Characterization of the Chicken Aldolase B Gene*

Diane G. Burgess‡ and Edward E. Penhoet§
From the Department of Biochemistry, University of California, Berkeley, California 94720

Vertebrates possess three isozyme forms of fructose diphosphate aldolase. We have isolated two overlapping chicken genomic clones which encode the liver-specific form of this enzyme; we have identified the 5' and 3' ends of this gene by a combination of primer extension analysis and S1 mapping; and we have determined the entire nucleotide sequence of this gene including 1400 base pairs (bp) of sequence from the regions flanking the 5' and 3' ends of the gene. The transcriptional unit for the aldolase B gene spans 8700 bp and contains eight intervening sequences, including a 4600-bp intron in the 5' non-coding region. On the basis of results from Southern genomic hybridizations, the aldolase B gene appears to be present only once per haploid genome. No differences were detected in the mRNA structure between RNA from three tissues expressing aldolase B (liver, kidney, and small intestine). Various features of the 5' flanking region are discussed, including a partial homology with the 5' non-coding region from rabbit aldolase A.

Fructose 1,6-bisphosphate aldolase (E.C. 4.1.2.13) is a glycolytic enzyme which catalyzes the reversible cleavage of fructose 1,6-bisphosphate to form dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. Vertebrates possess three aldolase isozymes which resemble each other in molecular weight, subunit structure, and overall catalytic reaction; both in vivo and in vitro they readily form mixed hybrids (1). Differences in amino acid sequence (2, 3), immunological properties (4), and catalytic activities make it clear that they are encoded by separate genes; however, overall conservation in amino acid sequence suggests that they are probably derived by divergent evolution from a common ancestral gene.

The catalytic properties of the isozymes differ according to the specialized roles which they perform (5). The isozymes are expressed in well-characterized tissue-specific patterns which reflect these roles. Aldolase A is the isozyme most efficient in glycolysis; it is found in muscle and other tissues with a high rate of glycolysis. It is also the isozyme most predominant in mammalian embryos. Aldolase B is adapted to participate in fructose metabolism and gluconeogenesis; it is expressed only in those tissues active in these pathways (liver, kidney, and small intestine). Aldolase C has intermediate catalytic properties; it is found in brain and nervous tissue. It is also the most prevalent isozyme in avian embryos.

We are interested in understanding the mechanisms involved in regulating the tissue-specific expression of this family of genes. We have previously reported the nucleotide sequence for cloned cDNAs from rabbit aldolase A (2) and human aldolase B (3) and are presently analyzing their respective gene structures. By comparing these genes with a non-mammalian aldolase gene, we can identify conserved regions which may be important in the tissue-specific regulation of gene expression. For this reason, we have characterized the aldolase B gene from chicken. In this paper we report the nucleotide sequence of the entire chicken aldolase B gene including the regions flanking the transcriptional unit on both the 5' and 3' ends.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes were purchased from New England Biolabs and Bethesda Research Laboratories. HindIII was the gift of J. C. Rahibowitz of this department. DNA polymerase I was from Bethesda Research Laboratories. DNA polymerase I Klenow fragment was from Boehringer Mannheim, T4 ligase was from New England Biolabs, reverse transcriptase was from Life Sciences Inc., and S1 nuclease was from Miles Laboratories Inc. Oligonucleotide primers were prepared by Mickey S. Urdea of Chiron Corp., Emeryville, CA. Radiochemicals were purchased from Amersham Corp. Guanidinium thiocyanate was from Pharmacia Biotech, Inc., P-L Biochemicals. Nitrocellulose BA55 paper and filters were from Schleicher and Schuell and autoradiography was done using Kodak XAR-5 x-ray film.

Screening a Chicken Genomic Library—The chicken genomic library was the generous gift of James D. Engel of Northwestern University. It was constructed by ligating a partial HaellII and AluI restriction digest into Charon 4A arms (6). The library was screened by the in situ plaque hybridization technique of Benton and Davis (7) at a density of 10,000-20,000 plaques/8.5-em plate. Approximately 1.5 × 10⁵ plaques were examined in two sets of screenings. The probe used for the initial screening was a 500-nucleotide PstI fragment from a rabbit aldolase A cDNA clone, pRM218 (2); this probe encodes amino acids 140-306 of aldolase. The second screening was performed with a 1200-nucleotide HindI fragment from the human aldolase B cDNA clone pHL413 (3); this probe contains the coding region for amino acids 105-363 and 400 nucleotides of 3' untranslated region. Both probes were purified by electrophoresis from agarose gels and radiolabeled by nick translation (8). Filters from the first screening were hybridized for 36 h at 55 °C in 1 M NaCl, 50 mM Tris-Cl, pH 7.5, 1 mM EDTA, 5 × Denhardt's (0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 0.1% sodium dodecyl sulfate, and 50 μg/ml salmon sperm DNA. Filters from the second screening were treated in a similar fashion except hybridization and washes were performed at 60 °C. All work with recombinant DNA molecules was done under P2 levels of containment in accordance with the NIH Guidelines for Recombinant DNA Research.

