APPLIED AND ENVIRONMENTAL MICROBIOLOGY, July 1996, p. 2303–2310 Vol. 62, No. 7
0099-2240/96/$04.00+0

Cloning of the Aldehyde Reductase Gene from a Red Yeast, *Sporobolomyces salmonicolor*, and Characterization of the Gene and Its Product

KEIKO KITA,[1]* KOJI MATSUZAKI,[1] TETSU HASHIMOTO,[1] HIDESHI YANASE,[1] NOBUO KATO,[1]†
MAX CHING-MING CHUNG,[2] MICHIHIKO KATAOKA,[3] AND SAKAYU SHIMIZU[3]

Department of Biotechnology, Tottori University, 4-101 Koyama, Tottori 680,[1] and Department of Agricultural Chemistry, Kyoto University, Sakyo-ku, Kyoto 606,[3] Japan, and Department of Biochemistry and Bioprocessing Technology Center, National University of Singapore, Singapore 0511, Singapore[2]

Received 21 February 1996/Accepted 29 April 1996

An NADPH-dependent aldehyde reductase (ALR) isolated from a red yeast, *Sporobolomyces salmonicolor*, catalyzes the reduction of a variety of carbonyl compounds. To investigate its primary structure, we cloned and sequenced the cDNA coding for ALR. The aldehyde reductase gene (*ALR*) comprises 969 bp and encodes a polypeptide of 35,232 Da. The deduced amino acid sequence showed a high degree of similarity to other members of the aldo-keto reductase superfamily. Analysis of the genomic DNA sequence indicated that the *ALR* gene was interrupted by six introns (two in the 5′ noncoding region and four in the coding region). Southern hybridization analysis of the genomic DNA from *S. salmonicolor* indicated that there was one copy of the gene. The *ALR* gene was expressed in *Escherichia coli* under the control of the *tac* promoter. The enzyme expressed in *E. coli* was purified to homogeneity and showed the same catalytic properties as did the enzyme from *S. salmonicolor*.

Aldehyde reductase (ALR) (EC 1.1.1.2), aldose reductase (EC 1.1.1.21), and carbonyl reductase (EC 1.1.1.184) catalyze the NADPH-dependent reduction of a variety of carbonyl compounds and are widely distributed in mammalian and plant tissues. The amino acid sequences of aldose reductases and ALRs show significant similarity, but that of carbonyl reductase does not show similarity to the other sequences. These enzymes are members of the aldo-keto reductase superfamily (3), which includes prostaglandin F synthase (29), ρ-crystalline (8), δ-3-ketosteroid 5-β-reductase (17), a soybean reductase (30), and chlordecone reductase (34). The yeast GCY gene product (16), the p100/11E gene product of *Leishmania major* (21), and mouse androgen-dependent protein (19), whose functions are not known, also show a high degree of sequence similarity with members of the aldo-keto reductase superfamily. The physiological roles of the aldo-keto reductases have not been established. It is suggested that under physiological conditions aldose reductase participates in osmoregulation, but under hyperglycaemic conditions it contributes to the onset and development of severe complications in diabetes.

We isolated an NADPH-dependent ALR from the red yeast *Sporobolomyces salmonicolor*, which in addition to catalyzing common substrates of ALRs (35) can catalyze the asymmetric reduction of ethyl 4-chloro-3-oxobutanoate to ethyl (*R*)-4-chloro-3-hydroxybutanoate, a promising chiral building block for chemical synthesis of L-carnitine. L-Carnitine deficiency is a common secondary problem associated with many metabolic diseases, and oral administration of the compound produces a drastic improvement in the condition of certain patients. On the basis of the N-terminal amino acid sequence, we deduced that the *Sporobolomyces* ALR belongs to the aldo-keto reductase superfamily (13).

In this study, we cloned and analyzed the cDNA sequence of the ALR gene (*ALR*). The complete primary structure of ALR was compared with those of other proteins in the family. We also describe the structure of the genomic DNA coding for ALR and overexpression of the enzyme in *Escherichia coli.*

MATERIALS AND METHODS

Microorganisms and culture conditions. *S. salmonicolor* (AKU4429) was used as the DNA donor. The organism was cultivated at 30°C in YPG medium containing 5% glucose, 1% peptone, and 1% yeast extract. *E. coli* JM109 [*recA supE endA hsdR gyrA relA thi* Δ(*lac-proAB*) F′ (*traD proAB*$^+$ *lacI*^{q} *lacZ*Δ*M15*)], *E. coli* MV1184 [*ara* Δ(*lac-proAB*) *rpsL thi* (ϕ80 *lacZ*Δ*M15*) Δ(*srl-recA*) *306*::Tn*10*(Tet^r)], and *E. coli* HB101 [*supE44 hsdS20*(r_B^- m_B^-) *recA13 ara-14 proA2 lacY1 galK2 rpsL21 xyl-5 mtl-1*] were used as host cells. *E. coli* cells were cultured at 37°C in Luria-Bertani (LB) medium containing 1% Bacto-tryptone (Difco Laboratories, Detroit, Mich.), 0.5% Bacto-yeast extract (Difco Laboratories), and 1% NaCl, pH 7.0. When necessary, ampicillin (100 μg/ml) and kanamycin (70 μg/ml) were added to the medium.

Enzymes and chemicals. Restriction enzymes were purchased from Takara Shuzo Co., Ltd. (Kyoto, Japan), and Toyobo (Osaka, Japan). *Taq* DNA polymerase was purchased from Takara Shuzo Co., Ltd. Sequenase and chemiluminescent DNA sequencing kits were purchased from Toyobo. [γ-^{32}P]ATP and [α-^{32}P]dCTP were purchased from Amersham Inc.

Amino acid sequencing. ALR was purified from cells of *S. salmonicolor* as described previously (35). The enzyme was digested with lysyl endopeptidase (Wako Pure Chemicals, Osaka, Japan) in 30 mM Tris-HCl, pH 9.0, containing 4 M urea for 6.5 h at 30°C at a substrate/enzyme ratio of 200:1. The peptides obtained on lysyl endopeptidase digestion were separated by high-performance liquid chromatography (HPLC) on a Cosmosil 5C18-AR column (4.6 by 150 mm; Nacalai Tesque, Kyoto, Japan) equilibrated with 0.1% trifluoroacetic acid and eluted with a linear acetonitrile gradient at a flow rate of 1.0 ml/min. Some peptides were further rechromatographed on a Cosmosil 5C8-300 column (4.6 by 150 mm) with 0.1% trifluoroacetic acid and eluted with a linear acetonitrile gradient at a flow rate of 1.0 ml/min. The sequence was determined with an Applied Biosystems model 477A gas–liquid-phase protein sequencer. The phenylthiohydantoin amino acid derivatives were separated and identified with an Applied Biosystems model 120A on-line phenylthiohydantoin analyzer.

