Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 2738–2742, May 1984
Biochemistry

Complete amino acid sequence for human aldolase B derived from cDNA and genomic clones

(evolution/isozyme family/enzyme mechanism)

WILLIAM H. ROTTMANN, DEAN R. TOLAN, AND EDWARD E. PENHOET*

Department of Biochemistry, University of California, Berkeley, CA 94720

Communicated by Howard K. Schachman, January 20, 1984

ABSTRACT **Several aldolase B clones from a human liver cDNA library have been identified by using a rabbit aldolase A cDNA as a hybridization probe. The most complete of these, pHL413, is 1389 base pairs long and covers ≈80% of the length of the mRNA, including 90% of the translated region. The cDNA, pHL413, was used to identify a genomic clone, λHG313, which encoded the remaining amino acids of human aldolase B. We demonstrate that the amino acid and nucleotide sequences of aldolase are strongly conserved even between different isozymes. Furthermore, in the 3′-untranslated regions of the mRNAs for the B isozyme of human and rat there is an extensive stretch of homology. Aldolase B lacks a cysteine at positions 72 and 338 and lacks a histidine at position 361. These residues, which are present in rabbit aldolase A, have previously been proposed to take part in catalysis. Our findings suggest that this may not be the case.**

Fructose-1,6-bisphosphate aldolase (EC 4.1.2.13) is a glycolytic enzyme that catalyzes the reversible conversion of fructose 1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. The enzyme consists of a tetramer of identical 40,000-dalton subunits. In vertebrates, three isozyme forms exist. These can be distinguished by their electrophoretic and catalytic properties (1). The amino acid sequence of the region around the active site lysine has been determined for aldolases from several sources (2–8). The similarities among these sequences suggest that the amino acid sequence of aldolase is greatly conserved throughout vertebrate evolution. The differences show that they are indeed distinct proteins, the products of a family of related genes. The study of the genes for this enzyme is of interest because the expression of isozyme forms is regulated during development and because they represent a poorly characterized class of genes, the so-called "housekeeping genes," which are expressed in all cells.

Mammalian tissues express the aldolase isozymes in a well-characterized pattern (9). The developing embryo produces aldolase A, which continues to be expressed in many tissues of the adult. In some tissues—for example, muscle—adult expression of aldolase A is dramatically increased over that of the embryo. Here it may represent as much as 5% of total cellular protein (10). In adult liver, kidney, and intestine, aldolase A expression is repressed, and the aldolase B isozyme is produced. In brain and nervous tissue, aldolase A and the third isozyme, aldolase C, are coordinately expressed in roughly equal quantities. Additionally, in some types of cells aldolase expression may be stimulated by insulin (11) or steroids (12). These examples show that control of aldolase expression takes several forms. Investigation of aldolase expression may provide insights into a variety of mechanisms of gene regulation.

The study of the expression of aldolase B is of particular interest. Regulation of the aldolase genes is altered in hepatoma cells (13, 14). In transformed liver cells production of the adult isozyme, B, is shut off and the embryonic form, aldolase A, is produced again. Furthermore, in humans deficiencies of aldolase B result in a genetic disorder, hereditary fructose intolerance (15, 16). Determination of the nature of the defect(s) that cause hereditary fructose intolerance at the genetic level and investigation of the regulation of the gene during development and oncogenesis require nucleic acid probes specific for aldolase B.

This paper describes the isolation and characterization of several cDNA clones for human aldolase B. From the nucleotide sequence of the cDNAs and a genomic clone, we deduce the entire amino acid sequence for this previously uncharacterized protein. This information allows us to confirm the existence of a high degree of homology between the aldolase isozymes throughout their sequences. The derived amino acid sequence contains information relevant to a mechanism of catalysis that has been proposed for aldolase.

MATERIALS AND METHODS

Materials. Poly(G) was from Collaborative Research. ^{32}P-labeled nucleotides were from Amersham. The Klenow fragment of *Escherichia coli* DNA polymerase I came from New England Nuclear. The primer for sequence analysis reactions was prepared by Chiron (Emeryville, CA). Nitrocellulose filters were purchased from Schleicher & Schuell. *Bam*HI, *Eco*RI, *Pst* I, T4 DNA ligase, and *E. coli* DNA polymerase were from New England BioLabs.

Human Liver cDNA Library. The human liver cDNA library was generously supplied by R. Hallewell (Chiron) in the form of supercoiled recombinant plasmid. The vector used was a derivative of pBR322. Double-stranded cDNA was prepared from poly(A)$^+$ RNA from human adult liver according to the method of Villa-Kamaroff *et al.* (17) and was inserted into the *Pst* I site of this plasmid by oligo(dG·dC) tailing (R. Hallewell, personal communication).