Preparation of Phage DNA—Phage stocks of Charon 4A recombinants were prepared from plate lysates. Phage were precipitated with polyethylene glycol as described by Yamamoto et al. (9) and purified through two CsCl step gradients. DNA was released by incubating phage for 1 h in 50% deionized formamide, 100 mM Tris-Cl, pH 8.5, 10 mM EDTA (10). The DNA was then precipitated at room temper-

* This work was supported by Grant NIH-CA20347 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported during part of these studies by a National Science Foundation predoctoral fellowship. Present address: Department of Genetics, University of California, Berkeley, CA 94720.

§ To whom correspondence should be addressed.
Characterization of the Chicken Aldolase B Gene

ature with ethanol, extracted with phenol, and ethanol precipitated a final time.

Restriction Mapping and Localization of the Aldolase Gene—The region of the genomic clones encoding aldolase was located by hybridizing pRM223 (2), a rabbit aldolase A cDNA clone containing rabbit amino acids 61-363 and 110 nucleotides of 3' noncoding region, to a Southern blot to Size-fractionated EcoRI digest of XC11. The probe was hybridized to three EcoRI fragments (1400, 830, and 670 bp in length). These fragments were cloned into an M13 vector and sequenced.

A restriction map of the Charon 4A recombinant clones was prepared by subcloning a set of restriction fragments into pUC vectors (15) and individually mapping each subclone with restriction enzymes. Restriction digests were carried out under conditions recommended by the suppliers. The location of the pUC subclones with respect to the genomic clones was determined by hybridizing radiolabeled subclones back to Southern blots (11) of DNA from XC11.1 and XC4 restriction digests.

Preparation of Plasmid DNA—Plasmid DNA from pUC subclones was purified in small quantities by the alkaline extraction procedure of Birnboim (13) or in large quantities by the procedure of Godson and Vapnek (14) followed by a final purification over a CsCl-ethidium bromide isopycnic density gradient. Chloramphenicol amplification of the plasmid DNA was omitted because this step resulted in very poor plasmid yields for certain pUC recombinant subclones.

Preparation of Radiolabeled Probes—Double-stranded DNA was radiolabeled by nick translation (8) and single-stranded M13 recombinant templates were radiolabeled by primer extension (15).

DNA Sequence Analysis—DNA digested with restriction enzymes was subcloned into the M13 vectors mp8, mp9, mp10, mp11, mp18, and mp19 (16, 17) and sequenced using the dideoxy chain termination method of Sanger et al. (18).

Preparation of poly(A) RNA—RNA was purified from fresh tissue from White Leghorn chickens by the guanidinium thiocyanate procedure of Chirgwin et al. (19). Poly(A) RNA was selected by chromatography over oligo(dT)-cellulose as described by Aviv and Leder (20). Muscle poly(A) RNA from 19-day-old embryos was the gift of Juan C. Mendible, Universidad Central de Venezuela.

Preparation of single-stranded DNA.—DNA was separated by electrophoresis on formaldehyde agarose gels (21) and transferred to nitrocellulose paper (22). Hybridization was carried out in 5 × SSC, 1 × Denhardt’s, 5 mg/ml HEPES, pH 7.0, 1 mM EDTA, 0.2 mg/ml sonicated calf thymus DNA, and 50% deionized formamide in the presence of 2.5 × 10^6 to 1 × 10^7 cpm/ml of a radiolabeled probe.

Primer Extension—Poly(A) RNA (15 μg) was mixed with 1 pmol of a single-stranded deoxyoligonucleotide primer (5′GGAGACATGTCGGTGGG3′ or 5′CTCGGCTGATAAGATCCC′′), denatured at 80 °C, and quick-cooled on ice. Primer was annealed to RNA in a 10-min incubation at 80 °C in 50 mM Tris·Cl, pH 8.7, 7.5 mM KCl, 8 mM MgCl₂, and 0.075 mM dithiothreitol. The primer extension reaction was performed in the same buffer in a 20-μl volume containing 1 mM dGTP, 1 mM dATP, 1 mM dTTP, 0.075 mM dCTP, 40 μCi of [α-32P]dCTP (410 Ci/mmol), and 12.5 units of reverse transcriptase. The reaction was allowed to proceed for 15 min at 42 °C and chased for another 10 min by the addition of dCTP to 1 mM. The reaction was stopped by the addition of EDTA to 25 mM, protein was removed by phenol/chloroform extraction, and nucleic acid was recovered by ethanol precipitation. The sample was dissolved in 60% formamide, denatured at 95 °C, and subjected to electrophoresis on an 80% deionized formamide. The reaction was diluted into 280 mM NaCl, 50 mM sodium acetate, pH 4.6, 4.5 mM ZnSO₄, and 20 μg/ml sonicated calf thymus DNA and unincorporated DNA was digested for 30 min with 2500 units of S1 nuclease at 37 °C.

