gag cct gtc act tgt gtt cca gga gcc aaa cca gca ATG TCT TCA CCT GCA CAA CCT GCA GTC CCT GCC CCG CTG GCC AAC (89)
 .a ... ... ... ... ... ..c ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ..A ... ... G.. (89)
                                                                                                                     (D)

 L   K   I   Q   H   T   K   I   F   I   N   N   E   W   H   D   S   V   S   S   K   K   F   P   V   L   N   P   A   T    44
TTG AAG ATT CAA CAT ACC AAG ATC TTT ATA AAC AAT GAA TGG CAT GAT TCG GTG AGC AGC AAG AAA TTT CCT GTC CTT AAC CCT GCA ACT (179)
... ... ... ... ... ... ... ... ..C ... ... ... ... ... ..C A.. ..A ... ... G.. ... ... ... ..A ..T ... ... ... ... ... (179)
                                                            (N)             (G)

 E   E   V   I   C   H   V   E   E   G   D   K   A   D   V   D   K   A   V   K   A   A   R   Q   A   F   Q   I   G   S    74
GAG GAG GTC ATC TGC CAC GTG GAA GAA GGG GAC AAG GCT GAT GTT GAC AAA GCT GTG AAG GCC GCA AGA CAG GCT TTC CAG ATT GGC TCC (269)
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ..T (269)

 P   W   R   T   M   D   A   S   E   R   G   R   L   L   N   K   L   A   D   L   M   E   R   D   R   L   L   L   A   T   104
CCA TGG CGC ACC ATG GAT GCT TCA GAG AGG GGC CGC CTG CTG AAC AAG CTG GCT GAC TTA ATG GAG AGA GAT CGT CTG CTG CTG GCT ACA (359)
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... (359)

 M   E   S   M   N   A   G   K   V   F   A   H   A   Y   L   L   D   V   E   I   S   I   K   A   L   Q   Y   F   A   G   134
ATG GAA TCG ATG AAT GCT GGG AAA GTC TTT GCT CAT GCA TAC CTG TTG GAT GTA GAG ATC AGC ATA AAA GCA TTA CAG TAC TTC GCA GGC (449)
... ..G G.A C.C ... .G. ... ... ... ... ..C A.. ... ... T.. .C. ... T.. .GA GG. T.. ... ... ... ... A.. ... .GT ... ... (449)
        (A) (L)     (G)                     (N)             (S)     (L) (G) (G) (C)                 (K)     (C)

 W   A   D   K   I   H   G   Q   T   I   P   S   D   G   N   I   F   T   Y   T   R   R   E   P   I   G   V   C   G   Q   164
TGG GCT GAC AAG ATC CAT GGC CAA ACA ATA CCC AGT GAT GGA AAC ATA TTC ACT TAT ACA AGG CGT GAA CCT ATT GGG GTG TGT GGC CAA (539)
... ... ... ... ..T ... ..T ... ... ... ..A ... ... ... G.. ..T ... ... ... ... ..A ... ... ... ... ..A ... ... ... ... (539)
                                                        (D)

 I   I   P   W   N   G   P   L   I   I   F   T   W   K   L   G   P   A   L   S   C   G   N   T   V   V   V   K   P   A   194
ATC ATC CCT TGG AAT GGT CCA TTG ATT ATA TTC ACT TGG AAG TTA GGC CCT GCC CTT AGC TGT GGG AAC ACT GTG GTT GTC AAG CCA GCA (629)
... ... ..C ... ... TT. ... A.. C.C ..G ... .T. ... ... A.. ... ... ... ... ... ... ... ..T ..C ... ... ... ... ... ... (629)
                    (F)     (M) (L) (M)     (I)         (I)

 E   Q   T   P   L   T   A   L   H   M   A   S   L   I   K   E   A   G   F   P   P   G   V   V   N   I   V   P   G   Y   224
GAG CAA ACT CCT CTC ACA GCT CTT CAC ATG GCA TCT TTA ATA AAA GAG GCA GGG TTT CCT CCT GGC GTG GTG AAC ATT GTC CCT GGT TAT (719)
... ... ... ... ... ..G ... ... ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... (719)
                                    (L)

 G   P   T   A   G   G   A   I   S   S   H   M   D   I   D   K   V   S   F   T   G   S   T   E   V   G   K   L   I   K   254
GGG CCA ACT GCA GGG GGA GCC ATC TCC TCC CAC ATG GAC ATC GAC AAA GTG TCC TTC ACA GGA TCA ACA GAG GTT GGC AAA TTA ATC AAG (809)
... ... ... ... ... .C. ... ... ... ..T ... ... ..T G.. ... ..G ... G.. ... ..T ... ... ... C.. ... ... ..G ... ... ... (809)
                    (A)                             (V)             (A)                     (Q)

 E   A   A   G   K   S   N   L   K   R   V   T   L   E   L   G   G   K   S   P   C   I   V   F   A   D   A   D   L   D   284
GAA GCT GCA GGG AAA AGC AAT CTG AAG AGA GTC ACC CTG GAG CTG GGG GGA AAG AGC CCT TGC ATT GTG TTT GCA GAT GCC GAC TTG GAC (899)
... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... (899)

 S   A   V   E   F   A   H   Q   G   V   F   F   H   Q   G   Q   I   C   V   A   A   S   R   L   F   V   E   E   S   I   314
AGT GCT GTT GAG TTT GCA CAC CAA GGA GTG TTC TTC CAC CAG GGT CAG ATT TGT GTT GCA GCG TCC AGG CTT TTT GTT GAG GAA TCA ATT (989)
.T. ... ... ... ... ... ... ..T ... ... ..T .AT ..T ..A ..C ..A TGC ... ..C ... ..A ... C.. A.. ... ... ... ..G ... G.. (989)
(I)                         (H)             (Y)                 (C)                         (I)                     (V)