Screening of the cDNA Library by Colony Hybridization. *E. coli* strain MM294 was transformed with supercoiled DNA from the plasmid library (18). Colonies were grown on nitrocellulose filters (≈1000 colonies per filter) and screened according to the procedure of Grunstein and Hogness (19) by using a rabbit aldolase A cDNA as a hybridization probe. The cDNA probe encoded the 3′ two-thirds of the rabbit aldolase A mRNA (20). The cDNA insert was excised by restriction endonuclease digestion using *Pst* I, gel purified, and ^{32}P-labeled to a specific activity of 5×10^7 cpm/μg by nick-translation (21). Duplicate nitrocellulose filters with DNA fixed to them were prehybridized with 0.9 M NaCl/0.09 M sodium citrate, pH 7.0, 5× concentrated Denhardt's solution [Denhardt's solution: 0.02% (wt/vol) each of polyvinylpyrrolidone, Ficoll, and bovine serum albumin] 10 μg of denatured sonicated *E. coli* DNA per ml, and 5 μg of poly(G) per

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bp, base pair(s).

*To whom reprint requests should be addressed.

ml at 55°C for 3 hr. The filters were then placed in a fresh sample of the same solution, to which denatured hybridization probe had been added. The concentration of probe was 4×10^5 cpm/ml and 2 ml was used per filter. Hybridization was permitted to take place at 55°C for 16 hr. The filters were washed repeatedly in 0.30 M NaCl/0.03 M sodium citrate, pH 7.0, at 45°C, and then exposed to film. Filters were also probed by using a human aldolase B cDNA, pHL216. When using this probe, the same procedures were followed, except that prehybridization, hybridization, and washes were done at 68°C.

Purification of Plasmid DNA. Selected bacterial clones were grown to $OD_{550} = 0.6$ and then further incubated for 20 hr in the presence of 60 μg of chloramphenicol per ml. The DNA was then harvested by the lysozyme/$NaDodSO_4$ lysis method described by Godson and Vapnek (22), and the supercoiled plasmid DNA was purified by CsCl/ethidium bromide isopycnic density gradient centrifugation.

Screening of Human Genomic Library by Plaque Hybridization. λ Charon 4a containing human genomic DNA (prepared by T. Maniatis' laboratory; ref. 23) was used to infect *E. coli* strain BHB2600, at a density of ≈2×10^4 plaques per plate. DNA from the plaques was transferred to nitrocellulose filters and hybridized to a 1.1-kilobase internal cDNA fragment of pHL413 that was radiolabeled by nick-translation (21). Hybridization was done at 58°C in 0.9 M NaCl/0.09 M sodium citrate, pH 7.0. Approximately 8×10^5 plaques were screened.

Purification of Phage DNA. A 500-ml culture of *E. coli* strain BHB2600 at $OD_{550} = 0.50$ was inoculated with 10^{12} plaque-forming units of λHG313. When the bacteria lysed, the phage were collected by precipitation with polyethylene glycol and purified by centrifugation on discontinuous CsCl gradients (24). The DNA was released from the phage particles by treatment with formamide and was precipitated with ethanol.

Purification of Poly(A)$^+$ RNA. HeLa cells were homogenized in 6 M guanidine thiocyanate, and the RNA was purified by centrifugation through 5.7 M CsCl (25). Poly(A)$^+$ RNA was selected by fractionation on an oligo(dT)-cellulose column (26). Poly(A)$^+$ RNA from adult human liver tissue was a gift from R. Hallewell.

Blot Hybridization of RNA. Poly(A)$^+$ RNA was denatured with formaldehyde and electrophoresed on 1.0% agarose gels containing 6% formaldehyde and 20 mM potassium phosphate at pH 7.0. The RNA was transferred to nitrocellulose filters (27) and probed with nick-translated plasmid pHL413 in a buffer of 50% deionized formamide/0.75 M NaCl/0.075 M sodium citrate/10 mM Hepes, pH 7.0/200 μg of sonicated calf thymus DNA per ml/1× concentrated Denhardt's solution. For stringent hybridization, the filters were incubated in this buffer at 55°C and washed in 0.30 M NaCl/0.03 M sodium citrate, pH 7.0, at 65°C. Nonstringent hybridization took place at 37°C, and the filters were washed with 0.03 M NaCl/0.03 M sodium citrate, pH 7.0, at 55°C.

Sequence Analysis. The cDNA inserts were excised from the plasmids by restriction endonuclease digestion using *Pst* I, according to the procedure recommended by the manufacturer. The inserts or their fragments were gel purified and subcloned into the bacteriophage vector M13mp8 at the *Pst* I site (28). Similarly, two 3.4-kilobase *Eco*RI–*Bam*HI fragments were purified from a digestion of λHG313 and ligated into *Eco*RI-, *Bam*HI-digested M13mp8. Sequence analysis was performed by the dideoxy technique of Sanger *et al.* (29) with a commercially prepared oligonucleotide complementary to M13mp8 as a primer for the Klenow fragment of *E. coli* DNA polymerase I.