* Corresponding author. Mailing address: Department of Biotechnology, Tottori University, 4-101 Koyama, Tottori 680, Japan. Phone: (81-857) 31-5277. Fax: (81-857) 31-0881. Electronic mail address: kita@bio.tottori-u.ac.jp.

† Present address: Department of Agricultural Chemistry, Kyoto University, Sakyo-ku, Kyoto 606, Japan.

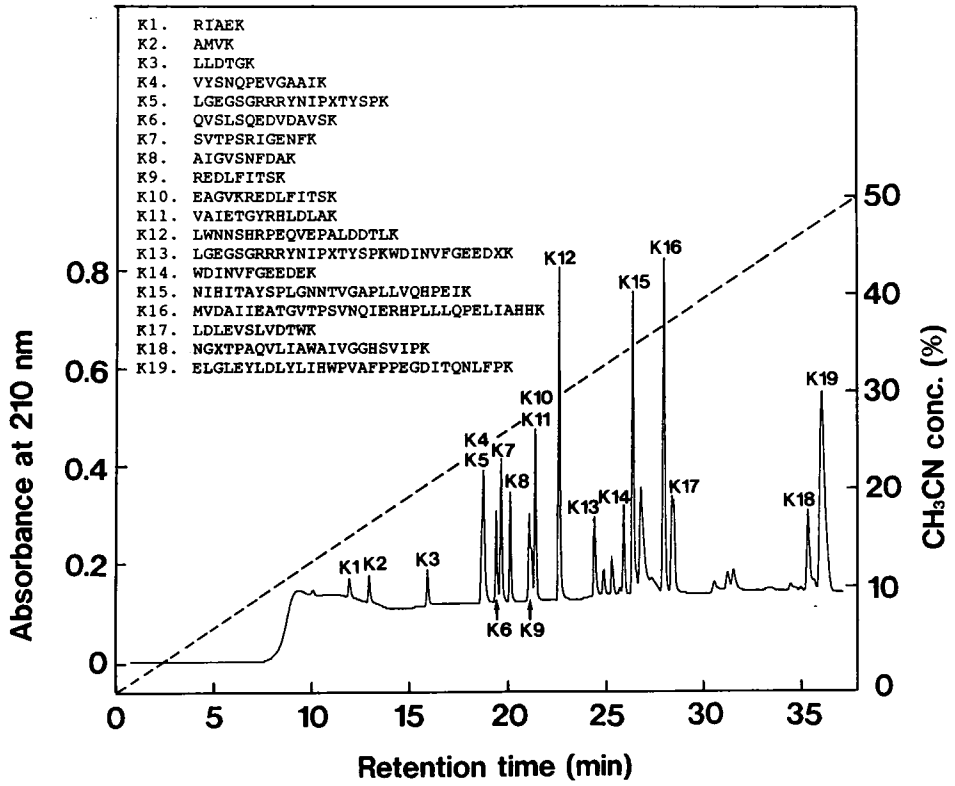

FIG. 1. Separation of lysyl endopeptidase-digested peptides. The reverse-phase HPLC column (Cosmosil 5C18-AR, 4.6 by 150 mm) was equilibrated with 0.1% trifluoroacetic acid and developed with a linear acetonitrile gradient at a flow rate of 1.0 ml/min. The solid and dotted lines show the A_{210} and the acetonitrile concentration, respectively.

Construction of a genomic DNA library in λEMBL3. Cell walls of *S. salmonicolor* were lysed by the method of Shiraishi and Fujii (27). *S. salmonicolor* cultivated in 500 ml of YPG medium was suspended in 20 ml of 50 mM glucose–25 mM Tris-HCl, pH 8.0–10 mM EDTA, and egg white lysozyme was added to a final concentration of 4 mg/ml. After the solution had been incubated at 37°C for 1 h with gentle shaking, the cells were collected by centrifugation and then resuspended in 80 ml of 1.2 M sorbitol–50 mM sodium acetate buffer, pH 5.0–0.1 mM EDTA. Meicelase P-1 (Meiji Seika Co., Ltd., Tokyo, Japan) was added to a final concentration of 18.75 mg/ml, and then the solution was incubated at 37°C for 16 h with gentle shaking. The cells were collected by centrifugation, and the pellet was resuspended in 20 ml of 50 mM Tris-HCl, pH 8.0–1 mM EDTA–1% sodium dodecyl sulfate (SDS) and then incubated at 70°C for 10 min. Next, proteinase K was added to a final concentration of 0.25 mg/ml and the mixture was incubated at 37°C for 2 h. The solution was extracted with phenol-chloroform and then dialyzed against 10 mM Tris-HCl, pH 8.0–1 mM EDTA–0.15 M NaCl. After RNase A treatment (30 μg/ml) at 37°C for 2 h, the solution was extracted with phenol-chloroform and then dialyzed against 10 mM Tris-HCl, pH 8.0–1 mM EDTA and stored at 4°C. Partial *Sau*3AI digests were fractionated by agarose gel electrophoresis. Fractions containing fragments of 9 to 23 kb were recovered from the gel and inserted into *Bam*HI-digested λEMBL3 with T4 DNA ligase. The ligated DNA was packaged into λ phage capsids with an in vitro packaging kit (Gigapack II; Stratagene, La Jolla, Calif.). After adsorption to *E. coli* P2392, the cells were plated onto LB agar plates.

Construction of a cDNA library of *S. salmonicolor*. Total RNA was prepared from *S. salmonicolor* by the method of Sherman et al. (25), and poly(A)$^+$ RNA was purified with a Dynabeads mRNA purification kit (Nihon Dynal, Tokyo, Japan). Oligo(dT)-primed double-stranded cDNA was synthesized with a cDNA synthesis kit (Stratagene), ligated with dephosphorylated λ ZAP II DNA digested with *Eco*RI and *Xho*I, and then packaged in vitro.

Screening of the library. On the basis of the amino acid sequence of ALR, two oligonucleotides [1, 5′-TA(T/C)GG(G/A/T/C)AC(G/A/T/C)TGGCA(G/A)GC-3′; 2, 5′-TGGGA(T/C)AT(A/T/C)AA(T/C)GT(G/A/T/C)TT(T/C)GG-3′] were synthesized as probes to screen the library. Hybridization was performed at 45°C for 16 h in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–10× Denhardt's solution–0.05 mg of denatured herring sperm DNA per ml–^{32}P-labeled oligonucleotide. Washes were carried out at 50°C in 6× SSC–0.1% SDS and then in 2× SSC–0.1% SDS. The genomic DNA library was screened with oligonucleotides 1 and 2, and the cDNA library was screened with oligonucleotide 2.