 Y   D   E   F   V   R   R   S   V   E   R   A   K   K   Y   I   L   G   N   P   L   N   S   G   I   N   Q   G   P   Q   344
TAC GAT GAG TTT GTG AGG AGG AGT GTG GAG CGG GCT AAG AAA TAC ATT CTA GGA AAT CCT CTG AAC TCC GGA ATA AAT CAA GGT CCT CAG (1079)
..T ... ... ... ... .AA ... ... ..T ... ..A ... ... ... ..T G.. ..T ... ... ... ... .C. C.A ... ... ... ... ..C ... ... (1079)
                    (K)                                     (V)                     (T) (P)

 I   D   K   E   Q   H   N   K   I   L   G   L   I   E   S   G   K   K   E   G   A   K   L   E   C   G   G   G   R   W   374
ATT GAC AAG GAA CAA CAC AAT AAA ATA CTT GGT CTC ATT GAG AGT GGG AAG AAA GAA GGA GCC AAG CTG GAG TGT GGT GGA GGT CGC TGG (1169)
... ... ... ... ... ..T G.. ... ... ..C .A. ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ..A ... ... (1169)
                        (D)             (D)

 G   N   K   G   F   F   V   Q   P   T   V   F   S   N   V   T   D   E   M   R   I   A   K   E   E   I   F   G   P   V   404
GGG AAC AAA GGC TTC TTT GTC CAA CCC ACA GTC TTC TCC AAT GTG ACT GAT GAG ATG CGC ATT GCC AAA GAG GAG ATA TTT GGA CCA GTG (1259)
... ... ... ... ... ... ..G ..G ... ... ..G ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... (1259)

 Q   Q   I   M   K   F   K   S   M   D   D   V   I   K   R   A   N   N   T   T   Y   G   L   A   A   G   V   F   T   K   434
CAA CAA ATC ATG AAG TTT AAA TCT ATG GAT GAT GTG ATC AAG AGA GCA AAC AAC ACT ACC TAT GGT TTA GCA GCA GGA GTC TTC ACT AAA (1349)
... ... ... ... ... ... ..G ... G.A ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... C.. ... ... ... (1349)
                                (V)                                                                     (L)

 D   L   D   K   A   I   T   V   S   S   A   L   Q   A   G   M   V   W   V   N   C   Y   L   A   V   P   V   Q   C   P   464
GAC CTG GAT AAG GCC ATC ACT GTG TCA TCT GCT CTG CAG GCT GGG ATG GTG TGG GTG AAC TGC TAT TTG GCT GTC CCT GTC CAA TGC CCA (1439)
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... G.. ... ... ..T ... ... ... A.. ATG T.G T.A .C. ..G ... ..C (1439)
                                                            (V)                         (M) (M) (L) (S) (A)

 F   G   G   F   K   M   S   G   N   G   R   E   L   G   E   H   G   L   Y   E   Y   T   E   L   K   T   V   A   M   Q   494
TTT GGT GGA TTC AAG ATG TCT GGA AAT GGG CGA GAA CTG GGC GAA CAT GGT CTT TAT GAA TAC ACT GAG CTC AAG ACA GTT GCA ATG CAA (1529)
..C ... ... ... ... ... ... ... ... ..A A.. ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... A.G (1529)
                                                                                                                    (K)

 I   S   Q   K   N   S   *
ATA TCT CAG AAG AAC TCA TAA agaagccagc agagtaaaga gaaactctca gcagtggctc cacatctcat acagtcacca g-aaagtggtg ttttgctata   (1630)
... ... ... ... ... ..C ... .......... .......... .......... .........a ........c. .t...a.... .c.t...c... ....at....   (1631)

agttcttctt ctgttgagtt ctttaaacgt aaggaattca tcagtattac tgttaatcat agaaaagacg tagcttaatc tggcagagtc agt-gtcttct aatccgtgaa (1740)
.t.......c .a.....-.. tc....g.aa .......... .....g.... ...c.c..... a....c.t. .g........ caa....t.. .t.cac..... ...at....c (1741)

cccaattcat aacccagaac aaaaggatac attaaggtgc aagctctctg taactctgtc atgataaggt gctttccatt gtagctactt gtctagagta cacacttggc (1850)
....g...t. .t..a....t .g.......g ..at.a.c.. .......... ......c... ....cc.... .......... .......... a....aca.. .t..t....t (1851)

aaggaggacc agtcctgatt taagctctgt cccacactga ttccttgaac tatgtaccat gcatgataac tgaaaagtta gctgctctat tcctcggggt tattcctctg (1960)
g........t ...tg...c. .......... ...t..g... c........g ..ctc....c a...a..g.. ..c.g...c. ........g. ...c.a..-- ---------- (1949)

agtgttgtga aatattttcc agaatgccat gactgcttgt caaatgaaat gcctccctgt aattcgaatg caaagcttaa taaaggcaat ctcgcatgat (2060)
--........ .........t ......t... .c........ .......... ....ag.... ....a...c. .......... .......cc ..t....... (2047)
```

© 1999 Biochemical Society

Table 2 Amino acid sequence identities among ALDH coding regions

	Mouse ALDH-PB	Mouse ALDH1	Rat ALDH-PB	Rat ALDH1	Human ALDH1
Mouse ALDH-PB	—	90	92	91	84
Mouse ALDH1		—	89	96	87
Rat ALDH-PB			—	89	83
Rat ALDH1				—	86
Human ALDH1					—

```
Aldh-pb (mouse)  ATGTCTTAAA AAGTCCAGAA  AATTATAACT  GGCTTTGAGT  GTCCACAGCA -233
Aldh1 (mouse)    T..G...GGT ..A.TA.TTC  .TC.G...A.  A.TGAGC...  C....G.-T.
ALDH-PB (rat)    T..G..-.GT ..A.TA.TTC  .TC.GCC.-.  ..TG.A.GA.  AG..T.C-T-
ALDH1 (human)    TC.G...TGT ..A.TA.TTC  .TC.GC..A.  A.TGCACT..  C....GG-T.