Computer Analysis. DNA and protein sequences were analyzed by using the programs of H. Martinez with the University of California, Berkeley, VAX/UNIX system.

RESULTS

A number of cDNA clones were selected from 4×10^4 human liver cDNA transformants in two separate screenings. A rabbit aldolase A cDNA and later one of the human liver cDNAs were used as hybridization probes. The largest of the clones isolated, pHL413, was used to hybridize to various RNAs (Fig. 1). Under stringent conditions pHL413 hybridized to a 1700 ± 25-nucleotide RNA from human liver but failed to hybridize to RNA from HeLa cells that produce aldolase A (30). Under low stringency a rabbit aldolase A cDNA will hybridize to this HeLa RNA showing a 1725 ± 25-nucleotide mRNA for aldolase A. We conclude that pHL413 and the other overlapping clones encode the human aldolase B mRNA and not aldolase A, which is found in liver as ≈5% of the total aldolase (9).

The cDNA inserts of four partially overlapping clones were subcloned into the *Pst* I site of the bacteriophage vector M13mp8 and subjected to sequence analysis. The clones showed homology to the rabbit aldolase A cDNA used for a probe and corresponded to the coding region from amino acid 34 to the carboxyl terminus of the protein, with an additional 395 nucleotides of the 3'-untranslated region. Fig. 2 shows how three of these clones, pHL413, pHL216, and pHL201, correspond to the full-length message. The two clones pHL216 and pHL201 overlapped each other by 8 base pairs (bp) and both overlapped the 1389-bp pHL413 clone.

To determine the first 33 amino acids, it was necessary to sequence a genomic clone of human aldolase B. Costanza *et al.* have sequenced a 111-bp human aldolase B cDNA, R38, which predicts a *Bam*HI site at the codons for amino acids 28 and 29 (31). The regions adjacent to the unique *Bam*HI site of the insert of λHG313 were subjected to sequence analysis. Fig. 3 shows the overlap of this genomic DNA sequence with other human aldolase B DNA sequences. The sequences of λHG313 and pHL413 overlapped by only 12 bp. This was due to the existence of an intervening sequence that interrupts the coding region of the gene at amino acid 37 (unpublished data). However, the clone R38 (30) bridges this junction (Fig. 3). The excellent agreement between the nucleotide sequences of λHG313, R38, and pHL413 suggests that they all correspond to the same protein. Three discrepancies were found. These can be attributed to allelic differences or errors in sequence analysis. The high degree of overlapping sequence indicated that this fragment of genomic DNA was from the human aldolase B gene where the codons for the first 37 amino acids were encoded.

The amino acid sequence for human liver aldolase derived from these clones can be compared to the known partial amino acid sequences of aldolases from other sources. The tryptic peptide containing the active site lysine (position 229) has been sequenced for several aldolases (2–8). In Fig. 4, the aldolase B amino acid sequences from ox, rat, rabbit, and human are compared to the same peptide of aldolases A and C. The aldolase B isozymes have amino acid residues in common that distinguish them from aldolases A and C. At

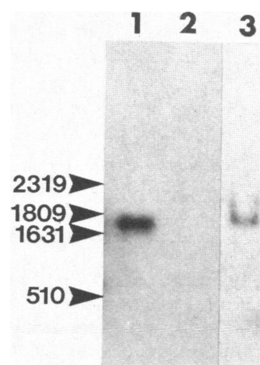

FIG. 1. Hybridization of aldolase cDNAs to poly(A)$^+$ RNA. Lane 1, 1 μg of liver poly(A)$^+$ RNA probed with pHL413 under stringent conditions; lane 2, 10 μg of HeLa poly(A)$^+$ RNA probed with pHL413 under stringent conditions; lane 3, 10 μg of HeLa poly(A)$^+$ RNA probed with rabbit aldolase A cDNA under nonstringent conditions. The positions of markers indicating size in nucleotides are shown on the left.