Subcloning and DNA sequencing. Phage particles were purified with LambdaSorb phage adsorbent (Promega, Madison, Wis.), and DNA was extracted with phenol-chloroform. The subfragments generated were cloned into M13mp18, M13mp19, pUC118, and pUC119 to provide templates, and the sequences were deduced from the data for both strands by the dideoxynucleotide chain termination method (15, 23).

Expression of ALR in *E. coli*. The PCR technique was used to subclone the *ALR* gene. The two synthetic primers used for this protocol were an N-terminal primer (5′-GTTGAATTCGGAGGTATTATATGGTCGGCACTACTACCCT C-3′) containing a Shine-Dalgarno sequence (underlined nucleotides) (26) flanked by an *Eco*RI site and a C-terminal primer (5′-ATTCTGCAGCTACT TGATCTTCACGGCGTTC-3′) flanked by a *Pst*I site. Plasmid pSAL2 was used as the template. The PCR mixture (100 μl) comprised 50 pmol of each primer, 200 μM each deoxynucleoside triphosphate, 10 mM Tris-HCl (pH 8.3) (at 25°C), 1.5 mM $MgCl_2$, 0.01% (wt/vol) gelatin, 1 ng of template DNA, and 2 U of *Taq* DNA polymerase. The reaction mixture was overlaid with mineral oil, and the reaction was carried out by using a Perkin-Elmer Cetus thermal cycler. Initial template denaturation was programmed for 5 min at 94°C. The profile (1 min at 61°C, 2 min at 72°C, and 30 s at 94°C) was repeated for 35 cycles, and the final 72°C extension step was increased to 5 min. The PCR-generated DNA fragment was ligated into pUC118 cleaved with *Eco*RI and *Pst*I and then transformed into *E. coli* JM109. After ampicillin selection, several clones were picked up and plasmid DNA was examined by restriction analysis. The *ALR* gene was digested with *Eco*RI and *Pst*I and then subcloned into pKK223-3 (7).

Purification of ALR expressed in *E. coli*. *E. coli* HB101(pKAR) cells were grown at 37°C in 5 liters of LB medium containing ampicillin for 12 h. The cells (20 g [wet weight]) were harvested by centrifugation and then resuspended in buffer A comprising 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 10 mM 2-mercaptoethanol. The cells were disrupted by sonication, and the debris was removed by centrifugation at 12,000 × *g* for 30 min at 4°C. The supernatant was loaded onto a 200-ml DEAE-Sephacel (Pharmacia Biotech Inc., Piscataway, N.J.) column equilibrated with buffer A containing 10 mM Tris-HCl, pH 7.4, and 0.1 mM dithiothreitol. The protein was eluted with a total 1,600-ml gradient of 0 to 0.4 M NaCl in buffer A. The ALR was purified to homogeneity by this procedure.

Analytical method. The activity of ALR and the optical purity of ethyl 4-chloro-3-hydroxybutanoate were determined as described elsewhere (13). The reductase activity was measured by using *p*-nitrobenzaldehyde as the substrate. Protein was measured by means of a Bio-Rad protein assay (Japan Bio-Rad

