Plant Physiol. (1986) 82, 1076–1080
0032-0889/86/82/1076/05/$01.00/0

The Complete Amino Acid Sequence for the Anaerobically Induced Aldolase from Maize Derived from cDNA Clones[1]

Received for publication July 29, 1986 and in revised form September 11, 1986

PHILIP M. KELLEY[2] AND DEAN R. TOLAN*
Department of Biochemistry, University of California, Berkeley, California 94720

ABSTRACT

A cDNA library was synthesized from maize anaerobic root mRNA and screened with cDNA specific to the anaerobically induced *Zea mays* cytoplasmic aldolase. At least 1% of the cDNA of the library corresponded to maize cytoplasmic aldolase. The sequence of four overlapping cDNA clones encoded a protein of molecular weight 38,611 homologous to aldolase. These cDNAs were polymorphic at three bases and one of these cDNAs had a different, shorter 3′-untranslated region. No known eukaryotic poly(A) addition site was detected. The derived amino acid sequences of maize was compared to the sequence of aldolase of trypanosome, *Drosophila*, and two mammalian isozymes, A and B. Of these, maize cytoplasmic aldolase was found to have the highest homology (55%) with rabbit aldolase A.

Anaerobiosis of maize seedlings results in the selective synthesis of cytoplasmic aldolase and at least 19 other proteins, including two alcohol dehydrogenase isozymes, pyruvate decarboxylase, glucose phosphate isomerase and sucrose synthase (C Bennett, personal communication) (9, 12, 13, 15, 23). This selective synthesis of the anaerobic proteins of maize is the result of the selective translation of mRNA coding for the anaerobic proteins (24) and the accumulation of anaerobic specific mRNA (9). Among a set of cDNA clones shown to be anaerobic specific, a 160 bp[3] cDNA, pZMX71 was found to hybrid select mRNA encoding a protein of approximately 40,000 mol wt that was selectively precipitated by an antiserum specific for maize cytoplasmic aldolase (9).

We have synthesized a maize anaerobic cDNA library and purified a set of aldolase specific cDNAs. We have determined the nucleotide sequence of these cDNAs and a derived amino acid sequence corresponding to a protein of mol wt of 38,611. We have compared the sequence of this maize aldolase to several vertebrate and invertebrate aldolases and found significant overall homology, 55% with rabbit aldolase A. Specific regions of the protein, such as the active site, showed much higher homology.

MATERIALS AND METHODS

Preparation and Characterization of mRNA. Approximately 2000 seeds (*Zea mays* Berkeley Fast) were soaked 8 h in distilled H_2O at 23°C, then germinated on moist paper towels at 27°C for 4 d. Seedlings were collected and submerged 20 h in 15 L of distilled H_2O. Primary roots were cut off and frozen in liquid N_2. Frozen roots were used immediately or stored at −80°C.

mRNA was isolated by the method of Chirgwin *et al.* (4) with the following modifications. Thirty g of frozen roots (from 2000 seedlings) were pulverized to a fine powder with a mortar and pestle and suspended in 150 ml of freshly prepared: 6 M guanidine thiocyanate, 5 mM sodium citrate (pH 7.0), 0.14 M 2-mercaptoethanol, and 0.5% (w/v) sarcosyl. This solution was mixed 5 min at 50% normal speed in a Waring Blendor. The resulting slurry was filtered through four layers of cheese cloth then centrifuged at 6,000 rpm in a Sorvall SS-34 rotor for 5 min at 4°C. One g of CsCl was added for each 2.5 ml of the supernatant fraction and layered onto a 10 ml cushion of 5.7 M CsCl, 0.1 M EDTA. RNA was pelleted by centrifugation for 12 h at 25,000 rpm in a Beckman SW 27.1 rotor at 20°C. The supernatant fraction was discarded and the RNA pellet was dissolved in 500 μl distilled H_2O. The RNA was made 0.25 M with sodium acetate and precipitated at −80°C with 2.5 volumes of ethanol. Precipitated RNA was solubilized in distilled H_2O and was further purified by chromatography on oligo-dT cellulose, as described in Maniatis *et al.* (18) with the exception that SDS was not used.