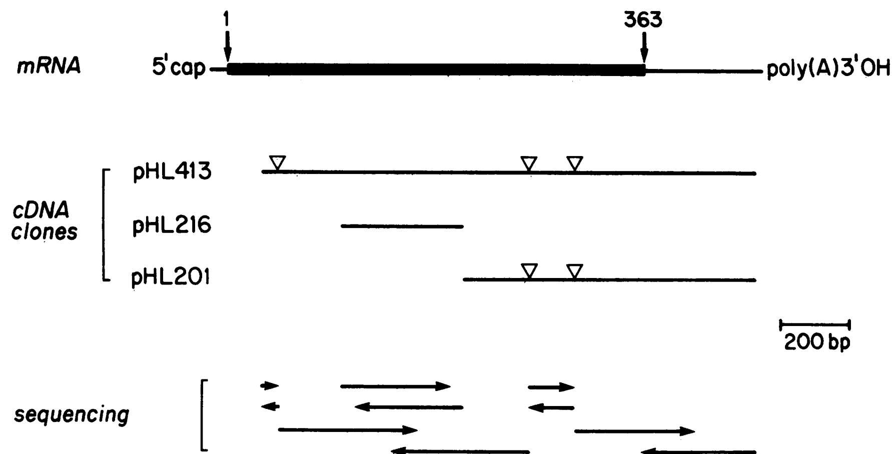

FIG. 2. Physical map of cDNA inserts relative to mRNA. The presumed structure of the aldolase B mRNA is at the top, showing the positions of the codons for the first and last amino acids of the protein. The lengths of the 5'- and 3'-untranslated regions are not known. Underneath this the cDNA inserts are depicted, showing how they correspond to the mRNA and each other. pHL201 and pHL216 overlap by 8 nucleotides (data not shown). Internal *Pst* I sites are indicated by ▽. The sequence analysis strategy is described at the bottom. The horizontal arrows indicate the direction of sequence analysis from the internal and flanking *Pst* I sites. Each arrow represents a sequence that was determined in two independent trials. Forty percent of the DNA was subjected to sequence analysis in both directions.

position 217, the B isozymes have an asparagine where A and C have a serine, and at position 241 there is a lysine where A and C have a glutamic acid or glutamine residue. The derived sequence of the cloned human aldolase was identical to the ox and rabbit aldolase B in this region, suggesting that the clones correspond to the B isozyme.

The sequence reported here along with the protein translation is summarized in Fig. 5. This sequence corresponds to the entire amino acid sequence of aldolase B and 88% of the length of the mRNA. The structure of the message is not known precisely. Assuming that the poly(A) tail is 100 nucleotides long, there remain ≈100 nucleotides in the 5'- and 3'-untranslated regions of the message that are yet to be determined.

DISCUSSION

Tsutsumi *et al.* (8) have published a 1200-nucleotide cDNA clone of rat aldolase B, which corresponded to the mRNA from the codon for amino acid 184 through to the poly(A) tail. Comparison of the sequences of the rat and human aldolase cDNA clones showed that they share an overall nucleotide homology of 88% in the coding region. The amino acid sequences derived from the DNA sequences are 94% homologous. In addition, there is some conservation of sequences in the 3'-nontranslated region. Sequences that are ≈70% homologous to the underlined regions of Fig. 5 can be found in the 3'-untranslated region of the rat aldolase B cDNA, ending 20 nucleotides upstream from the polyadenylylation signal. These sequences have remained similar during the 6×10^7 years since these species' lineages diverged (32). This region may possibly represent some feature of control of aldolase B expression or reflect the slower rates of evolutionary change for these genes. This question cannot be resolved at this time.

The complete amino acid sequence of rabbit aldolase A has been known for some time (2). With the information that has been made available by sequence analysis of these aldolase B clones, some interesting comparisons may be made between the isozymes. Though there are some salient differences, which will be discussed below, the amino acid sequence homology between aldolases A and B is quite high. There is an overall homology of 68% between human aldolase B and rabbit aldolase A. The greatest homology occurs in the middle region of the sequence, near the active site lysine. Between positions 165 and 265, 84% of the amino