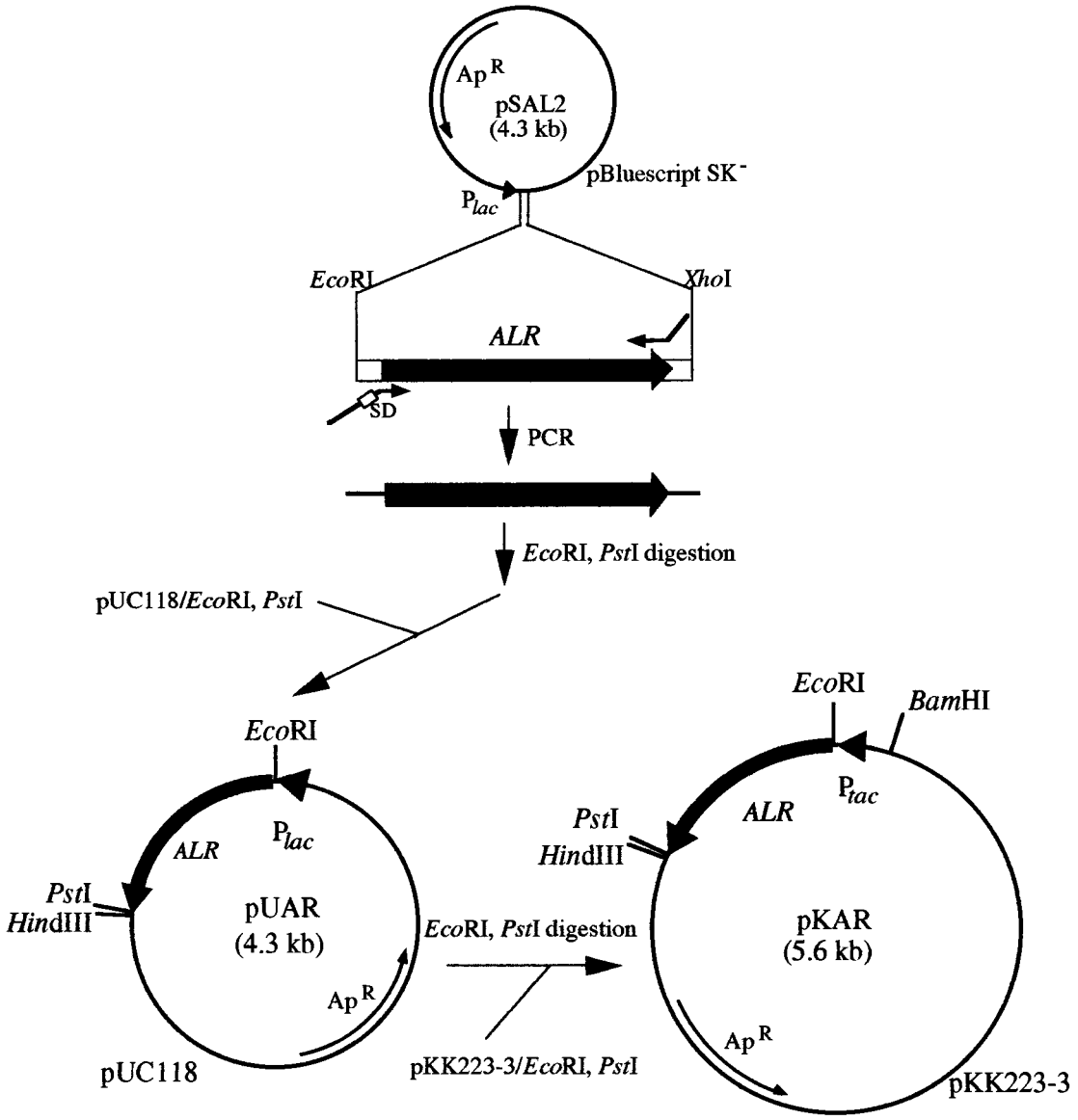

FIG. 2. Structure of plasmid pKAR. pUAR was constructed by insertion of the 1-kb *Pst*I-*Eco*RI fragment generated by PCR into the *Pst*I-*Eco*RI site of pUC118. The *Pst*I-*Eco*RI fragment was cut out from pUAR and then ligated into pKK223-3 digested with *Pst*I and *Eco*RI to obtain pKAR. The translated regions are indicated by filled boxes, and untranslated regions are indicated by open boxes. *ALR*, ALR gene; Ap^R, β-lactamase gene; SD, Shine-Dalgarno sequence; P_{lac}, *lac* promoter; P_{tac}, *tac* promoter.

Laboratories, Tokyo, Japan) with bovine serum albumin as the standard. SDS-polyacrylamide gel electrophoresis (PAGE) was performed by the method of Laemmli (14). The purified enzymes were analyzed by SDS-PAGE and then transferred to a nitrocellulose membrane by the standard procedure (22). The Western blots (immunoblots) were probed first with a purified antibody and then with anti-rabbit immunoglobulin G coupled to alkaline phosphatase. Blocking and probing with antibodies were performed as described by Sambrook et al. (22). Isoelectric focusing was carried out by using precast isoelectric focusing gels (pH 3.5 to 9.5), according to the manufacturer's (Pharmacia Biotech Inc.) instructions. The standard methods described by Sambrook et al. (22) were used for DNA manipulations.

Nucleotide sequence accession number. The sequence of *ALR* is available from GenBank under accession number U26463.

RESULTS AND DISCUSSION

Partial amino acid sequence. ALR purified from cells of *S. salmonicolor* (35) was digested with lysyl endopeptidase. The 19 peptides (K1 to K19) obtained were separated by reverse-

TABLE 1. Intron-processing signals of the *ALR* gene from *S. salmonicolor*

Species and/or *ALR* intron	5′ Splice site	Internal conserved sequence	3′ Splice site
S. salmonicolor			
Intron 1	GTACGT	GTCAC	CAG
Intron 2	GTGAGT	CTGAC	CAG
Intron 3	GTACGC	CTGAT	CAG
Intron 4	GTTCGT	CTGAC	CAG
Intron 5	GTGCGT	CTCAT	CAG
Intron 6	GTAAGC	CTCAT	CAG
Higher eukaryotes[a]	GTAAGT	$CT^G_AA^T_C$	TAG

[a] The *S. salmonicolor* splicing signals are shown with the corresponding consensus sequence in higher eukaryotes (1).

```
GAGCTCGAAG CGGCTCGAGA AGGACAATGG GGAAGCGGTG AAGACGGAGG AAGCAGAACG    60
ATCAGGTACT CGCATATCTT GATGTCGTAG ACGGTGCTCC ACGGGTGTCG GGTTACTAAG   120
GGGAAGCGAA GGAGGCAGCT CTACATACTA GGTTGACGAA GGGGATCAGG CCTCGCGCAT   180
GAGGCGACCG TCTGGGTGAC AGCTCTTGTC GCTTTCGATC TTAGCGTTTT TCTTTCCGTG   240
CTTCTCAGCA GTGGAAAATT GGCAGTGAAC GGGTAGAAAG ACCGGCCACG TCAGCCCAGA   300
ATAGAGCGTA GCTCTGCCGT GGGCGGCGCC GTCATCGATG GTCTAACCCT CGCGGCCAGC   360
TCCGCCCTCG CGGACCGCTT GTTCCTCTTC CTTCTTCACT CTTTACTCCA CCCCCCCGCC   420
                                                                 Δ
AACAACCCCT GCACGCATgt acgtcctctg ccctcgtggt cttctctttc ttcccttgcg   480
gtcttccccc cgagctcacg gtcgtcgttc ttctcctcgc cggaccccca ctcctccacc   540
                                    intron 1
tcgcgttccc cgtcacgaca ttccgattcc ccctcgctct cagCACACAT gtgagttgag   600
cctctctcat ctctctcgaa agctcatctt ccgccgtgca agctgtgctc atcttccgtc   660
                                    intron 2
tcttctgacc ccaaccgcct cttccctcgg gcttctcagC ATGGTCGGCACTACTACCCTC   721
                                          (M) V  G  T  T  T  L      7

AACACTGGCGCTTCCCTCGAGCTCGTCGGCTACGGCACGTGGCAGgtacgtcgatactcctttcc   786
N  T  G  A  S  L  E  L  V  G  Y  G  T  W  Q                       22

tctcctggtttgaccgcgctgatcgaacgggcacttcagGCAGCACCGGGCGAGGTGGGCCAGGG   851
      intron 3                          A  A  P  G  E  V  G  Q  G     31

CGTCAAGGTCGCCATCGAGACTGGATACCGTCACCTCGACCTTGCCAAGgttcgctccctctgac   916
 V  K  V  A  I  E  T  G  Y  R  H  L  D  L  A  K                     47

caccgcacgtatctgggagtcaactgactcgaccctactcccgcagGTCTACTCGAACCAACCTG   981
      intron 4                                    V  Y  S  N  Q  P  E   54

AGGTTGGTGCCGCCATCAAGGAGGCTGGCGTCAAGCGCGAGGACCTCTTCATCACCTCGAAGCTC  1046
  V  G  A  A  I  K  E  A  G  V  K  R  E  D  L  F  I  T  S  K  L      75

TGGAACAACTCGCACCGCCCGGAGCAGGTCGAGCCTGCCCTTGACGACACCCTCAAGGAGCTCGG  1111
W  N  N  S  H  R  P  E  Q  V  E  P  A  L  D  D  T  L  K  E  L  G     97

CCTCGAGTACCTCGACCTTTACCTCATTCACTGGCCCGTCGCGTTCCCGCCCGAGGGCGACATCA  1176
 L  E  Y  L  D  L  Y  L  I  H  W  P  V  A  F  P  P  E  G  D  I  T    119

CCCAGAACCTCTTCCCGAAGGCCAACGACAAGGAGGTCAAGCTCGACCTGGAGGTCAGCCTCGTC  1241
  Q  N  L  F  P  K  A  N  D  K  E  V  K  L  D  L  E  V  S  L  V     140

GACACGTGGAAGGCGATGGTCAAGCTTCTCGACACTGGCAAGGTCAAGGCGATCGGCGTTTCCAA  1306
D  T  W  K  A  M  V  K  L  L  D  T  G  K  V  K  A  I  G  V  S  N    162

CTTCGACGCGAAGATGGTCGACGCCATCATCGAGGCTACCGGCGTGACCCCCTCCGTCAACCAGA  1371
 F  D  A  K  M  V  D  A  I  I  E  A  T  G  V  T  P  S  V  N  Q  I    184

TCGAGCGTCACCCTCTCCTTCTCCAGCCCGAGCTCATCGCCCACCACAAGGCCAAGAACATgtgc  1436
  E  R  H  P  L  L  L  Q  P  E  L  I  A  H  H  K  A  K  N  I        204

gttatcgctcgccattcgtttcgagctgtccaccgctgttccgactgaccttctcatcgtcttcc  1501
                          intron 5

ccatgggcttcagTCACATTACCGCATACTCTCCTCTCGGTAACAACACCGTCGGCGCGCCTCTT  1566
              H  I  T  A  Y  S  P  L  G  N  N  T  V  G  A  P  L      221

CTTGTCCAGCACCCGGAGATCAAGCGCATCGCCGAGAAGAACGGCTGCACGCCCGCTCAGGTgta  1631
L  V  Q  H  P  E  I  K  R  I  A  E  K  N  G  C  T  P  A  Q  V       242

agctctcttcaccgtctcccttccttccaattgcgcttctcgttctctgtgaattcgagctcatg  1696
                          intron 6

ccttgcggtctatgcagCCTCATTGCCTGGGCCATCGTTGGCGGCCACTCGGTTATCCCCAAGTC  1761
                  L  I  A  W  A  I  V  G  G  H  S  V  I  P  K  S    258

GGTCACCCCCTCCCGCATTGGCGAGAACTTCAAGCAGGTCTCGCTCTCGCAGGAGGACGTCGATG  1826
 V  T  P  S  R  I  G  E  N  F  K  Q  V  S  L  S  Q  E  D  V  D  A   280

CCGTCAGCAAGCTCGGCGAGGGTTCGGGCCGCAGGCGCTACAACATCCCCTGCACGTACTCGCCC  1891
  V  S  K  L  G  E  G  S  G  R  R  R  Y  N  I  P  C  T  Y  S  P     301

AAGTGGGACATCAACGTCTTTGGCGAGGAGGACGAGAAGTCGTGCAAGAACGCCGTGAAGATCAA  1956
K  W  D  I  N  V  F  G  E  E  D  E  K  S  C  K  N  A  V  K  I  K    323

GTAGAAGGTCT CGGGTGCCCG GGTACTCTGT TAAAAGCAAT CCCAATCAGT TCGGGGCGT   2016
 *

CCTTTCAAAG AGCTGCACCC TAGCGCTTGT GAGGAAAGCA TGCGGCGAGA AGGCCCAGCA  2076
GCTTACGGTG TGTTGGCCTC AATGCGTAGT CACTTGAAAA AGAAAGGGTT TTCTTCAAGC  2136
CTGGGCCACC TTGACAACCC CTCAATGCGC TAAATTGACG CTGGAGCGGA GCTTAGACGA  2196
GCCGCGGTTC GGCACGCTCC GAGGAGCACC AACTTGCGGG GGGGTTAATC TCCCGTCTTC  2256
TCTTCGCCTG TCGCTCTGAG CTTGCGTTGC TGCGTACGGG GTCACATCTT CAG         2309
```