Hybridization to Genomic DNA—High molecular weight genomic DNA was prepared from chicken heart by the procedure of Robbins et al. (23). Genomic DNA was digested with restriction enzymes, size-fractionated on a 0.8% agarose gel at 4 °C in 40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, pH 7.3, and transferred to nitrocellulose paper. The filters were prehybridized for several hours at 50 °C in 50% deionized formamide, 5 × SSC, 5 × Denhardt’s, 50 mM sodium phosphate, pH 6.5, 20 mM glycine, and 500 μg/ml single-stranded sonicated salmon sperm DNA. The genomic DNA was hybridized with the subclone p3α (see Fig. 1), which had been radiolabeled to a specific activity of 6 × 10⁶ cpm/μg. The subclone p3α contains a 1325-bp insert which includes the entire last exon, 205 bp of the last intron, and 720 bp of 3′ flanking region. Hybridization was performed at 50 °C in 50% deionized formamide, 5 × SSC, 1 × Denhardt’s, 20 mM sodium phosphate, pH 6.5, 200 μg/ml sonicated single-stranded salmon sperm DNA, and 10% dextran sulfate. The filter was washed at 50 °C in 0.2 × SSC, 0.1% sodium dodecyl sulfate.

Computer Analysis—The DNA sequence was analyzed on a VAX/UNIX system using computer programs supplied by Hugo Martinez (University of California at San Francisco Bioinformatics Computation Laboratory).

RESULTS

Isolation and Restriction Mapping of Genomic Aldolase Clones—Two recombinant clones were isolated from a chicken genomic library using probes derived from mammalian aldolase cDNA clones. These genomic clones were designated XC11.1 and XC4. The clones were shown to be overlapping by restriction mapping analysis (see Fig. 1). Within the 7100 bp of overlap no differences were found in restriction endonuclease sites. Together these clones span a region of 27.6 kb. To facilitate the characterization of the aldolase gene contained in these clones, restriction fragments of the genomic clones were subcloned into pUC vectors. The location of these subclones is diagrammed in Fig. 1.

Determination of the Isozyme Identity of the Genomic Clone—In order to determine which aldolase isozyme is encoded by this gene, the tissue-specific pattern of expression of these clones was examined by hybridization to RNA from different tissues. Fig. 2 shows that liver RNA hybridizes to the subclone pBH (see Fig. 1) under conditions of high stringency. When the stringency conditions are lowered, RNA hybridizes from all three tissues examined (liver, muscle, and brain). The liver specific hybridization of this gene identifies it as encoding the aldolase B isozyme.

DNA Sequence Analysis—The strategy used to sequence the entire chicken aldolase B gene is shown in Fig. 3. The DNA sequence determined for the 5′ flanking and coding regions is shown in Fig. 4. The entire sequence, including the intron regions, is presented in the Miniprint Supplement (Fig. 1a).

The exon/intron structure of the gene was determined by comparing the amino acid sequence, as deduced from the nucleotide sequence, with the amino acid sequence for rabbit aldolase A. The abbreviations used are: bp, base pair; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; PIPES, 1,4-piperazinediethanesulfonic acid; kb, kilobase.
Characterization of the Chicken Aldolase B Gene

FIG. 1. Map of restriction endonuclease cleavage sites in λC11.1 and λC4 and location of subcloned fragments. a, location of fragments from λC11.1 and λC4 that were subcloned into pUC8 and pUC9 vectors. Abbreviations correspond to the following restriction endonuclease cleavage sites: E (EcoRI), H (HindIII), B (BamHI), P (PstI), and Bg (BglII). b, relative orientation of λC11.1 with respect to λC4. Black boxes represent sequences present in mRNA, c, map of restriction endonuclease cleavage sites. Vertical lines represent restriction sites.

Fig. 2. Northern hybridization of mRNA from liver, muscle, and brain with an aldolase B probe. Poly(A) RNA was separated on a 1% agarose gel, transferred to a nitrocellulose filter, and hybridized with the chicken aldolase B probe pH (see Fig. 1) at low stringency (30 °C, lanes 1 and 2; 37 °C, lanes 3 and 4; 60 °C, lanes 5 and 6) in 50% formamide, 0.75 M NaCl. Lanes 1, 3, 5, and 7, liver poly(A) RNA; lanes 2 and 4, 19-day-old embryo muscle poly(A) RNA; lanes 6 and 8, brain poly(A) RNA. The molecular weight markers are BglII pBR322 (2319 and 809 nucleotides) and HindIII pBR322 (1631 and 517 nucleotides). (Muscle from 19-day-old embryos synthesizes aldolase A, the adult form, rather than aldolase C, the embryonic form (51)).

aldolase A (2). This approach was feasible because of the high conservation of amino acid sequence between these two proteins (over 70%). The size determined for each exon was confirmed by S1 mapping (results not shown). The exact boundaries of the exons were identified according to the GT/AG rule of Brentnach et al. (25). In all cases the exon-intron junction sequences closely fit the consensus sequence derived by Mount (26).