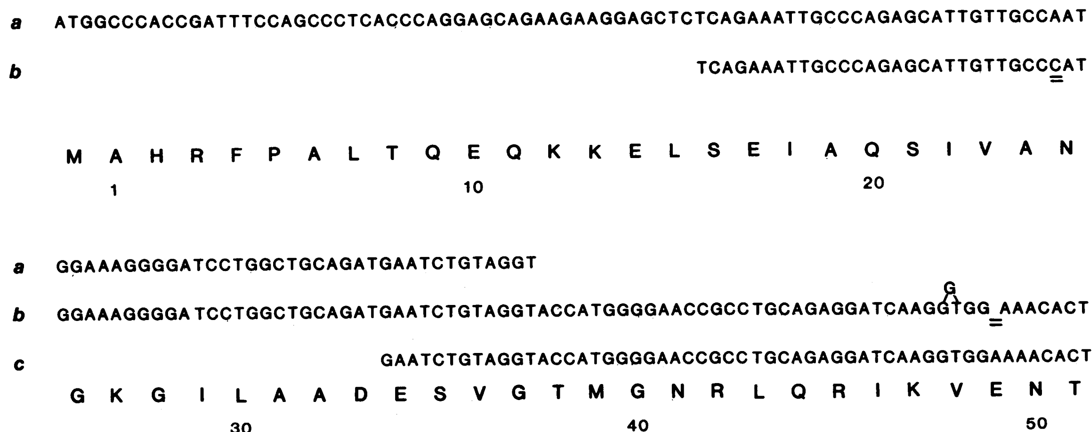

FIG. 3. Comparisons of genomic DNA and cDNA nucleotide sequences for the amino-terminal region of human aldolase B. (*a*) Sequence of coding region from λHG313. (*B*) Sequence of a cDNA determined by Costanza *et al.* (31). (*c*) Sequence from the 5' portion of pHL413. Differences in sequences are underlined or looped out of the alignment. Amino acids −1 to 51, derived from the sequences of λHG313 and pHL413, are shown underneath the nucleotide sequences, designated in the one-letter code.

					220										230										240		
Rabbit A	Ala	Leu	Ser	Asp	His	His	Ile	Tyr	Leu	Glu	Gly	Thr	Leu	Leu	Lys	Pro	Asn	Met	Val	Thr	Pro	Gly	His	Ala	Cys	Thr	Gln	Lys
Frog A	Ala	Leu	Ser	Asx	His	His	Val	Tyr	Leu	Glx	Gly	Thr	Leu	Leu	Lys	Pro	Asx	Met	Val	Thr	Ala	Gly	Asx	Ala	Cys	Thr	Glx	Lys
Cod A	Ala	Leu	Ser	Asp	His	His	Val	Tyr	Leu	Gln	Gly	Thr	Leu	Leu	Lys	Pro	Asp	Met	Val	Thr	Ala	Gly	His	Ser	Cys	Thr	Gln	Lys
Rabbit C	Ala	Leu	Ser	Asx	His	His	Ile	Tyr	Val	Glx	Gly	Thr	Leu	Leu	Lys	Pro	Glx	Met	Val	Thr	Pro	Gly	Asx	Ala	Cys	Thr	Glx	Lys
Rabbit B	Ala	Leu	Asn	Asp	His	His	Val	Tyr	Leu	Glu	Gly	Thr	Leu	Leu	Lys	Pro	Asn	Met	Val	Thr	Ala	Gly	His	Ala	Cys	Thr	Lys	
Ox B	Ala	Leu	Asn	Asp	His	His	Val	Tyr	Leu	Glu	Gly	Thr	Leu	Leu	Lys	Pro	Asn	Met	Val	Thr	Ala	Gly	His	Ala	Cys	Thr	Lys	Lys
Rat B	Ala	Leu	Asn	Asp	His	His	Val	Tyr	Leu	Glu	Gly	Thr	Leu	Leu	Lys	Pro	Asn	Met	Leu	Thr	Ala	Gly	His	Ala	Cys	Thr	Lys	Lys
Human	Ala	Leu	Asn	Asp	His	His	Val	Tyr	Leu	Glu	Gly	Thr	Leu	Leu	Lys	Pro	Asn	Met	Val	Thr	Ala	Gly	His	Ala	Cys	Thr	Lys	Lys

FIG. 4. Comparison of the amino acid sequences near the active site lysine for several aldolases. The boxes enclose residues that are common to all B aldolases and that distinguish them from other isozymes. Rabbit A, ref. 3; frog A, ref. 5; cod A, ref. 6; rabbit C, ref. 2; rabbit B and ox B, ref. 7; rat B, ref. 8.

acid residues are conserved between rabbit aldolase A and human aldolase B.