Construction of the cDNA Library. All steps involving recombinant DNA were done according to NIH directive. A cDNA library was constructed using methods described by Maniatis *et al.* (18). The dC-tailed cDNA was fractionated over a Sepharose CL-4B column (0.5 × 7 cm) equilibrated in annealing buffer. Fractions of about 250 μl were taken and annealed with Pst I restricted and dG-tailed pUC8 (28) in equimolar ratios to a final concentration of 1 ng/μl. Annealed DNA was stored at −20°C. Annealed cDNA (2 μg) was used to transform 0.3 ml of *Escherichia coli* strain JM83 or JM103 competent cells and plated on LB plates containing 35 μg/ml isopropyl thiogalactoside and 0.0033% (w/v) 5-bromo-4-chloro-3-indolyl galactoside. Plasmid DNA was purified from eight randomly chosen transformants derived from individual Sepharose fractions. The library derived from the leading fraction with an average insert size of 1100 bp was used to isolate aldolase cDNA clones.

Isolation of cDNA Clones. The cDNA library was plated on nitrocellulose filters (Schleicher and Schuell, BA85) at a density of about 1000 per plate. Two copies of the original were made. These two copies were baked, the filters were then soaked in 6×SSC (20×SSC is 3 M NaCl, 0.3 M sodium citrate adjusted to pH 7.0) for 5 min, then washed 2 h at 42°C in 50 mM Tris-Cl (pH 8.0), 1 M NaCl, and 0.1% (w/v) SDS. The washed filters were prehybridized 12 h at 65°C with 50 mM sodium phosphate buffer, 5× Denhardts, 1 M NaCl, 0.2% (w/v) SDS, 1 mM EDTA, 140 μg/ml salmon sperm DNA, 50 μg/ml poly(A), 50 μg/ml poly(C), and 6.25 μg/ml sonicated pUC8. Gel purified insert DNA was nick translated, heat denatured at 100°C for 10 min, then added to the prehybridization mix and incubated with filters

[1] Supported by grant GM-32344 from the National Institutes of Health.

[2] Present address: School of Biological Sciences, University of Nebraska-Lincoln, 348 Manter Hall, Lincoln, NE 68588-0118; phone 402-472-1683.

[3] Abbreviations: bp, base pair; *Adh1*, alcohol dehydrogenase gene 1; b, base; ds, double stranded; ss, single stranded.

for 12 h at 65°C. These filters were then washed for 10 min three times in 2×SSC + 0.1% (w/v) SDS at room temperature and 1 h three times in 2×SSC + 0.1% (w/v) SDS at 65°C. The filters were radioautographed, colonies were located on the original filter and rescreened. Plasmid DNA was isolated as previously described (7). The library was also screened using a radiolabeled deoxyribo-oligonucleotide specific for maize aldolase under the same conditions as above except incubations were at 50°C and washes were at 42°C in 0.2×SSC.

Sequence Analysis. Purified insert from cDNA clones was digested with *Pst* I, *Taq* I, *Alu* I, *Hae* III, *Kpn* I, or *Sst* I and subcloned into M13mp8 and/or M13mp9. Alternatively, cDNA clones were digested with restriction endonucleases and individual restriction fragments were gel purified before cloning into M13 vectors. pZM1085 was sequenced by the method of Sanger *et al.* (25) by subcloning *Taq* I, *Alu* I, and *Hae* III total digests into M13mp8. pZM1154 was sequenced by subcloning a *Kpn* I/*Pst* I digest into M13mp18 or M13mp19 and by internal priming using synthetic deoxynucleotides as primers. The sequence for pZM205 was obtained by subcloning restriction fragments obtained from *Sst* I and *Kpn* I/*Pst* I restriction digestions and by internal priming. The location of the primers and their sequence is shown in Figure 2. Certain compressed regions of the sequence were resolved by substitution of 7-deaza-dGTP for the dGTP in the reaction mixtures. This reagent precludes certain intrastrand base pair interaction (26). The cDNA insert from pZMX71 was subcloned into *Pst* I cut pUC8, then cut with *EcoR* I and *Hind* III and the fragment containing the cDNA insert was sequenced using the method of Maxam and Gilbert (20). Computer analysis of DNA and protein sequences was accomplished using the programs of H. Martinez (University of California, San Francisco) with the University of California, Berkeley VAX/UNIX system.