Aldh-pb (mouse)  GTTGCATATA CACCCTTAAC  ACAGATTT--  ---GGCTTGG  TATGTTAATT -188
Aldh1 (mouse)    .--AAC.TC. GGG.TGAGGT  G...G..CCT  TAA..A..TA  C..G.A..GG
ALDH-PB (rat)    ------CT.G G.A.T----T  G......CAG  GAA..A..TA  C..T.A..GG
ALDH1 (human)    C--AA..TCG ATG.TGG.G.  ..T.G...CT  TAA..A..TA  AG.T.A..GT

Aldh-pb (mouse)  CATCTGCCAA TAGTGTAGAA  GAACCTCTTG  GTGGAGTGGT  GCCCTTTCA- -139
Aldh1 (mouse)    ..AAG..TTT CA.CCC-T.G  .TGTTA.-AA  ...-......  ...---.TCA
ALDH-PB (rat)    ..AAG..TCC CT.CCCCT.G  .TTTTA.CAA  ...-...A..  ......----
ALDH1 (human)    ..AAG..TTC CT.CCC-T.G  .TGTTA.-AA  A.A-...A..  .T.G...CCT

                                        CCAAT box       <- Oct1 ->
Aldh-pb (mouse)  -GCCCTGCCC TGAGTCTGCC  CATCCAATCA  TATCCCGATA  TGCAAATGAC  -90
Aldh1 (mouse)    T......... ..........  ..........  ..........  ..........
ALDH PB (rat)    ---....... ..--....G.  ..........  ...G......  ....G-----
ALDH1 (human)    .TTTT...T. .....T..TT  .........G  .....GAG..  .......A.A

                             ATA
Aldh-pb (mouse)  CCTTAGTGCA TGCAGAGAAA  AAGGAG----  -------CAA  GTGCCCTTTC  -51
Aldh1 (mouse)    .......... ......T...  ......----  -------...  .G..T.....
ALDH-PB(rat)     ---------- -.....T...  ......----  -------...  .T..T.....
ALDH1 (human)    .T....CC.G ......T...  ..A.GAACAA  ATAAAGC...  .T..T..A..