Our results indicate that the nucleic acid sequences of the aldolase A and B genes have been strongly conserved as well. Since all vertebrates express both these isozymes, the genes probably have been distinct and have had the opportunity to diverge for at least 5×10^8 years (32). Yet, sequences from aldolases A and B, even from different species, are readily able to hybridize to each other, showing that significant portions of the nucleotide sequences are quite similar to each other. Comparison of the nucleotide sequence for human aldolase B with the cDNA for rabbit aldolase A (20) shows that the nucleotide sequences are conserved in much the same pattern as the amino acid sequences. The overall homology of the coding sequences is 70%, though in the region corresponding to the most highly conserved amino ac-

```
-1   1                                      10                                                          20
Met Ala His Arg Phe Pro Ala Leu Thr Gln Glu Gln Lys Lys Glu Leu Ser Glu Ile Ala Gln Ser Ile Val Ala Asn Gly Lys Gly Ile
ATG GCC CAC CGA TTT CCA GCC CTC ACC CAG GAG CAG AAG AAG GAG CTC TCA GAA ATT GCC CAG AGC ATT GTT GCC AAT GGA AAG GGG ATC  (90)

30                                          40                                          50
Leu Ala Ala Asp Glu Ser Val Gly Thr Met Gly Asn Arg Leu Gln Arg Ile Lys Val Glu Asn Thr Glu Glu Asn Arg Arg Gln Phe Arg
CTG GCT GCA GAT GAA TCT GTA GGT ACC ATG GGG AAC CGC CTG CAG AGG ATC AAG GTG GAA AAC ACT GAA GAG AAC CGC CGG CAG TTC CGA  (180)

60                                          70                                          80
Glu Ile Leu Phe Ser Val Asp Ser Ser Ile Asn Gln Ser Ile Gly Gly Val Ile Leu Phe His Glu Thr Leu Tyr Gln Lys Asp Ser Gln
GAA ATC CTC TTC TCT GTG GAC AGT TCC ATC AAC CAG AGC ATC GGG GGT GTG ATC CTT TTC CAC GAG ACC CTC TAC CAG AAG GAC AGC CAG  (270)

90                                          100                                         110
Gly Lys Leu Phe Arg Asn Ile Leu Lys Glu Lys Gly Ile Val Val Gly Ile Lys Leu Asp Gln Gly Gly Ala Pro Leu Ala Gly Thr Asn
GGA AAG CTG TTC AGA AAC ATC CTC AAG GAA AAG GGG ATC GTG GTG GGA ATC AAG TTA GAC CAA GGA GGT GCT CCT CTT GCA GGA ACA AAC  (360)

120                                         130                                         140
Lys Glu Thr Thr Ile Gln Gly Leu Asp Gly Leu Ser Glu Arg Cys Ala Gln Tyr Lys Lys Asp Gly Val Asp Phe Gly Lys Trp Arg Ala
AAA GAA ACC ACC ATT CAA GGG CTT GAT GGC CTC TCA GAG CGC TGT GCT CAG TAC AAG AAA GAT GGT GTT GAC TTT GGG AAG TGG CGT GCT  (450)

150                                         160                                         170
Val Leu Arg Ile Ala Asp Gln Cys Pro Ser Ser Leu Ala Ile Gln Glu Asn Ala Asn Ala Leu Ala Arg Tyr Ala Ser Ile Cys Gln Gln
GTG CTG AGG ATT GCC GAC CAG TGT CCA TCC AGC CTC GCT ATC CAG GAA AAC GCC AAC GCC CTG GCT CGC TAC GCC AGC ATC TGT CAG CAG  (540)

180                                         190                                         200
Asn Gly Leu Val Pro Ile Val Glu Pro Glu Val Ile Pro Asp Gly Asp His Asp Leu Glu His Cys Gln Tyr Val Thr Glu Lys Val Leu
AAT GGA CTG GTA CCT ATT GTT GAA CCA GAG GTA ATT CCT GAT GGA GAC CAT GAC CTG GAA CAC TGC CAG TAT GTT ACT GAG AAG GTC CTG  (630)

210                                         220                                         230
Ala Ala Val Tyr Lys Ala Leu Asn Asp His His Val Tyr Leu Glu Gly Thr Leu Leu Lys Pro Asn Met Val Thr Ala Gly His Ala Cys
GCT GCT GTC TAC AAG GCC CTG AAT GAC CAT CAT GTT TAC CTG GAG GGC ACC CTG CTA AAG CCC AAC ATG GTG ACT GCT GGA CAT GCC TGC  (720)

240                                         250                                         260
Thr Lys Lys Tyr Thr Pro Glu Gln Val Ala Met Ala Thr Val Thr Ala Leu His Arg Thr Val Pro Ala Ala Val Pro Gly Ile Cys Phe
ACC AAG AAG TAT ACT CCA GAA CAA GTA GCT ATG GCC ACC GTA ACA GCT CTC CAC CGT ACT GTT CCT GCA GCT GTT CCT GGC ATC TGC TTT  (810)

270                                         280                                         290
Leu Ser Gly Gly Met Ser Glu Glu Asp Ala Thr Leu Asn Leu Asn Ala Ile Asn Leu Cys Pro Leu Pro Lys Pro Trp Lys Leu Ser Phe
TTG TCT GGT GGC ATG AGT GAA GAG GAT GCC ACT CTC AAC CTC AAT GCT ATC AAC CTT TGC CCT CTA CCA AAG CCC TGG AAA CTA AGT TTC  (900)

300                                         310                                         320
Ser Tyr Gly Arg Ala Leu Gln Ala Val Ala Leu Ala Ala Trp Gly Gly Lys Ala Ala Asn Lys Glu Ala Thr Gln Glu Ala Phe Met Lys
TCT TAT GGA CGG GCC CTG CAG GCA GTA GCA CTG GCT GCC TGG GGT GGC AAG GCT GCA AAC AAG GAG GCA ACC CAG GAG GCT TTT ATG AAG  (990)

330                                         340                                         350
Arg Ala Met Ala Asn Cys Gln Ala Ala Lys Gly Gln Tyr Val His Thr Gly Ser Ser Gly Ala Ala Ser Thr Gln Ser Leu Phe Thr Ala
CGG GCC ATG GCT AAC TGC CAG GCG GCC AAA GGA CAG TAT GTT CAC ACG GGT TCT TCT GGG GCT GCT TCC ACC CAG TCG CTC TTC ACA GCC (1080)

360         363
Cys Tyr Thr Tyr ter
TGC TAT ACC TAC TAG GGTCCAATGCCCGCCAGCCTAGCTCCAGTGCTTCTAGTAGGAGGGCTGAAAGGGAGCAACTTTTCCTCCAATCCTGGAAATTCGACACAATTAGATTTG (1194)

AACTCGCTGGAAATACAACACATGTTAAATCTTAAGTACAAGGGGGAAAAAATAAATCAGTTATTTGAAACATAAAAATGAATACCAAGGACCTGATCAAATTTCACACAGCAGTTTCC (1313)

TTGCAACACTTTCAGCTCCCCATGCTCCAGAATACCCACCCAAGAAAATAATAGGCTTTAAAACAATATCGGCTCCTCATCCAAAGAACAACTGCTGATTGAAACACCTCATTAGCTGA (1432)

GTGTAGAGAAGTGCATCTTATGAAACAGTCTTAGCAGTGGTAGGTTGGGAAGGAGATAG                                                            (1491)
```