FIG. 3. DNA and deduced amino acid sequences of the ALR. The exon sequences are shown in uppercase letters. The intron sequences are shown in lowercase letters and are labeled 1 to 6. Amino acids matching the amino acids identified on N-terminal sequence analysis (13) and internal peptide sequence analysis of the ALR are in boldface. The position of the polyadenylation site is shown by double underlining. The triangle indicates the 5′ end of pSAL2. A pair of arrows shows the inverted repeat.

phase HPLC, and the amino acid sequences of these peptides were analyzed (Fig. 1).

Isolation of a genomic *ALR* DNA clone. Total DNA was isolated from *S. salmonicolor* cells and partially digested. DNA fragments of predominantly 9 to 23 kb were inserted into λEMBL3. To screen for the *ALR* gene, two oligonucleotide probes based on the partial amino acid sequence were synthesized. The library (70,000 plaques) was screened, and seven positive clones, to which both oligonucleotides hybridized, were obtained. DNA extracted from the phages was digested with restriction enzymes, and Southern hybridization analysis was carried out with the two oligonucleotide probes. The most intense signals were obtained with clone SAL-6, whose structure was analyzed further.

A 4.4-kb *Bgl*II fragment was subcloned from SAL-6 into pUC119, and the nucleotide sequence of the leftmost 2.3-kb region was determined. Open reading frames in this sequence could potentially encode portions of peptide derived from direct amino acid sequencing of the protein. However, some peptide sequences were interrupted by an inserted unrelated DNA sequence. These results strongly suggested the presence of introns in the *ALR* gene.

Isolation of an ALR-encoding cDNA and nucleotide sequence analysis. The cDNA library constructed from *S. salmonicolor* mRNA was screened by using the same oligonucleotide probes used for screening of the genomic DNA library. Two positive clones were isolated from among 20,000 plaques. The nucleotide sequence was determined for the 1.0-kb insert in one of the positive clones, pSAL2 (Fig. 2), and was compared with the genomic DNA sequence. Figure 3 shows the nucleotide sequence of the 2.3-kb genomic DNA encoding ALR and the deduced amino acid sequence of ALR. The cDNA sequence revealed one open reading frame (969 bp, 323 amino acids) which corresponded to the amino acid sequence of ALR, and the deduced amino acid sequence from the cDNA was identical to the partial amino acid sequence determined by peptide sequencing (Fig. 1). The predicted molecular mass (35,232 Da, excluding the initial methionine) was nearly identical to the molecular mass (35,000 Da) estimated by SDS-PAGE (35).

The genomic DNA sequence covering the region encoding ALR was 1,260 bp long and was interrupted by four introns. This is the first report of the presence of introns in genomic DNA of the red yeast *S. salmonicolor*. The cDNA sequence contained a poly(A) tail added after a G residue located 55 bp downstream of the translational stop codon. However, a potential polyadenylation signal (AATAAA) was not observed within this region. An inverted repeat was present 109 bp downstream of the polyadenylation site. Comparison of the genomic and cDNA sequences indicated that the ALR message started at a point at least 283 bp upstream of the translation start codon and that two introns existed in the 5′ nontranslated region (Fig. 4). Eukaryotic promoter-like elements such as a TATA box and a CAAT sequence were not apparent upstream of the mRNA start site.