Determination of the 3’t End of the Gene—The 3’t end of the aldolase gene was identified by S1 mapping analysis using liver poly(A) RNA to protect a radiolabeled M13 subclone, KE9, which includes the entire coding region of the last exon and 1100 additional nucleotides on the 3’t side. As shown in Fig. 5a, a fragment of 492–493 nucleotides was protected from S1 nuclease. Since the last three nucleotides of the previous exon are identical to the three nucleotides in the intron just upstream from this exon, the actual size of the last exon is only 489–490 base pairs. According to Fitzgerald and Shenk (27), polyadenylation usually initiates at an adenosine residue. Using this additional criterion, the most likely position for the 3’t end of this exon is an adenosine residue located 28 nucleotides downstream from the polyadenylation signal ‘AAUAAA’. A minor S1 protected fragment 503 nucleotides in length was also seen; this indicates that a minor poly(A) addition site is located 10 nucleotides downstream from the major polyadenylation site.

Determination of the 5’t End of the Gene—The size of the 5’t untranslated region was determined by primer extension analysis using synthetic oligodeoxynucleotide primers complementary to the RNA sequence near the 5’t end of the coding region (see Fig. 4). From the size of the extended products, the 5’t noncoding region was found to be 73 nucleotides in length (Fig. 6).

The size of the first coding exon was determined by S1 mapping analysis using radiolabeled M13 subclones (Fig. 5b) which included either the entire 112 nucleotides of coding region from this exon (EH3) or only the first 90 nucleotides (UG15); both clones also contained over 1.5 kb of DNA upstream from the coding region (see Fig. 7). In both cases, the result was consistent with the inclusion of only 10 nucleotides of 5’t noncoding region in the first coding exon. Since the 5’t noncoding region is 73 nucleotides long, the remaining 63 nucleotides of 5’t noncoding region must be in a separate exon. Consistent with this expectation, an intron-exon junction acceptor sequence was found at a position 10 nucleotides upstream from the coding region.

The first exon was roughly located to a region 2050–7600 nucleotides upstream from the second exon by the hybridization of the pUC subclone plio (see Fig. 1) to liver poly(A) RNA (Fig. 8a). A more precise location was determined by S1 mapping analysis of M13 recombinant clones containing a nested set of inserts which span the region 7.5 kb upstream from the first coding exon (see Fig. 7). Fig. 5c shows that a 61–68 nucleotide segment of the subclone AS8 was protected from digestion by S1 nuclease. AS8 contains an insert derived from a region 4600 nucleotides upstream from the first coding exon. An overlapping M13 subclone, BT12, did not contain this protected piece, which narrowed the location of the first exon to a 480-bp region. The region within this 480-bp region was analyzed and a TATA box (TATAAAA) was found, which is 70–75 nucleotides upstream from the first coding exon. Analysis of the sequence in this region also explained the bimodal distribution pattern seen for the protected pieces in Fig. 5c: the first six nucleotides in the intron region immediately adjacent to the putative first exon, GTAATG, closely match the first six nucleotides of exon II,
exon, and UG11, which contains the second exon. Fig. 8 shows that RNA from all three tissues hybridizes to both probes. In the liver message, a Northern blot of RNA from all three tissues was hybridized with POE1, which contains the first intron is diagrammed in Fig. 3.

To confirm the location of exon I, S1 mapping was performed on an M13 recombinant clone, RT11, which extends only as far as a PstI site 11 nucleotides downstream from the putative 5' end of the message. As expected, the protected piece was approximately 11 nucleotides shorter than the piece protected by AS8 (Fig. 5c).

To ensure that this exon was part of the aldolase gene, the M13 subclone POE1 was hybridized to a Northern blot of liver RNA. POE1 contains a 93-bp PstI-KpnI insert which includes all but the first 11 nucleotides of exon I. Fig. 6 shows that an RNA hybridizes which is the same size as the RNA hybridizing to UG11, an M13 recombinant clone containing only the second exon.

The 63-bp size of this exon, as determined by primer extension analysis, places the 5' end of the message at a guanine residue 28 bp downstream from the TATA box. A cluster of pyrimidines is present just 3' to this site, which is consistent with the consensus sequence for transcriptional start sites encoded by the first exon.

To ensure that no small restriction fragments were missed, sequencing was performed on an M13 recombinant clone, RT11, which extends from 143 to 837 bp. The position of each exon and intron was determined by S1 mapping to mRNA from different tissues, corresponds to the pattern expected for the B isozyme. Under high stringency conditions: the first product includes the first six nucleotides of the second exon. The coincidental correspondence between these two sequences allows for two possible S1 products under our experimental conditions: the first product includes the first six nucleotides and the second product, due to the 1-bp mismatch, ends right at the exon-intron boundary.

To determine whether the 5' noncoding regions are the same size, RNA from small intestine was analyzed by primer extension. Fig. 6 shows that the product is identical in size to the product synthesized off of liver RNA. To ensure that no additional exons are present, RNA from all three tissues was hybridized with the radiolabeled subclones p1n2 and p4a (see Fig. 1), which together span a 5700-bp region directly upstream from the area analyzed by S1 mapping. No additional transcripts were detected in either of these subclones (results not shown). Thus, no tissue-specific differences were detected in the structure of aldolase B mRNA in the region encoded by the first exon.