FIG. 5. Combined nucleotide sequences from HG313 and pHL413 with derived amino acid sequence for human aldolase B. The amino acid numbering is according to Lai *et al.* (3). The sequence from −1 to 37 is from λHG313, and the sequence from 34 to the end is from pHL413. The underlined segments are also found in the untranslated region of the rat aldolase B message (8). The oligo(dG·dC) tails flanking the cDNA have been omitted.

ids, the homology reaches 75%. There are no significant homologies between the 3'-untranslated regions of the messages of these isozymes.

In addition to providing information on the structures of the proteins and their messages, these sequence analysis data allow some conclusions to be drawn concerning the mechanism of catalysis by aldolase. Several amino acid residues have been proposed to take part in substrate binding and catalysis on the basis of enzyme inhibition by covalent modification of rabbit aldolase A (3). A useful test of the accuracy of the proposed schemes would be to see which of these residues are conserved between the different aldolases. The high degree of conservation of the amino acid sequences of the isozymes indicates that the structure of aldolase is under severe constraints. Amino acid residues directly involved in the catalytic process would unlikely be changed as the different isozymes diverged.

Several other amino acid residues besides those around the active site lysine 229 (Fig. 3) are also conserved between aldolases A and B. Arg-55, Lys-146, and Arg-148 have been implicated in binding of the substrate (33, 34) and are all present in the new sequence. Likewise, Tyr-363 is conserved in aldolase B. However, some important differences exist. Lai *et al.* (3) have proposed that Lys-107, Cys-72 or Cys-338, and His-361 take part in a proton transfer system that is necessary for catalysis. While Lys-107 is found in the human aldolase B sequence, position 72 is now a serine, position 338 is an alanine, and position 361 is a tyrosine. The published rat aldolase sequence corroborates our findings at positions 338 and 361.

The question of whether a cysteine is essential for aldolase activity has long been debated. Though modification of several cysteine residues can be correlated with loss of activity (35), no one has conclusively shown that this is due to loss of a catalytic moiety. However, Anderson (36) has reported that in sturgeon aldolase A, Cys-338 does not exist and, furthermore, that Cys-72 can be modified by certain agents without loss of enzyme activity. On the basis of these data, Anderson has suggested that cysteine is not a catalytically essential residue, but the observed enzyme inactivation caused by chemical modification of Cys-72 and Cys-338 is due to steric hindrance or changes in the conformation of the aldolase subunits. Our data support this hypothesis because human aldolase B lacks cysteine at both of these sites, although we cannot eliminate the possibility of another cysteine in the enzyme taking part in catalysis.

The involvement of a histidine at position 361 in catalysis stems from substantial evidence indicating that modification of His-361 inhibits the activity of aldolase A (37, 38). Our data clearly indicates a TAT coding for tyrosine at this position. The fact that this has been found in two independent clones, pHL201 and pHL413, as well as in a rat aldolase B clone (8), suggests that this is not an artifact of cloning. It may be that the explanation is similar to that suggested above for the cysteine residues; modification of the residue in question lessens enzymatic activity due to alterations of the protein conformation, not because the residue actually takes part in the catalytic process.