The nucleotide sequences of the six introns in the *ALR* gene of *S. salmonicolor* were compared with those reported for higher eukaryotes. Table 1 shows that the splicing signals of *S. salmonicolor* at the 5′ splice site, branch point, and 3′ splice site are conserved relative to each other and to those of higher eukaryotes (18). The spacing between the branch points and the 3′ splice sites in the two 5′ noncoding region introns was 27 to 30 nucleotides, while that in the structural-gene introns was 16 to 21 nucleotides. Among the aldo-keto reductases analyzed so far, both the cDNA and genomic DNA sequences have been analyzed only for human aldose reductase (3, 11). The structural gene for human aldose reductase is interrupted by nine introns. The intron positions were compared between two proteins, but no similarity was found except for the first introns (in both *S. salmonicolor* and the human gene) and between the fourth (*S. salmonicolor*) and seventh (human) ones.

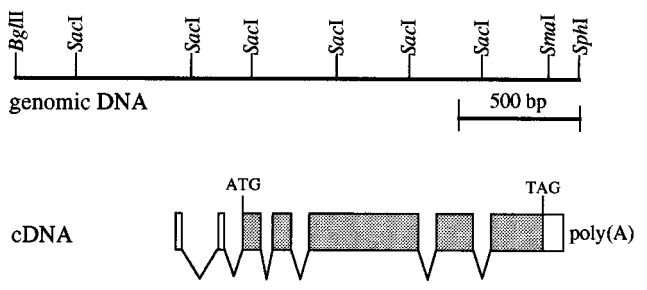

FIG. 4. Exon-intron structure of the *ALR* gene. A restriction map of the genomic region containing the ALR locus along with the structure of the spliced cDNA is presented. The translated regions are indicated by shaded boxes, and the untranslated regions are indicated by open boxes. The start and stop codons are indicated, as well as the polyadenylation site.

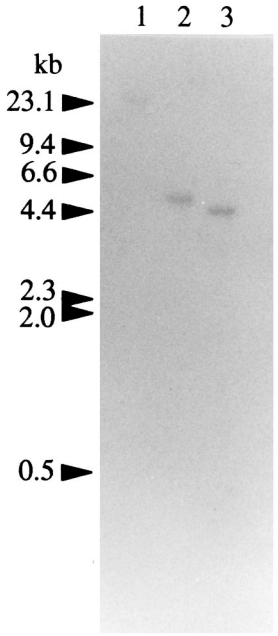

FIG. 5. Southern hybridization analysis of genomic DNA from *S. salmonicolor*. DNA was digested with *Bam*HI (lane 1), *Bgl*II (lane 2), or *Sma*I (lane 3) and then fractionated on a 0.7% agarose gel. The DNA was transferred to Hybond-N^+ membrane and probed with oligonucleotide 2.

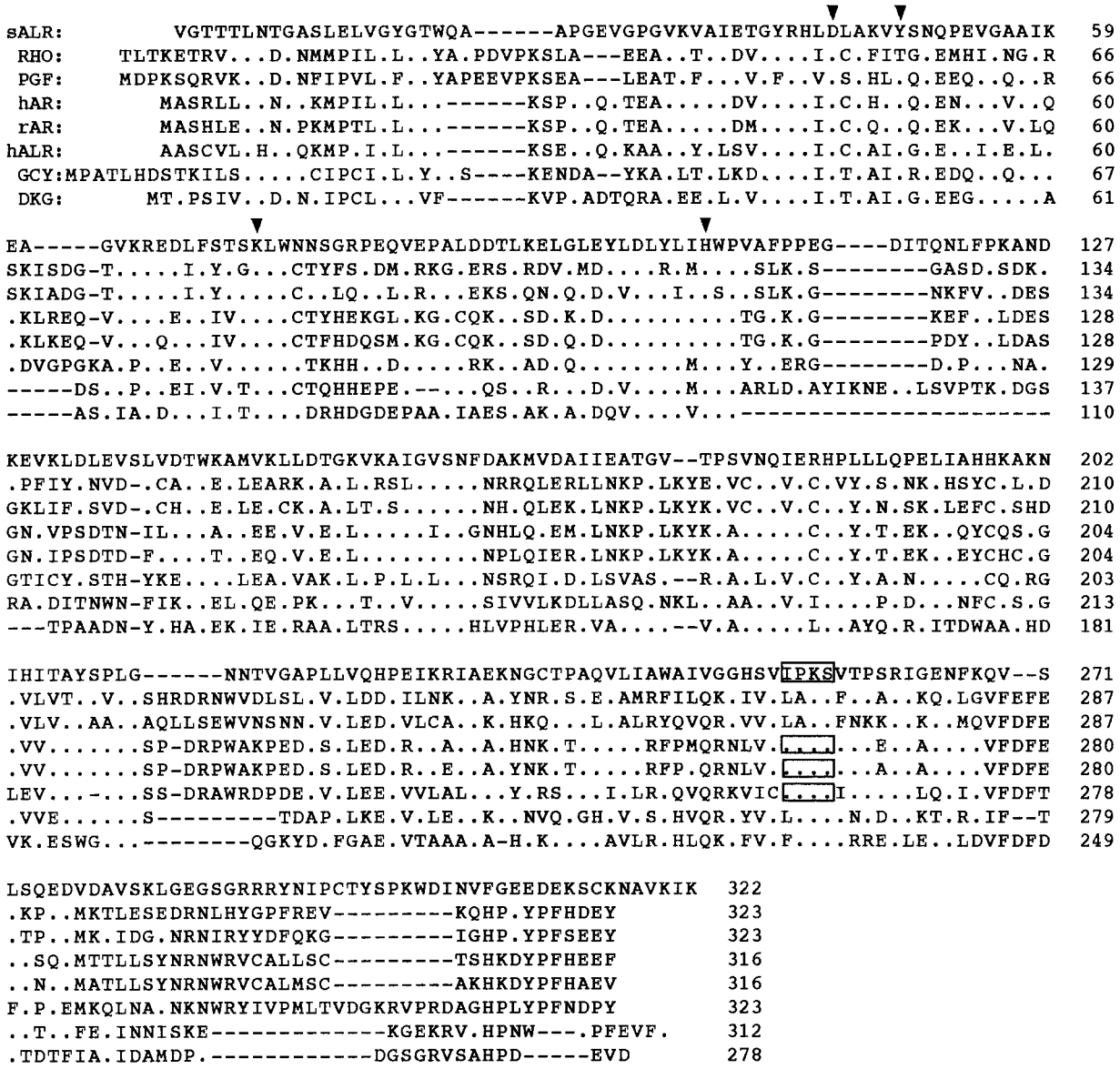

FIG. 6. Comparison of the deduced amino acid sequences of aldo-keto reductases. The sequences of ALR from *S. salmonicolor* (sALR), ρ-crystallin (RHO) (9), prostaglandin F synthase (PGF) (29), human aldose reductase (hAR) (3), rat aldose reductase (rAR) (8), human ALR (hALR) (3), the yeast nuclear gene product (GCY) (16), and 2,5-diketogluconic acid reductase from a *Corynebacterium* sp. (DKG) (12) are aligned. Dots represent amino acids identical to those in the ALR from *S. salmonicolor*. Gaps in the aligned sequences are also indicated (−). The putative active-site residues are indicated by black triangles. The predicted NADPH-binding motifs, IPKS, are boxed.