Aldh-pb (mouse)  AGAACCATTT TGCTGAGCCT  GTCACTTGTG  TTCCAGGAGC  CAAACCAGCA   -1
Aldh1 (mouse)    .......A.. ..........  .....C....  ..........  ..........
ALDH-PB (rat)    .......A.. ..........  .....C.T..  ..........  ..........
ALDH1 (human)    .......AA. .........A  .....C....  ..........  .G..T...A.
```

Figure 3 Comparison of 5′-flanking region sequences of mouse *Aldh-pb*, mouse *Aldh1*, rat *ALDH-PB* and human *ALDH1* genes

Nucleotide identities (dots) and deletions (dashes) are indicated. The *Aldh-pb* sequence is numbered as −1 from the immediate 5′-end of the initiation codon ATG. The known transcriptional initiation site for human *ALDH1* is underlined [41]. Potential regulatory sequences are indicated, i.e. the CCAAT box, Oct1 and ATA (TATA box).

two overlapping clones (f5 and f16) of a novel mouse gene, *Aldh-pb*, which exhibited the homologous splice positions as contained in exons 3–12 of *Aldh1*. Genomic walking was performed to obtain 5′- and 3′-clones (f5-2 and f11 respectively), and it was determined that *Aldh-pb* spans 35 kb and contains 13 exons (Figure 1B). Southern blot analysis of *Bam*HI- or *Sst*I-treated total genomic DNA with each respective cDNA probe revealed band patterns consistent with the restriction maps of both cloned genes.

The intron/exon boundaries of *Aldh1* and *Aldh-pb* were mapped by sequencing the exons and portions of the adjacent introns, and comparing with the cDNA sequences (see below). All of the intron/exon boundaries within both genes follow the GT/AG rule except for intron 9 of *Aldh1*, which has a GC/AG boundary sequence (Table 1a and 1b). For both genes, eight of the twelve codon interruptions of the exons are in phase O (between codons).

Cloning of full-length cDNAs encoding mouse *Aldh1* and *Aldh-pb*

In order to confirm the predicted amino acid sequences and assignments of intron/exon boundaries, we performed PCR cloning of ALDH1 and ALDH-PB cDNAs using as the template either mouse testis cDNA library DNA or mouse liver total

Figure 2 Nucleotide and deduced amino acid sequences of mouse ALDH-PB cDNA and sequence alignment with mouse ALDH1

The ALDH-PB and ALDH1 nucleotide and deduced amino acid sequences obtained in this laboratory are shown at the top and the bottom of the alignment respectively. Nucleotide 1 is the A of the longest cloned ALDH-PB cDNA. Amino acid residue 1 was assigned to the serine at the second codon downstream of the translation initiator methionine codon. Indicated are nucleotide identities (dots) and deletions (dashes). Amino acid residues in ALDH1 that differ from those of ALDH-PB are shown in parentheses. A stop codon is indicated by an asterisk. The potential poly(A)$^{+}$ signal (AATAAA) is indicated by a single underline.

© 1999 Biochemical Society

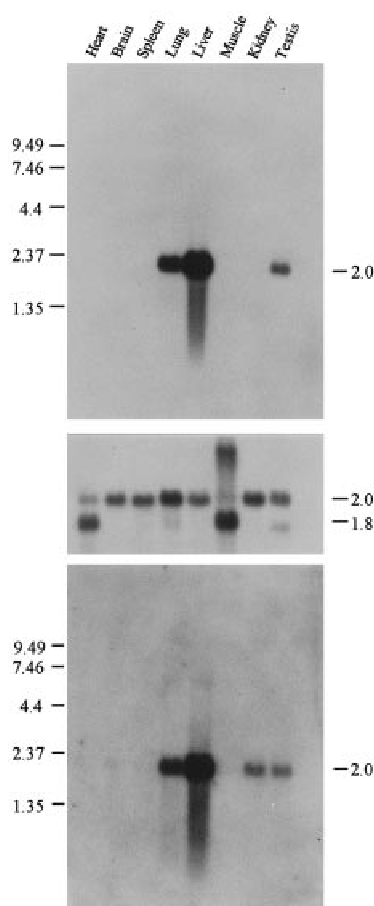

Figure 4 Tissue expression patterns of mouse *Aldh1* and *Aldh-pb*

Shown are Northern blot results from multiple adult mouse tissues with an ALDH1-specific probe (top panel) and an ALDH-PB-specific probe (bottom panel). Indicators of the molecular size of the probes (in kb) are shown on the left and right sides of the panels. A 2.0-kb band that varied in intensity among tissues was seen for each respective probe. A mouse β-actin cDNA probe, which identifies mRNAs of 2.0 and 1.8 kb, was used as a control to confirm the integrity of the mRNA samples (central panel).

RNA. Figure 2 shows the full-length cDNAs for mouse ALDH-PB and ALDH1, both of which are just over 2.0 kb in length. Both cDNA sequences have consensus polyadenylation signals (AATAAA) within 20 bases of the polyadenylated [poly(A)$^{+}$] tail. Our ALDH1 cDNA coding sequence is consistent with a previously published ALDH1 sequence [28], except at seven positions, namely nt 66, 67, 161, 182, 303, 701 and 1418, which result in amino acid discrepancies at residues 10, 86 and 457. In addition, there are three nucleotide discrepancies in the 3′-UTR of the two sequences. For both ALDH1 and ALDH-PB, the nucleotide sequences of the cDNAs isolated here were identical with their corresponding exonic sequences from genomic clones. Further analysis may reveal whether there is any significance in the nucleotide polymorphisms observed in the ALDH1 cDNAs described both previously and in the present study.

The percentages of positional identities between the amino acid sequences deduced from the mouse *Aldh-pb*, mouse *Aldh1*, rat *ALDH-PB* [38], rat *ALDH1* [10,12,31] and human *ALDH1* [39,40] genes are 83 % to 96 % (Table 2). Mouse ALDH1 shares 96 % of its amino acid sequence with rat ALDH1 and 87 % identity with human ALDH1, thus demonstrating that it is the mouse homologue for this gene. The novel ALDH-PB cDNA sequence shows a high percentage of positional identity (90–92 %) in the coding region and 66–73 % identity at the less-restricted 3′-UTR with mouse and rat ALDH1. The sizes of the 5′- and 3′-UTRs of the novel ALDH-PB transcripts are 44 and 510 nt respectively, similar to those of other class I ALDH transcripts. A very high level of overall amino acid sequence identity was noted between mouse ALDH-PB and rat ALDH-PB (92 %), as well as high sequence identity between these two proteins in small pockets (for example, residues 116–125 of ALDH-PB and rat ALDH-PB have an 8-out-of-10 match, but either of these proteins display only a 4-out-of-10 match with the homologous regions of mouse or human ALDH1, and a 5-out-of-10 match with rat ALDH1). From these findings, as well as additional similarities listed below, we conclude that the new gene identified in our cloning attempts, *Aldh-pb*, is the mouse homologue of rat *ALDH-PB*. A human homologue for *ALDH-PB* has not been described.