These clones can be used to further probe the λ library of human genomic DNA, in an attempt to obtain complete genes for the human aldolases. In addition to the gene coding for aldolase B, these probes should be able to hybridize to genes for aldolases A and C as well as any pseudogenes for aldolase that might exist.

We are very grateful to Rob Hallewell for his gifts of the human cDNA library and RNA. We also thank Karen Erdley for her help in preparing the manuscript. This work was supported by Grant NIH-CA20347 from the National Institutes of Health.

1. Penhoet, E., Kochman, M., Valentine, R. & Rutter, W. J. (1967) *Biochemistry* **6**, 2940–2949.
2. Horecker, B. L., Tsolas, O. & Lai, C. Y. (1975) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), Vol. 7, pp. 213–258.
3. Lai, C. Y., Nakai, N. & Chang, D. (1974) *Science* **183**, 1204–1206.
4. Benfield, P. A., Forcina, B. G., Gibbons, I. & Perham, R. N. (1979) *Biochem. J.* **183**, 429–444.
5. Ting, S.-M., Sia, C. L., Lai, C. Y. & Horecker, B. L. (1971) *Arch. Biochem. Biophys.* **144**, 485–490.
6. Lai, C. Y. & Chen, C. (1971) *Arch. Biochem. Biophys.* **144**, 467–475.
7. Ting, S.-M., Lai, C. Y. & Horecker, B. L. (1971) *Arch. Biochem. Biophys.* **144**, 476–484.
8. Tsutsumi, K., Mukai, T., Hikada, S., Miyahara, H., Tsutsumi, R., Tanaka, T., Hori, K. & Ishikawa, K. (1983) *J. Biol. Chem.* **258**, 6537–6542.
9. Lebherz, H. G. & Rutter, W. J. (1969) *Biochemistry* **8**, 109–121.
10. Penhoet, E. E., Kochman, M. & Rutter, W. J. (1969) *Biochemistry* **8**, 4396–4402.
11. Spiegelman, B. M. & Green, H. (1980) *J. Biol. Chem.* **255**, 8811–8818.
12. Mainwaring, W. I. P., Mangan, F. R., Irving, R. A. & Jones, D. A. (1974) *Biochem. J.* **144**, 413–426.
13. Shapira, F., Dreyfus, J. C. & Shapira, G. (1963) *Nature (London)* **200**, 995–996.
14. Adelman, R. C., Morris, H. P. & Weinhouse, S. (1967) *Cancer Res.* **27**, 2408–2413.
15. Shapira, F., Shapira, G. & Dreyfus, J. C. (1961) *Enzymol. Biol. Clin.* **1**, 170–175.
16. Froesch, E. R., Wolf, H. P., Baitsch, H., Prader, A. & Labart, A. (1963) *Am. J. Med.* **34**, 151–167.
17. Villa-Kamaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naker, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3727–3731.
18. Dagert, M. & Ehrlich, S. D. (1979) *Gene* **6**, 23–28.
19. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
20. Tolan, D. R., Amsden, A. B., Putney, S. D., Urdea, M. S. & Penhoet, E. E. (1984) *J. Biol. Chem.* **259**, 1127–1131.
21. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
22. Godson, G. N. & Vapnek, D. (1973) *Biochim. Biophys. Acta* **299**, 516–520.
23. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
24. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L. & Treiber, G. (1970) *Virology* **40**, 734–744.
25. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
26. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
27. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
28. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
29. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5487.
30. Gliksman, R., Ghosh, N. K. & Cox, R. P. (1977) *Enzyme* **22**, 416–419.
31. Costanza, F., Castagnoli, L., Dente, L., Arcari, P., Smith, M., Costanzo, P., Rangei, G., Izzo, P., Peitropaolo, T. C., Bougueleret, L., Cimino, F., Salvatore, F. & Cortese, R. (1983) *EMBO J.* **2**, 57–61.
32. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure (1972–1978)*, (National Biomedical Research Foundation, Georgetown Univ. Medical Center, Washington, DC).
33. Hartman, F. C. & Brown, J. B. (1976) *J. Biol. Chem.* **251**, 3057–3062.
34. Patthy, C., Varadi, A., Thesz, J. & Kovacs, K. (1979) *Eur. J. Biochem.* **99**, 309–313.
35. Steinman, H. M. & Richards, F. M. (1970) *Biochemistry* **9**, 4361–4372.
36. Anderson, P. J. (1972) *Can. J. Biochem.* **50**, 111–119.
37. Hoffee, P., Lai, C. Y., Pugh, E. L. & Horecker, B. L. (1966) *Proc. Natl. Acad. Sci. USA* **57**, 107–113.
38. Hartman, F. C. & Welch, M. H. (1974) *Biochem. Biophys. Res. Commun.* **57**, 85–92.