A blot of *S. salmonicolor* genomic DNA digested with restriction enzymes showed the same pattern when probed with oligonucleotide 2 used for selection of the genomic clone (Fig. 5), indicating that the *S. salmonicolor* genome contains a single *ALR* gene.

Comparison of the ALR amino acid sequence with those of other aldo-keto reductases. The deduced amino acid sequence of ALR was compared with other protein sequences in the GenBank database. A high level of identity was found for other proteins belonging to the aldo-keto reductase superfamily (3). Identities between *S. salmonicolor* ALR and the bovine (24), rat (8), human (3), rabbit (10), and barley (2) aldose reductases; human ALR (3); and xylose reductase from *Pichia stipitis* (1) were 45.1, 45.0, 43.7, 42.9, 39.5, 43.9, and 37.7%, respectively. In general, the aldo-keto reductases have greater similarity at their N-terminal regions than at their C-terminal regions (Fig. 6).

Structural models of human aldose reductase suggested that Tyr-49 acts as the acid base catalyst and Asp-44, Lys-79, and His-112 play an important role in facilitating the hydride transfer (4, 28). All four of these amino acid residues are conserved throughout the superfamily, including the ALR from *S. salmonicolor.*

These enzymes have a strict requirement for NADPH. The tetra-amino acid motif IPKS is conserved among these NADPH-dependent reductases, and the lysine residue in this motif is involved in NADPH binding (5). Although the motif is present in *S. salmonicolor* ALR (boxed in Fig. 6), some amino acid residues were not conserved around the motif. The crystal structures of aldose reductases complexed with NADPH or an

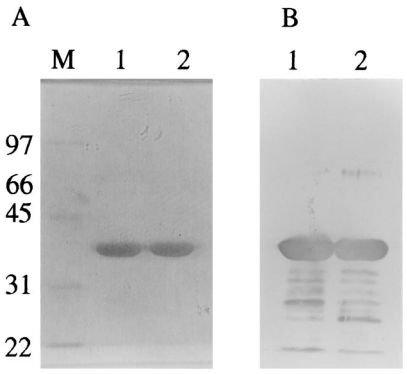

FIG. 7. Western blotting analysis of ALR purified from *E. coli*. ALRs purified from *S. salmonicolor* (lane 1) and *E. coli* (lane 2) were separated by 0.1% SDS–10% PAGE (A) and analyzed by Western blotting (B) with antibodies raised against ALR from *S. salmonicolor*. Molecular mass standards (M) are indicated in kilodaltons.

analog of the cofactor show that the enzyme undergoes a large conformational change upon binding to NADPH. The change involves the reorientation of loop 7 to a position which appears to lock the coenzyme into place (6, 20, 33). The putative loop 7 in the ALR from *S. salmonicolor* is shorter than that in other aldo-keto reductases. The K_m values for NADPH were estimated to be 2 to 4 and 37.5 μM for mammalian reductases (31, 32) and *S. salmonicolor* reductase (13), respectively. The difference in the K_m values could be due to amino acid sequence differences around the NADPH-binding region.

Construction of an ALR-overproducing strain. An *E. coli* strain carrying ALR cDNA, pSAL2, showed reductase activity. As neither a putative promoter sequence nor a ribosome-binding site precedes the *ALR* gene, the protein would be expressed as a fusion protein with β-galactosidase derived from the vector. The *ALR* gene was introduced into high-level expression vector pKK223-3, and the recombinant plasmid, pKAR (Fig. 2), was transformed into *E. coli* JM109 cells. The specific activity of ALR toward *p*-nitrobenzaldehyde in a cell extract of *E. coli* JM109(pKAR) was 19.69 U/mg of protein, that is, 10 and 5 times higher than those in *E. coli* JM109 (pSAL2) and *S. salmonicolor* (35), respectively.

Analysis of ALR purified from *E. coli*. The recombinant ALR was purified 5.8-fold from *E. coli* HB101(pKAR) cells, with a yield of 72.6%. Cellular content of the enzyme was more than 17% of the total extractable protein. The purified enzyme gave a single band on SDS-PAGE, and Western blotting analysis showed that the purified recombinant enzyme reacted with the ALR antibody in a manner similar to that of ALR from *S. salmonicolor* (Fig. 7). The specific activity (104.5 U/mg), K_m value (0.64 mM), and V_{max} value (370 μmol/min/mg) of the recombinant ALR toward *p*-nitrobenzaldehyde were the same as those for the ALR purified from *S. salmonicolor* (13, 35). The isoelectric point of the recombinant enzyme was 4.7, which was the same as that of the yeast enzyme (35). The sequence of the first 10 amino acids of the purified enzyme was determined by protein sequencing; the determined sequence was the same as that of the enzyme purified from *S. salmonicolor* (13). The substrate specificity for typical aldehydes and stereoselectivity for ethyl 4-chloro-3-oxobutanoate of the enzyme from *E. coli* were the same as those of ALR from *S. salmonicolor*. Since *E. coli* does not contain a detectable amount of reducing activity toward ethyl 4-chloro-3-oxobutanoate, we think that this system could produce large amounts of ALR. This process with a recombinant strain of *E. coli* could constitute an economical means for the production of ethyl (*R*)-4-chloro-3-hydroxybutanoate.

ACKNOWLEDGMENTS

We thank T. Morikawa, Research Center, Denki Kagaku Kogyo, for his help in the amino acid sequencing analysis.