5′-flanking regions of the *Aldh1* and *Aldh-pb* genes

The nucleotide sequences of the 5′-flanking regions (about 300 nt) for *Aldh1* and *Aldh-pb* were aligned with those of rat *ALDH-PB* [38] and human *ALDH1*, for which the transcription-initiation site has been previously determined [41] (Figure 3). The 5′-flanking region of the *Aldh-pb* gene (between nt −156 and −1 relative to the initiation codon) is closely related to the other rodent genes, showing 81–90 % sequence identity, and has 66 % sequence identity with the human *ALDH1* 5′-flanking region. Between nt −157 and −300, *Aldh-pb* demonstrates essentially no sequence similarity with the other three genes (23–29 % identity), although sequence similarity does exist between the other three genes in this region (50–77 % identity). In each sequence a perfect match to the CCAAT box is conserved, located about 100–120 bp upstream of the initiation codon. Also, an octameric binding motif, Oct1 (ATGCAAAT), is conserved in all except rat *ALDH-PB*, and an ATA motif (TATA box) is well conserved among all except mouse *Aldh-pb*, which shows a 1-bp mismatch in a critical T residue.

Aldh1 and *Aldh-pb* tissue-specific expression patterns

Expression of *Aldh1* and *Aldh-pb* was examined by Northern blot analysis of adult mouse tissues using gene-specific 3′-UTR fragments as probes (Figure 4). A 2.0 kb mRNA that varied in intensity among tissues was seen for both *Aldh1* and *Aldh-pb*. The mRNA band size is compatible with the sizes of the full-length cDNAs obtained for both genes. The highest levels of expression for both mouse genes were found in the liver, lung and testis, but no mRNA was detected for either gene in the heart, spleen, or muscle. ALDH-PB mRNA was also strongly expressed in the kidney and weakly expressed in the brain, but no ALDH1 mRNA was detected in these tissues. This expression pattern for mouse *Aldh-pb* is shared by rat *ALDH-PB*, providing further support for a close relationship between the two genes [38].

Test of whether *Aldh1* or *Aldh-pb* can function in retinoic acid synthesis in *Xenopus* embryos

In order to examine the ability of either ALDH1 or ALDH-PB to contribute to retinoic acid synthesis *in vivo*, *Xenopus* embryos (2–4 cell stage) were micro-injected with *in vitro*-transcribed mRNAs for each gene to overexpress the respective enzyme. After injection, embryos were incubated to the blastula stage (stage 8) and intact embryos were assayed for the presence of retinoic acid using a bioassay in which the embryo is cultured on

© 1999 Biochemical Society

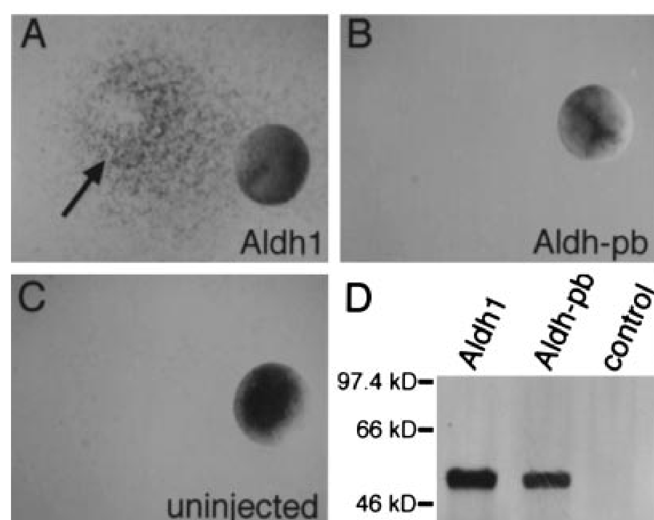

Figure 5 Retinoic acid bioassay following injection of ALDH1 or ALDH-PB mRNAs into *Xenopus* embryos

Xenopus embryos at the 2–4 cell stage were injected with mRNA for either mouse ALDH1 (**A**) or mouse ALDH-PB (**B**) (46 nl of RNA at 0.2 ng/nl in both cases), or were uninjected (**C**). This was followed by incubation to blastula stage 8 and bioassay for retinoic acid by incubation as explants on top of the F9-RARE-*lacZ* reporter cell line, which responds to retinoic acid by inducing *lacZ* expression. Retinoic acid diffusing from the embryo injected with ALDH1 mRNA is shown (by the arrow) to have induced *lacZ* expression surrounding the area where the embryo was incubated; the embryo was moved aside from its site of incubation for photography. No *lacZ* expression is observed for the embryo injected with ALDH-PB mRNA or the uninjected embryo. The quality of the mRNAs was analysed by *in vitro* translation using a rabbit reticulocyte lysate, followed by analysis of the products with SDS gel electrophoresis (**D**); lanes marked ALDH1 and ALDH-PB (0.4 μg of mRNA added in either case) show translation products of about 55 kDa in molecular mass, whereas the control lane (no mRNA added) has no endogenous band at that location.

top of a monolayer of retinoic-acid-reporter cells (F9-RARE-*lacZ*); retinoic acid released from the embryo was subsequently detected in the reporter cells by assaying for *lacZ* expression (β-galactosidase activity) driven by a retinoic acid response element. Embryos injected with mouse ALDH1 mRNA resulted in a strong detection of β-galactosidase activity in the surrounding reporter cells (Figure 5A), whereas embryos injected with mouse ALDH-PB displayed no detectable β-galactosidase activity (Figure 5B). Uninjected embryos at blastula stage 8 also displayed no detectable β-galactosidase activity with this bioassay, indicating that *Xenopus* embryos have no detectable endogenous retinoic acid at this stage (Figure 5C); endogenous retinoic acid was detected by stage 15 (results not shown). ALDH1 and ALDH-PB mRNAs were subjected to *in vitro* translation in a rabbit reticulocyte lysate, and both produced proteins of the correct molecular mass (approx. 55 kDa) with almost equal efficiency (Figure 5D). Overall, we found that injection of comparable amounts of ALDH1 and ALDH-PB mRNAs over a 5-fold concentration range resulted in retinoic acid detection in most ALDH1-injected embryos (35–82 %), but in none of the ALDH-PB-injected embryos or uninjected embryos (Table 3).