Amino Acid Composition. Maize cytoplasmic aldolase was purified from a Black Mexican Sweet suspension cell line as described previously (13). Protein was hydrolyzed in 5 N HCl for 1 h at 150°C, then derivatized with 3-phenyl-2-thiohydantoin and separated by HPLC (27). The protein was unreactive during Edman degradation, indicating that the amino terminal residue was blocked.

N-Terminus of CNBr Peptides. Maize aldolase was reduced and alkylated with 4-vinylpyridine, the carboxyl groups were amidated with dimethylethylenediamine, and the protein was cleaved with cyanogen bromide. The resulting mixture of peptides was subjected to N-terminal sequence analysis by five cycles of Edman degradation (27).

RESULTS

Synthesis and Characterization of a Maize Seedling Root Anaerobic Library in pUC8. A cDNA library was prepared from anaerobic maize seedling root mRNA. The construction of this library is summarized in Table I. The library contained 80,000 independent transformants derived from 500 ng of ds cDNA. An input of 10 μg of mRNA yielded 0.5 μg of high mol wt dC-tailed dsDNA for an efficiency of 16,000 colonies/μg of input mRNA. Seven clones were purified which hybridized to the maize cytoplasmic adolase cDNA pZMX71. Two of these clones, pZM1085 and pZM1154, were selected for further characterization. A deoxyoligonucleotide was synthesized based on sequence information obtained from pZM1085 (Fig. 1). The deoxyoligonucleotide corresponding to the 5′-most sequence was radiolabeled and used to screen the cDNA library. It hybridized to approximately 1% of the cDNA clones. The largest clone obtained from this screen, pZM205, was selected for further characterization.

Nucleotide Sequence Determination of a Maize Aldolase cDNA. The entire sequence of four cDNA clones was determined as shown in Figure 1. The composite nucleic acid sequence derived from pZMX71, pZM1085, pZM1154, and pZM205, and the derived amino acid sequence of the maize cytoplasmic adolase is shown in Figure 2. The derived amino acid sequence corresponded to a protein of 355 amino acids with a mol wt of 38,611. Polymorphisms were determined at three positions in the sequence, twice in the coding region and once in the 3′-noncoding region. pZM1154 ended with a different sequence, $A_{10}TTA_{19}$. The composite sequence accounted for 1406 bases. The 3′-noncoding region terminated with poly(A) and was found in two sizes, 299 and 233 nucleotides long. pZMX71 was found to be 160 bp long and corresponds to amino acid 174 to 226 of maize aldolase.

Table I. *Synthesis of Maize Anaerobic cDNA Library*

cDNA synthesis	Yield	Yield
	μg	%
Input maize anaerobic root poly(A) mRNA	10.0, mRNA	100
First strand synthesis	4.2, ssDNA	42
Second strand synthesis	4.8, dsDNA	24
S1-nuclease treatment	2.4, dsDNA	12
C-tailing		
Sepharose CL-4B	2.2, dsDNA	11
High mol wt cDNA annealed with G-tailed pUC8	0.5, dsDNA	2.5
Transformants obtained	80,000 colonies	
Efficiency	16,000 colonies/μg of input mRNA	
Average insert size	1,100 bp	

Amino Acid Composition of Cytoplasmic Maize Aldolase. A total amino acid composition was determined and is shown in Table II. The amino acid composition of the purified maize cytoplasmic aldolase is consistent with the composition predicted from the derived amino acid sequence. N-terminal analysis by Edman degradation of purified maize cytoplasmic aldolase indicated a blocked N-terminus. If the cDNA sequence encoded the N-terminal amino acid there would be only three peptides generated from CNBr cleavage. Subsequent N-terminal analysis would yield only 2 mol amino acids released/mol of protein. If the cDNA did not encode the N-terminus, then 3 mol of amino acids would be released. The maize aldolase protein was cleaved with CNBr, and N-terminal Edman degradation was performed. Each of five cycles yielded 2±0.3 pmol/pmol of protein. The predominant amino acids released in each cycle were consistent with the predicted amino acid sequence from the two internal CNBr peptides. By inference, the N-terminus must be the Met-Ser-Ala. pZM1154 was the only cDNA clone which encoded the N-terminal methionine. In addition, this clone had 21 nucleotides of the 5′-untranslated region.