**DISCUSSION**

We have isolated and characterized overlapping genomic clones derived from a chicken aldolase B gene. Several pieces of evidence identify these genomic clones as encoding the B isozyme form of aldolase. First of all, the genomic clones could not encode the A isozyme form since the deduced amino acid sequence does not match the amino acid sequence of the NH₂-terminal end of chicken aldolase A (29). Second, the tissue-specific pattern of expression of this gene, as determined by hybridization to mRNA from different tissues, corresponds to the pattern expected for the B isozyme. Under high stringency conditions: the first product includes the first six nucleotides and the second product, due to the 1-bp mismatch, ends right at the exon-intron boundary.
Characterization of the Chicken Aldolase B Gene

Fig. 4. Nucleotide sequence of the 5' flanking and coding regions of the chicken aldolase B gene. Nucleotides 5' to the putative transcriptional initiation site are numbered above the sequence by negative numbers, beginning with +1 as the point of transcription initiation; the derived amino acid sequence is numbered from 1 to 360 with the initiating methionine denoted as +1. The number of nucleotides deleted from each intron is noted. Several features referred to under "Discussion" are underlined: CCAAT boxes at nucleotides -100, -131, and -162; repeats at -41 and -118; sequences resembling an adenosine enhancer sequence at -480 and -550; and a region of the 5' untranslated region which is 75% homologous to a sequence in the 5' untranslated region of the rabbit aldolase A gene. Sequences in exon II used to prepare complementary oligonucleotides for primer extension are indicated by wavy lines.
polyacrylamide/urea sequencing gels. A set of MIS-dideoxy sequenc-

tides) from exon 11. 

stream from the end of the coding region. The protected product was

3, 

run in running reactions was run as size markers.

EHB, which includes the entire protein coding region (112 nucleo-

acrylamide/urea gel. 

c, S1 mapping of the first exon.

exon 11. 

omitted 

pieces were run in a 10% polyacrylamide/urea gel.

The presence of a second aldolase B gene would require the

simultaneous conservation of BglII, SstI, and BamHI sites

located 4.5–12.6 kb downstream from the 3′ end of the aldolase B gene. Unless there has been a very recent gene duplication event, this is unlikely. Similar evidence in this lab and by Besmond et al. (30) suggests that there is only one aldolase B gene present in the human genome. In addition, this gene possesses all of the features associated with active genes: a TATA box (TATTTAAT at −28), a typical AUG translation initiation region (31), reasonable intron-exon junction sequences (26), a termination codon, and a polyadenylation signal (32). It also has the set of amino acids which have both been implicated in the reaction mechanism and have been found in all aldolases so far sequenced.

Coding Region—Our results confirm and extend previous reports indicating that aldolase is a highly conserved protein (2, 3). The deduced amino acid sequence of chicken aldolase B shares an overall homology of 81% with human aldolase B (3) and 71% with rabbit aldolase A (2). Since the approximate time of divergence of these pairs is known (250 million years for the mammalian/avian split and 500 million years for the isozyme split (33, 34)), it is possible to calculate the rate of evolution of aldolase. Using the correction factors given in Dayhoff et al. (35), aldolase is found to accept about 7 amino acid mutations/100 residues/100 million years. This places it with other highly conserved proteins such as gastrin and trypsinogen (33).

The amino acid sequence is particularly highly conserved in the active site region of the enzyme. There is 93% homology between the chicken aldolase B (3) and human aldolase B sequences in this region. Also conserved are a number of amino acids which have been previously implicated in the reaction mechanism of aldolase A (36) (Arg-55, Lys-107, Lys-146, the active site Lys-229, and Tyr-363). We previously reported (3) that three other amino acids which have been implicated in the reaction mechanism (36) (Cys-72 or -338 and His-361) are not present in the human aldolase B sequence. The chicken aldolase B sequence follows the human aldolase B pattern at two of these positions (Ser-72 and Tyr-361) but retains the cysteine seen in aldolase A at position 338.

The sequence of the carboxyl-terminal 11 amino acids follows an isozyme-specific pattern. While there is only 36% homology in this region between chicken aldolase B and rabbit aldolase A (37), the two B isozymes share 91% homology. Enzyme modification studies have also shown this region to be important in determining isozyme-specific differences in catalytic properties. Removal of the COOH-terminal tyrosine (38) converts aldolase A into an enzyme with catalytic properties similar to aldolase B (lower rate of cleavage of fructose 1,6-diphosphate compared to fructose-1-phosphate cleavage).

Intervening Sequences—By comparing the deduced amino acid sequence with the amino acid sequence for rabbit aldolase A (2), the coding region of the chicken aldolase B gene was found to be interrupted by seven intervening sequences. An eighth intervening sequence was found in the 5′ noncoding region through a combination of S1 mapping and primer extension experiments. The size of this intervening sequence, almost 4600 bp, is far larger than any of the other intervening sequences. A number of other genes have also been shown to have intervening sequences in their 5′ noncoding regions, including ovalbumin (25) and several polypeptide hormones (cf. Refs. 39 and 40). Many of these 5′ noncoding region introns are unusually large; for example, the intron in the bovine pituitary glycoprotein hormone α-subunit gene is greater than 13 kb in size (41).