This work was supported in part by a grant-in-aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. **Amore, R., P. Kötter, C. Kuster, M. Ciriacy, and C. P. Hollenberg.** 1991. Cloning and expression in *Saccharomyces cerevisiae* of the NAD(P)H-dependent xylose reductase-encoding gene (*XYL1*) from the xylose-assimilating yeast *Pichia stipitis*. Gene **109:**89–97.
2. **Bartels, D., K. Engelhardt, R. Roncarati, K. Schneider, M. Rotter, and F. Salamani.** 1991. An ABA and GA modulated gene expressed in the barley embryo encodes an aldose reductase related protein. EMBO J. **10:**1037–1043.
3. **Bohren, K. M., B. Bullock, B. Wermuth, and K. H. Gabbay.** 1989. The aldo-keto reductase superfamily. J. Biol. Chem. **264:**9547–9551.
4. **Bohren, K. M., C. E. Grimshaw, C.-J. Lai, D. H. Harrison, D. G. Ringe, A. Petsko, and K. H. Gabbay.** 1994. Tyrosine-48 is the proton donor and histidine-110 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. Biochemistry **33:**2021–2032.
5. **Bohren, K. M., J. L. Page, R. Shankar, S. P. Henry, and K. H. Gabbay.** 1991. Expression of human aldose and aldehyde reductases. J. Biol. Chem. **266:** 24031–24037.
6. **Borhani, D. W., T. M. Harter, and J. M. Petrash.** 1992. The crystal structure of the aldose reductase · NADPH binary complex. J. Biol. Chem. **267:**24841–24847.
7. **Brosius, J., and A. Holy.** 1984. Regulation of ribosomal RNA promoter with a synthetic *lac* operator. Proc. Natl. Acad. Sci. USA **81:**6929–6933.
8. **Carper, D., C. Nishimura, T. Shinohara, B. Dietzchold, G. Wistow, C. Craft, P. Kador, and J. H. Kinoshita.** 1987. Aldose reductase and ρ-crystallin belong to the same protein superfamily as aldehyde reductase. FEBS Lett. **220:**209–213.
9. **Fujii, Y., K. Watanabe, H. Hayashi, Y. Urade, S. Kuramitsu, H. Kagamiyama, and O. Hayaishi.** 1990. Purification and characterization of ρ-crystallin from Japanese common bullfrog lens. J. Biol. Chem. **265:**9914–9923.
10. **Garcia-Perez, A., B. Martin, H. R. Murphy, S. Uchida, H. Murer, B. D. Cowley, Jr., J. S. Handler, and M. B. Burg.** 1989. Molecular cloning of cDNA coding for kidney aldose reductase. J. Biol. Chem. **264:**16815–16821.
11. **Graham, A., L. Brown, P. J. Hedge, A. J. Gammack, and A. F. Markham.** 1991. Structure of the human aldose reductase gene. J. Biol. Chem. **266:** 6872–6877.
12. **Grindley, J. F., M. A. Payton, H. Van De Pol, and K. G. Hardy.** 1988. Conversion of glucose to 2-keto-L-gulonate, an intermediate in L-ascorbate synthesis, by a recombinant strain of *Erwinia citreus*. Appl. Environ. Microbiol. **54:**1770–1775.
13. **Kataoka, M., H. Sakai, T. Morikawa, M. Katoh, T. Miyoshi, S. Shimizu, and H. Yamada.** 1992. Characterization of aldehyde reductase of *Sporobolomyces salmonicolor*. Biochim. Biophys. Acta **1122:**57–62.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
15. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19:**269–276.
16. **Oechsner, U., V. Magdolen, and W. Bandlow.** 1988. A nuclear yeast gene (GCY) encodes a polypeptide with high homology to a vertebrate eye lens protein. FEBS Lett. **238:**123–128.
17. **Onishi, Y., M. Noshiro, T. Shimosato, and K. Okuda.** 1991. Molecular cloning and sequence analysis of cDNA encoding delta(4)-3-ketosteroid 5-beta-reductase of rat liver. FEBS Lett. **283:**215–218.
18. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55:**1119–1150.
19. **Pailhoux, E. A., A. Martinez, G. M. Veyssiere, and C. G. Jean.** 1990. Androgen-dependent protein from mouse vas deferens. J. Biol. Chem. **265:**19932–19936.
20. **Rondeau, J. M., F. Tete-Favier, A. Podjarny, J. M. Reymann, P. Barth, J. F. Biellmann, and D. Moras.** 1992. Novel NADPH-binding domain revealed by the crystal structure of aldose reductase. Nature (London) **355:**469–472.
21. **Samaras, N., and T. W. Spithill.** 1989. The developmentally regulated p100/11E gene of *Leishmania major* shows homology to a superfamily of reductase genes. J. Biol. Chem. **264:**4251–4254.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
24. **Schade, S. Z., S. L. Early, T. R. Williams, F. J. Kezdy, R. L. Heinrikson, C. E. Grimshaw, and C. C. Doughty.** 1990. Sequence analysis of bovine lens aldose reductase. J. Biol. Chem. **265:**3628–3635.
25. **Sherman, F., G. R. Fink, and J. B. Hicks.** 1986. Laboratory course manual for methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
27. **Shiraishi, A., and H. Fujii.** 1978. *Sporobolomyces* cell wall lytic enzyme produced by a bacterium, no. 98. Nippon Nogeikagaku Kaishi **52:**553–559. (In Japanese.)
28. **Tarle, I., D. Borhani, W. D. K. Wilson, F. A. Quiocho, and J. M. Petrash.** 1993. Probing the active site of human aldose reductase. J. Biol. Chem. **268:**25687–25693.
29. **Watanabe, K., Y. Fujii, K. Nakayama, H. Ohkubo, S. Kuramitsu, H. Kagamiyama, S. Nakanishi, and O. Hayaishi.** 1988. Structural similarity of bovine lung prostaglandin F synthase to lens ε-crystallin of the European common frog. Proc. Natl. Acad. Sci. USA **85:**11–15.
30. **Welle, R., G. Schröder, E. Schiltz, H. Grisebach, and J. Schröder.** 1991. Induced plant responses to pathogen attack. Eur. J. Biochem. **196:**423–430.
31. **Wermuth, B., H. Burgisser, K. Bohren, and J.-P. Von Wartburg.** 1982. Purification and characterization of human-brain aldose reductase. Eur. J. Biochem. **127:**279–284.
32. **Wermuth, B., J. D. B. Munch, and J.-P. Von Wartburg.** 1977. Purification and properties of NADPH-dependent aldehyde reductase from human liver. J. Biol. Chem. **252:**3821–3828.
33. **Wilson, D. K., K. M. Bohren, K. H. Gabbay, and F. A. Quiocho.** 1992. An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. Science **257:**81–84.
34. **Winters, C. J., D. T. Molowa, and P. S. Guzelian.** 1990. Isolation and characterization of cloned cDNA encoding human liver chlordecone reductase. Biochemistry **29:**1080–1087.
35. **Yamada, H., S. Shimizu, M. Kataoka, H. Sakai, and T. Miyoshi.** 1990. A novel NADPH-dependent aldehyde reductase, catalyzing asymmetric reduction of β-keto acid esters, from *Sporobolomyces salmonicolor*: purification and characterization. FEMS Microbiol. Lett. **70:**45–48.