DISCUSSION

In this study we have cloned and characterized the cDNAs and genes encoding mouse ALDH1 and ALDH-PB. Upon the basis of gene structure and sequence identity, ALDH1 and ALDH-PB are both class I ALDHs with a close evolutionary relationship. The coding regions of the two genes predict proteins sharing 90 % amino acid sequence identity. The gene structures of both *Aldh1* and *Aldh-pb* consist of 13 exons with 12 introns interrupting their coding regions at exactly the same positions as seen in the human cytosolic class I ALDH gene *ALDH1* [41]. Many of the corresponding introns of the two mouse genes are very similar in size, also indicating a close relationship to each other. Although *Aldh1* and *Aldh-pb* appear to be very closely related genes, significant differences in their expression pattern and function described in the present study provide them with unique identities.

Mouse *Aldh1* and *Aldh-pb* were both expressed at high levels in liver, lung and testis, but *Aldh-pb*, unlike *Aldh1*, was also highly expressed in kidney and demonstrated a low level of expression in the brain. The expression pattern of mouse *Aldh-pb* thus matches that previously reported for rat ALDH-PB, except that the latter is expressed at a low level in liver [38]. Our observation of a high level of expression for both mouse genes in the liver, lung and testis, but no expression for either gene in the heart, spleen or muscle, is supportive of a role in epithelial function, since the former tissues, but not the latter, have large populations of epithelial cells.

A role for mouse ALDH1 in the oxidation of retinal to retinoic acid has previously been proposed by *in vitro* analysis of enzyme activity [8]. This is further supported by studies showing that human and rat ALDH1 homologues also catalyse retinoic acid synthesis *in vitro* [6,11,32]. However, *in vitro* enzyme assays may not reflect physiological conditions, making it unclear whether enzymes identified *in vitro* are playing a significant role in retinoic acid synthesis *in vivo*. We overexpressed mouse ALDH1 and ALDH-PB mRNAs in *Xenopus* embryos to provide an *in vivo* test of the role of these enzymes in retinoic acid synthesis. We found that endogenous retinoic acid was undetectable in *Xenopus*

Table 3 Retinoic acid bioassay

Following injection embryos were incubated until they reached stage 8 (blastula). At that time they were assayed for retinoic acid detection using the F9-RARE-*lacZ* bioassay. The number of embryos (n) used for scoring is shown in parentheses. The bioassay is qualitative, and some variation in the intensity of *lacZ* detection was noted in embryos injected with ALDH1 mRNA that showed a positive response. The lack of response in some ALDH1-injected embryos might be due to embryonic damage during injection. Retinoic acid could not be detected in embryos that were uninjected ($n = 45$).

	% Embryos with retinoic acid detection (n)	
Injection of mRNA (nl) mRNA species added ...	Mouse ALDH1	Mouse ALDH-PB
46	82 (17)	0 (49)
23	77 (61)	0 (60)
13.8	35 (23)	0 (25)
9.2	60 (45)	0 (26)

© 1999 Biochemical Society

embryos at blastula stage 8, as also shown by others [42], providing a large time window during which embryos could be manipulated and examined for effects on retinoic acid synthesis with essentially no background detection. We found that injection of mRNA for mouse *Aldh1*, but not *Aldh-pb*, induced easily detectable retinoic acid synthesis in *Xenopus* embryos using a bioassay that monitors diffusion of retinoic acid from individual embryos to *lacZ* reporter cells. Thus we have provided evidence that ALDH1 can in fact catalyse retinoic acid synthesis *in vivo* using normal endogenous concentrations of substrate (retinal), coenzyme (NAD^+) and other factors that might effect enzyme activity.

The retinoic acid bioassay described here is made possible by the fact that *Xenopus* eggs and embryos contain large amounts of the vitamin A substrate all-*trans*-retinal, the immediate precursor of the major bioactive retinoid all-*trans*-retinoic acid [43,44]. In contrast, all-*trans*-retinol, rather than all-*trans*-retinal, is the abundant form of vitamin A in mammalian embryos [45]. In addition, *Xenopus* embryos contain significant amounts of 4-oxo-retinal, the immediate precursor for the bioactive retinoid 4-oxo-retinoic acid [42]. It is not known if 4-oxo-retinal is a substrate for ALDH1, but the ability of this enzyme to utilize both all-*trans*-retinal and 9-*cis*-retinal [11,46] suggests that its active site might also accommodate 4-oxo-retinal, resulting in metabolism to 4-oxo-retinoic acid. The reporter cells detect the sum of all active carboxylated retinoids, including all-*trans*-retinoic acid, 4-oxo-retinoic acid, 9-*cis*-retinoic acid and didehydro-retinoic acid [47]. Thus it cannot be certain which bioactive retinoids are being produced by ALDH1 in this bioassay, but it is clear that conversion from an aldehyde form into a carboxylic acid form is essential. Our results also indicate that ALDH-PB is inactive in the production of all the carboxylated retinoids mentioned above, which collectively account for the vast majority of bioactive retinoids observed in vertebrate systems.