In general, no special functions have been attributed to 5′

conditions, subclones derived from the genomic clones hy-

briedize to RNA from liver but not from brain or muscle. Lastly, liver poly(A) RNA contains an RNA population which fully protects both the 5′ and 3′ untranslated regions of this gene from digestion by S1 nuclease.

We believe that this gene is actively transcribed. It appears to be the only aldolase B gene present in the chicken genome on the basis of genomic Southern hybridization experiments. The presence of a second aldolase B gene would require the

Fig. 5. S1 mapping analysis of the 5′ and 3′ ends of the chicken aldolase B gene. Radiolabeled M13 recombinant clones were hybridized to 5 μg of liver poly(A) RNA and digested with S1 nuclease. The protected pieces were separated electrophoretically in polyacrylamide/urea sequencing gels. A set of M13-dideoxy sequencing reactions was run as size markers. a, S1 mapping of the exon containing the 3′ end using the M13 clone KE9, which contains a 1320-bp EcoRI fragment including the 1100-nucleotide region downstream from the end of the coding region. The protected product was run in a 5% polyacrylamide/urea gel. b, S1 mapping of the first coding exon (exon II). Lane 1, protection of UG15, which includes the first 90 nucleotides of protein coding region. Lane 3, protection of EH3, which includes the entire protein coding region (112 nucleo-
tides) from exon II. Lane 2, as in lane 3, but with poly(A) RNA omitted as a control. The protected pieces were run in an 8% polyacrylamide/urea gel. c, S1 mapping of the first exon. Lane 2, protection of AS8, a 693-bp KpnI fragment located 4550 bp upstream from exon II. Lane 1, same as lane 2, but with poly(A) RNA omitted. Lane 4, protection of RT11, a subclone which contains the last 93 bp of the same KpnI fragment which was cloned into AS8 (see Fig. 7). Lane 3, same as lane 4, but with poly(A) RNA omitted. The protected pieces were run in a 10% polyacrylamide/urea gel.

Fig. 6. Primer extension analysis of the 5′ end of the chicken aldolase B gene. 10 μg of poly(A) RNA were hybridized to the synthetic oligodeoxyribonucleotide GAGACACGTGCTGGG (lanes 2 and 4) and extended with reverse transcriptase. In lanes 1 and 3, 10 μg of poly(A) RNA were extended with reverse transcriptase in the absence of primer. Lanes 1 and 2, liver poly(A) RNA. Lanes 3 and 4, small intestine poly(A) RNA.
Characterization of the Chicken Aldolase B Gene

FIG. 7. The nested set of M13 recombinant clones used to localize exons I and II by S1 mapping analysis. The location of each insert is shown in reference to a partial restriction map of the region analyzed. The last column indicates which exon, if any, was protected.

FIG. 8. Hybridization of probes from AC11.1 subclones to chicken poly(A) RNA. 1 µg of poly(A) RNA was separated in a 1% agarose gel, transferred to a nitrocellulose filter, and hybridized with radiolabeled AC11.1 subclones at 50°C in 50% formamide, 0.75 M NaCl. a, poly(A) liver RNA was hybridized with p1io (lane 1) and p1do (lane 2). The subclone p1do has a 2240-bp BamHI-EcoRI insert which contains the second exon (see Fig. 1). The subclone p1io has a 5450-bp EcoRI-BamHI insert which covers the region directly upstream from the 2240-bp BamHI-EcoRI insert in p1do (see Fig. 1). b, poly(A) RNA from liver (lanes 1 and 4), small intestine (lanes 2 and 5), and kidney (lanes 3 and 6) was hybridized with the M13 recombinant subclones UG11 (lanes 1–3) and POEl (lanes 4–6). UG11 has a 540-bp Sau3AI-EcoRI insert which contains exon 2. POEl has a 9.7-kb PstI-KpnI insert which includes 84% of exon I.

noncoding region introns. However, the mouse α-amyIase gene has been shown to use alternate promoters in a tissue-specific fashion in association with differential splicing of its 5′ noncoding region intron (42). Since aldolase B is expressed in several tissues (liver, kidney, and small intestine), we were interested in determining whether a similar regulatory mechanism might operate for the chicken aldolase B gene. Tsutsumi et al. (43) have recently shown that rat aldolase A mRNA shows tissue-specific heterogeneity in size, presumably due to heterogeneity in the 5′ or 3′ noncoding region. No such tissue-specific differences were found for chicken aldolase B mRNA.

FIG. 9. Hybridization of a chicken aldolase B probe to restriction digests of chicken genomic DNA. Chicken genomic DNA was digested with BglII (lane 1), SstI (lane 2), or BamHI (lane 3), size-fractionated in a 0.8% agarose gel, and transferred to a nitrocellulose filter. The Southern genomic blot was hybridized with the radiolabeled subclone p0a (see Fig. 1) at 50°C in 50% formamide, 0.75 M NaCl.

Both message size and length of the 5′ noncoding region are identical and the 5′ noncoding region exon used in the liver is also expressed in both kidney and small intestine tissues.

5′ Flanking Region—Since the 5′ flanking region appears to be important in controlling the developmental- and tissue-specific regulation of gene expression (cf. Ref. 44), we searched this region of the aldolase B gene for features with possible biological significance. A CCAAT box consensus sequence has been reported to be present in the −70 to −100 region of many genes, including all globin genes (45). A variation of the CCAAT box, CCAAC, was found at −100 and two other CCAAT boxes were found further upstream (at −131 and −162) (see Fig. 4). Two of these CCAAT boxes are surrounded
by short inverted repeats:

ACAGCACA-6 bp-CCCAT-12 bp-TTGTCTGT

and

CTGGGC-12 bp-CCACAC-2 bp-GCCGAG

Inverted and direct repeats are often found associated with promoter regions (cf. Ref. 46). Presumably the binding of regulatory molecules at two positions can produce a stronger signal than binding at one site alone. An example of an imperfect repeat was found at -118 and -41 (CTGGGCAAGGGA and CTGGGCAAGGA).

Upstream regions of several cellular genes have been shown to have enhancer-like properties. For instance, the 5' flanking regions of chymotrypsin and insulin genes show tissue-specific properties which behave in an orientation-independent manner (44). In the chicken aldolase B gene a sequence is present at -550 (see Fig. 4) which matches the adenovirus enhancer consensus sequence, [GGAAGTGG] (47). On a random basis, this sequence would be expected to occur in each strand only once in every 280,000 nucleotides. A second sequence was found nearby which matches the consensus sequence with the exception of one deleted base (AGGACTGAC at -480). Determination of the biological significance of these sequences awaits further studies.

5' Noncoding Region—Although the 5' noncoding region is not generally conserved, even between closely related genes (48), the chicken aldolase B gene shares 75% homology with a 24-nucleotide region in the 5' noncoding region of the rabbit aldolase A message (2).

CBB: GCAGCAGGGCTGCTCCAGAAC/ATAAGTACC

** * ** * * * * *

RDA: GCCGCAAAGACTCTGCTCCAGACCC A T C

It is interesting that the juxtaposition of these two sequences starts immediately 5' to the intervening sequence (slash) in the aldolase B gene and at the beginning of the 5' noncoding region in the rabbit aldolase A sequence (RBA). (The translation initiation codons have been underlined.) It is not known whether the aldolase A gene has an intervening sequence within its 5' noncoding region.

3' Noncoding Region.—A microheterogeneity in the length of the 3' noncoding region was found from S1 mapping experiments. A minor polyadenylation site occurs 10 nucleotides downstream from the major poly(A) addition site. This second site may be related to the existence of two polyadenylation signals in this region: AAUAAA and 11 nucleotides further downstream, the variant AAUUAAA (49). A small hairpin loop with a stability of -8.76 kcal/mol at 25 °C surrounds the first polyadenylation signal. Since not all polyadenylation signals are actually utilized (50) this hairpin loop could serve as a secondary recognition signal.

We previously compared the 3' noncoding regions from human and rat aldolase B genes (3) and found two 55-bp conserved regions near the 3' end. It is interesting that neither of these conserved regions are found in the chicken aldolase B gene. The first conserved region has either been deleted from the chicken gene or been inserted into the mammalian genes as sequences on either side of this conserved region are homologous to sequences which are contiguous in the chicken gene. The second conserved region, at the very end of the 3' noncoding region, is nonhomologous to the same region in the chicken gene. If these conserved regions have a function in the mammalian genes, the function may have been attained after the chicken and mammalian lineages diverged.

Acknowledgments—We are grateful to Drs. Juan C. Mendible for his help in the initial stages of this work, Mickey S. Urdea for his synthesis of oligonucleotide primers, James D. Engel for providing us with a chicken genomic library, and Hugo Martinez of the University of California at San Francisco Biomathematics Computation Laboratory for the use of his DNA sequence analysis programs. We appreciate many helpful discussions with Dr. Dean Tolan, Bruce Amstler, and William Rottmann.

REFERENCES

1. Penhoet, E., Kochman, M., Valentine, R., and Rutter, W. J. (1967) Biochemistry 6, 2940-2949

2. Tolan, D. R., Amstler, A. B., Putney, S. D., Urdea, M. S., and Penhoet, E. E. (1984) J. Biol. Chem. 259, 1127-1131

3. Rottmann, W. H., Tolan, D. R., and Penhoet, E. E. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 2738-2742

4. Penhoet, E. E., Kochman, M., and Rutter, W. J. (1969) Biochemistry 8, 4396-4409

5. Rutter, W. J., Rajkumar, T., Penhoet, E. E., Kochman, M., and Valentine, R. (1966) Ann. N. Y. Acad. Sci. 151, 102-117

6. Dodgen, J. B., Streamer, J., and Engel, J. D. (1979) Cell 17, 879-887

7. Benton, W. D., and Davis, R. W. (1977) Science (Wash. D. C.) 196, 185-182

8. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 257-285

9. Yamamoto, K. R., Alberts, B. M., Bentinger, R., Lawlorn, L., and Treiber, G. (1970) Virology 40, 734-744

10. Davis, R. W., Botstein, D., and Roth, J. R. (1980) Advanced Bacterial Genetics, pp. 106-107, Cold Spring Harbor Laboratorv, Cold Spring Harbor, NY

11. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517

12. Vieira, J., and Messing, J. (1972) Gene (Amst.) 19, 259-268

13. Gassm, H. C. (1983) Methods Enzymol. 100, 243-255

14. Godson, G. N., and Vapnek, D. (1973) Biochem. Biophys. Acta 299, 516-520

15. Hu, N., and Messing, J. (1982) Gene (Amst.) 17, 271-277

16. Messing, J., and Vieira, J. (1982) Gene (Amst.) 19, 269-276

17. Norrander, J., Kempe, T., and Messing, J. (1983) Gene (Amst.) 28, 101-106

18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467

19. Chirigwe, J. M., Pribulya, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299

20. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412

21. Rave, N., Crkvenjakov, R., and Boedtcher, H. (1979) Nucleic Acids Res. 6, 3555-3567

22. Thomas, P. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5201-5205

23. Favaloro, J., Treisman, R., and Kammen, R. (1980) Methods Enzymol. 65, 718-749

24. Robbins, J., Rosteck, P., Jr., Haynes, J. R., Freger, G., Cleary, M. L., Kalter, H. D., Smith, K., and Langel, J. B. (1979) J. Biochem. Chem. 254, 6187-6192

25. Breathnach, R., Benoist, C., O'Hare, K., and Chanbon, P. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 4853-4857

26. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472

27. Fitzgerald, M., and Shenk, T. (1981) Cell 24, 251-260

28. Breathnach, R., and Chanbon, P. (1981) Annu. Rev. Biochem. 50, 349-383

29. Lebherz, H. G., Bates, O. J., and Bradshaw, R. A. (1984) J. Biol. Chem. 259, 1132-1135

30. Bescond, M., Dreyfus, J.-C., Gregori, C., Frain, M., Zakin, M. M., Sala Treapat, J., and Kahn, A. (1983) Biochem. Biophys. Res. Commun. 117, 601-609

31. Kozak, M. (1984) Nucleic Acids Res. 12, 857-872

32. Proudfound, N. J., and Brownwee, G. G. (1976) Nature (London) 263, 211-214

33. McLaughlin, P. J., and Dayhoff, M. O. (1972) in Atlas of Protein Sequence and Structure 1972 (Dayhoff, M. O., ed) pp. 47-52, Natl. Biomedical Research Foundation, Washington, D. C.

34. Lebherz, H. G., and Rutter, W. J. (1969) Biochemistry 8, 199-211

35. Dayhoff, M. O., Ech, R. V., and Park, C. M. (1972) in Atlas of Protein Sequence and Structure 1972 (Dayhoff, M. O., ed) pp. 95, Natl. Biomedical Research Foundation, Washington, D. C.
Characterization of the Chicken Aldolase B Gene

36. Lai, C. Y., Nakai, N., and Chang, D. (1974) Science (Wash. D. C.) 183, 1204-1206
37. Freemont, P. S., Dunbar, B., and Fothergill, L. A. (1984) Arch. Biochem. Biophys. 228, 342-352
38. Drechsler, E. R., Boyer, P. D., and Kowalsky, A. G. (1959) J. Biol. Chem. 234, 2627-2634
39. Cordell, B., Bell, G., Tischer, E., DeNoto, F. M., Ullrich, A., Picter, R., Rutter, W. J., and Goodman, H. M. (1979) Cell 18, 533-543
40. Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kakidani, H., Jingami, H., and Numa, S. (1980) Nature (Lond.) 287, 752-755
41. Goodwin, R. G., Moncman, C. L., Rottman, F. M., and Nilson, J. H. (1983) Nucleic Acids Res. 11, 6873-6882
42. Young, R. A., Hagenbüchle, O., and Schibler, U. (1981) Cell 23, 451-458
43. Tsutsumi, R., Tsutsumi, K., Numazaki, M., and Ishikawa, K. (1984) Eur. J. Biochem. 142, 161-164
44. Walker, M. D., Edlund, T., Boulet, A. M., and Rutter, W. J. (1983) Nature (Lond.) 306, 557-561
45. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Barralle, F. E., Shouiders, C. C., and Proudfoot, N. J. (1980) Cell 21, 653-668
46. McKnight, S. L., Kingsbury, R. C., Spence, A., and Smith, M. (1984) Cell 37, 253-262
47. Hearing, P., and Shenk, T. (1983) Cell 33, 695-703
48. Kozak, M. (1981) Curr. Top. Microbiol. Immunol. 93, 81-123
49. Beget, S. M. (1984) Nature (Lond.) 309, 179-182
50. Moore, C. L., and Sharp, P. A. (1984) Cell 34, 581-591
51. Lebherz, H. G. (1975) in Isoenzymes (Markert, C. L., ed) Vol. III, pp. 253-279, Academic Press, San Francisco
Characterization of the Chicken Aldolase B Gene

Figure 1a. The complete nucleotide sequence of the chicken aldolase B gene, including intervening sequences and 3 flanking region sequences. The notation is the same as in figure 2, with the addition of numbers on the right of the figure corresponding to nucleotide position, beginning with the first nucleotide in the sequence at 1.