On the basis of our findings we conclude that mouse ALDH1 can participate in retinoic acid synthesis *in vivo*, but that the closely related ALDH-PB does not possess this property. This is likely to define a major difference in the physiological functions of these two *ALDH* genes. Another indication that ALDH1 functions in retinoic acid synthesis *in vivo* comes from our observation here of a lack of *Aldh1* expression in the adult mouse kidney, combined with our previous report that adult mouse kidney has undetectable retinoic acid using the bioassay described herein [48]. This is further supported by studies in the rat, which has abundant levels of both retinoic acid and ALDH1 in the adult rat kidney [11,49]. Together, these findings suggest a direct link between expression of *ALDH1* and production of retinoic acid in the adult kidney. The high level of *Aldh-pb* expression we observed in the adult mouse kidney, a tissue that lacks retinoic acid, also further strengthens the argument that this gene does not function in retinoic acid synthesis.

Our studies have provided additional information about the promoter regions of class I ALDH genes. A minimal promoter region for human *ALDH1* (nt -133 to -1 in Figure 3) has been defined [50], and this is highly conserved with the rat *ALDH-PB* promoter [38], as well as with the mouse *Aldh1* and *Aldh-pb* 5′-flanking regions described here. Therefore it is very likely that the promoters for the two mouse genes are located in the corresponding regions. Functional analysis of human *ALDH1* in different cell lines [50] indicates that this proximal promoter region directs cell-type-specific expression, and contains two major *cis*-regulatory elements, a CCAAT box conserved in mouse *Aldh1,* mouse *Aldh-pb* and rat *ALDH-PB*, and an Oct1 motif conserved only in *Aldh1* and *Aldh-pb*. It has been suggested that a 16-bp deletion in the rat *ALDH-PB* promoter, which destroys the Oct1 motif, might account for this gene being expressed at low levels in rat liver [51]. In support of this hypothesis, we have shown that mouse *Aldh-pb* has conserved the Oct1 motif, and demonstrates high expression in the liver. Since deletion of the CCAAT box in human *ALDH1* results in a significant decrease in promoter activity [50], it is likely that the CCAAT box plays a significant role in the transcriptional control of all these genes. Unlike the other promoters described above, which have a putative ATA motif (TATA box) located 20–40 bp downstream from the CCAAT box, the *Aldh-pb* promoter has a $T \rightarrow G$ base change at this location, suggesting that the ATA motif might not be a primary regulatory element. This is consistent with studies showing that deletion of the ATA box did not significantly decrease human *ALDH1* promoter activity [50].

In total, 23 conserved amino acid residues have been identified from sequence alignment of multiple ALDH isoenzymes of different classes and phylogenetic origins [52]. The three-dimensional structures of rat cytosolic ALDH3 [53] and human mitochondrial ALDH2 [54] in complex with NAD^+ have shown that the 23 strictly conserved residues are located in the catalytic pocket (Cys^{302}, Glu^{399} and Phe^{401}), at turns in the structure (Gly^{160}, Gly^{186}, Gly^{223}, Gly^{245}, Gly^{270}, Gly^{299}, Gly^{370}, Gly^{402}, Gly^{449}, Gly^{467}, Gly^{474}, Pro^{158}, Pro^{383} and Pro^{403}), or at locations that stabilize adjacent elements of secondary structure (Arg^{84}, Lys^{192}, Thr^{384}, Asn^{421}, Asn^{454} and Ser^{471}). Therefore it was proposed that other classes of ALDH might have similar structural folds, like ALDH2 and ALDH3. Our study shows that these 23 strictly conserved residues are found in both ALDH1 and ALDH-PB. It is thus very likely that these two mouse enzymes, which share 90 % amino acid sequence identity, exhibit similar folding structures and substrate-binding pockets. However, it has been suggested that different exon/intron regions of closely related ALDH genes are under different evolutionary pressures; i.e. human *ALDH7* and *ALDH8* [55]. We now describe a similar observation between mouse *Aldh1* and *Aldh-pb*, in which exons 4 and 12 show 74–80 % positional identity, which is lower than the overall sequence identity of 90 %. The ALDH three-dimensional structures indicate that these two exons encode sequences corresponding to an α-helix located on the surface of the molecule and a turn in the structure respectively. It is possible that these two exons encode protein domains or folds related to the specific functions of individual isoenzymes. Perhaps one or more amino acid differences at these locations can explain our observation that ALDH1, but not ALDH-PB, functions in retinoic acid synthesis.

In summary, our studies define two closely related mouse ALDH genes called *Aldh1* and *Aldh-pb*, demonstrate that mouse ALDH-PB and rat ALDH-PB appear to form a separate branch of class I ALDH that is distinct from human and rodent ALDH1, and show that ALDH1, but not ALDH-PB, functions in retinoic acid synthesis *in vivo*.

We are grateful to M. Wagner for the F9-RARE-*lacZ* reporter cell line, and to A. Yoshida for critical discussions. G. D. acknowledges support from NIH Grant AA09731.

REFERENCES

1 Means, A. L. and Gudas, L. J. (1995) Annu. Rev. Biochem. **64**, 201–233
2 Marshall, H., Morrison, A., Studer, M., Pöpperl, H. and Krumlauf, R. (1996) FASEB J. **10**, 969–978
3 Kastner, P., Mark, M. and Chambon, P. (1995) Cell **83**, 859–869
4 Wilson, J. G., Roth, C. B. and Warkany, J. (1953) Am. J. Anat. **92**, 189–217
5 Duester, G. (1996) Biochemistry **35**, 12221–12227

© 1999 Biochemical Society

6 Dockham, P. A., Lee, M.-O. and Sladek, N. E. (1992) Biochem. Pharmacol. **43**, 2453–2469
7 Yoshida, A., Hsu, L. C. and Dave, V. (1992) Enzyme **46**, 239–244
8 Lee, M.-O., Manthey, C. L. and Sladek, N. E. (1991) Biochem. Pharmacol. **42**, 1279–1285
9 McCaffery, P., Lee, M.-O., Wagner, M. A., Sladek, N. E. and Dräger, U. C. (1992) Development **115**, 371–382
10 Kathmann, E. C. and Lipsky, J. J. (1997) Biochem. Biophys. Res. Commun. **236**, 527–531
11 Labrecque, J., Dumas, F., Lacroix, A. and Bhat, P. V. (1995) Biochem. J. **305**, 681–684
12 Penzes, P., Wang, X. S., Sperkova, Z. and Napoli, J. L. (1997) Gene **191**, 167–172
13 Zhao, D., McCaffery, P., Ivins, K. J., Neve, R. L., Hogan, P., Chin, W. W. and Dräger, U. C. (1996) Eur. J. Biochem. **240**, 15–22
14 Wang, X. S., Penzes, P. and Napoli, J. L. (1996) J. Biol. Chem. **271**, 16288–16293
15 Yoshida, A., Rzhetsky, A., Hsu, L. C. and Chang, C. (1998) Eur. J. Biochem. **251**, 549–557
16 McCaffery, P., Posch, K. C., Napoli, J. L., Gudas, L. and Dräger, U. C. (1993) Dev. Biol. **158**, 390–399
17 Lindahl, R. (1992) Crit. Rev. Biochem. Mol. Biol. **27**, 283–335
18 Ambroziak, W. and Pietruszko, R. (1993) Adv. Exp. Med. Biol. **328**, 5–15
19 Greenfield, N. J. and Pietruszko, R. (1977) Biochim. Biophys. Acta **483**, 35–45
20 Inoue, K., Fukunaga, M. and Yamasawa, K. (1980) Pharmacol. Biochem. Behav. **13**, 295–297
21 Rawles, J. W., Rhodes, D. L., Potter, J. J. and Mezey, E. (1987) Biochem. Pharmacol. **36**, 3715–3722
22 Pereira, F., Rosenmann, E., Nylen, E., Kaufman, M., Pinsky, L. and Wrogemann, K. (1991) Biochem. Biophys. Res. Commun. **175**, 831–838
23 Radin, A. I., Zhoa, X. L., Woo, T. H., Colvin, O. M. and Hilton, J. (1991) Biochem. Pharmacol. **42**, 1933–1939
24 Yoshida, A., Hsu, L. C. and Yanagawa, Y. (1993) Adv. Exp. Med. Biol. **328**, 37–44
25 Banfi, P., Lanzi, C., Falvella, F. S., Gariboldi, M., Gambetta, R. A. and Dragani, T. A. (1994) Mol. Pharmacol. **46**, 896–900
26 Westlund, P., Fylling, A. C., Cederlund, E. and Jörnvall, H. (1994) FEBS Lett. **345**, 99–103
27 Algar, E. M. and Holmes, R. S. (1989) Prog. Clin. Biol. Res. **290**, 93–103
28 Rongnoparut, P. and Weaver, S. (1991) Gene **101**, 261–265
29 Chen, M., Achkar, C. and Gudas, L. J. (1994) Mol. Pharmacol. **46**, 88–96
30 McCaffery, P., Tempst, P., Lara, G. and Dräger, U. C. (1991) Development **112**, 693–702
31 Bhat, P. V., Labrecque, J., Boutin, J. M., Lacroix, A. and Yoshida, A. (1995) Gene **166**, 303–306
32 Posch, K. C., Burns, R. D. and Napoli, J. L. (1992) J. Biol. Chem. **267**, 19676–19682
33 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
34 Nieuwkoop, P. D. and Faber, J. (1994) Normal Table of Xenopus Laevis (Daudin), Garland Publishing, New York
35 Kay, B. K. (1991) in *Xenopus laevis*: Practical Uses in Cell and Molecular Biology, vol. 36 (Kay, B. K. and Peng, H. B., eds.), pp. 663–669, Academic Press, San Diego
36 Wagner, M., Han, B. and Jessell, T. M. (1992) Development **116**, 55–66
37 Ang, H. L., Deltour, L., Hayamizu, T. F., Zgombic-Knight, M. and Duester, G. (1996) J. Biol. Chem. **271**, 9526–9534
38 Dunn, T. J., Koleske, A. J., Lindahl, R. and Pitot, H. C. (1989) J. Biol. Chem. **264**, 13057–13065
39 Hempel, J., von Bahr-Lindstrom, H. and Jornvall, H. (1984) Eur. J. Biochem. **141**, 21–35
40 Hsu, L. C., Tani, K., Fujiyoshi, T., Kurachi, K. and Yoshida, A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 3771–3775
41 Hsu, L. C., Chang, W. C. and Yoshida, A. (1989) Genomics **5**, 857–865
42 Blumberg, B., Bolado, Jr., J., Derguini, F., Craig, A. G., Moreno, T. A., Chakravarti, D., Heyman, R. A., Buck, J. and Evans, R. M. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 4873–4878
43 Plack, P. A. and Kon, S. K. (1961) Biochem. J. **81**, 561–570
44 Creech Kraft, J., Schuh, T., Juchau, M. R. and Kimelman, D. (1994) Biochem. J. **301**, 111–119
45 Horton, C. and Maden, M. (1995) Dev. Dyn. **202**, 312–323
46 El Akawi, Z. and Napoli, J. L. (1994) Biochemistry **33**, 1938–1943
47 Yamamoto, M., Dräger, U. C. and McCaffery, P. (1998) Dev. Brain Res. **107**, 103–111
48 Ang, H. L., Deltour, L., Zgombic-Knight, M., Wagner, M. A. and Duester, G. (1996) Alcohol. Clin. Exp. Res. **20**, 1050–1064
49 Bhat, P. V., Poissant, L., Falardeau, P. and Lacroix, A. (1988) Biochem. Cell Biol. **66**, 735–740
50 Yanagawa, Y., Chen, J. C., Hsu, L. C. and Yoshida, A. (1995) J. Biol. Chem. **270**, 17521–17527
51 Chen, J., Yanagawa, Y. and Yoshida, A. (1996) Biochem. Genet. **34**, 109–116
52 Hempel, J., Nicholas, H. and Lindahl, R. (1993) Prot. Sci. **2**, 1890–1900
53 Liu, Z. J., Sun, Y. J., Rose, J., Chung, Y. J., Hsiao, C. D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. and Wang, B. C. (1997) Nat. Struct. Biol. **4**, 317–326
54 Steinmetz, C. G., Xie, P. G., Weiner, H. and Hurley, T. D. (1997) Structure **5**, 701–711
55 Hsu, L. C., Chang, W. C. and Yoshida, A. (1997) Gene **189**, 89–94

Received 3 December 1998/15 January 1999; accepted 4 February 1999

© 1999 Biochemical Society