DISCUSSION

We have synthesized a maize anaerobic cDNA library in the vector pUC8 containing 80,000 transformants. About 1% of these clones specifically hybridize to a cDNA clone corresponding to a maize anaerobic cytoplasmic aldolase. We have sequenced four overlapping cDNA clones which reveal an open reading frame corresponding to a 38,611 D protein of 355 amino acids. The N-terminus amino acid was deduced from the resultant N-terminal amino acids released CNBr peptides from purified aldolase protein.

Identification of Maize Cytoplasmic Aldolase cDNA. The cDNA clones were identified as sequences encoding maize aldolase because of the complete homology in nucleic acid sequence to the previously characterized clone pZMX71 (9) and

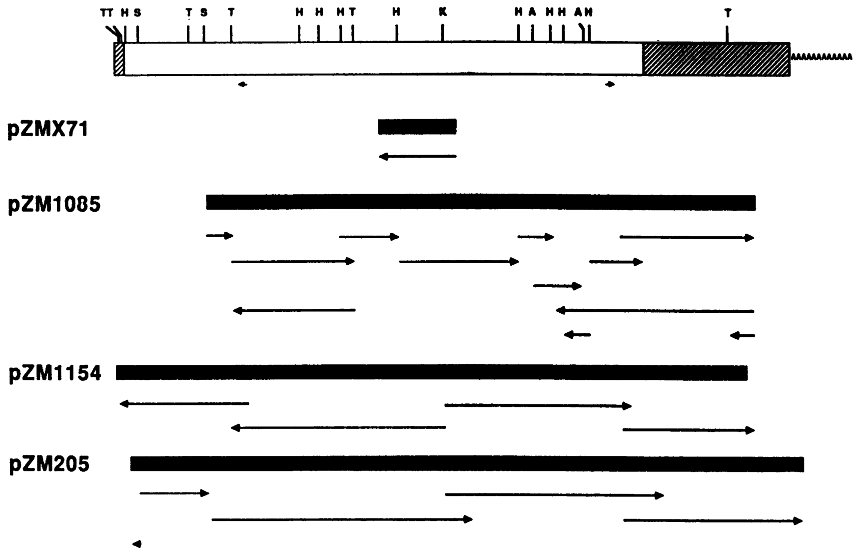

FIG. 1. Restriction map and sequencing strategy for maize cytoplasmic aldolase. The map shows in schematic form the physical structure of the mRNA with relevant restriction sites: T, *Taq* I; H, *Hae* III; S, *Sst* I; and K, *Kpn* I. The small arrows underneath the restriction map indicate the location of sequences complementary to the oligonucleotide primers. The overlapping cDNA clones are indicated by solid black bars. M13 subclones are shown with an arrow indicating the direction and extent of sequencing (78% of the sequence is in both directions).

```
                           1                                   10                                      20
                           Met Ser Ala Tyr Cys Gly Lys Tyr Lys Asp Glu Leu Ile Lys Asn Ala Ala Tyr Ile Gly Thr Pro Gly Lys
CCGUCGCGUCGAGCAUCGAUC      AUG UCG GCC UAC UGC GGA AAG UAC AAG GAU GAG CUC AUC AAG AAU GCU GCC UAC AUU GGC ACC CCU GGC AAG   (93)
1154                                   205
                      30                                      40                                      50
Gly Ile Leu Ala Ala Asp Glu Ser Thr Gly Thr Ile Gly Lys Arg Leu Ser Ser Ile Asn Val Glu Asn Val Glu Glu Asn Arg Arg Ala
GGU AUC CUU GCU GCU GAU GAG UCC ACU GGC ACC AUU GGC AAG CGC CUU UCC AGC AUC AAU GUC GAG AA[C] GUU GAG GAG AAC CGC CGU GCC   (183)

                      60                                      70                                      80
Leu Arg Glu Leu Leu Phe Cys Cys Pro Gly Ala Leu Gln Tyr Ile Ser Gly Val Ile Leu Phe Glu Glu Thr Leu Tyr Gln Lys Thr Lys
CUC CGU GAG CUC CUA UUC UGC UGC CCU GGU GCU CUC CAG UAC AUC AGC GGU GUG AUC CUC UUC GAG GAG ACC CUG UAC CAG AAG ACC AAG   (273)
                      1085
                      90                                      100                                     110
Asp Gly Lys Pro Phe Val Asp Val Leu Lys Glu Gly Gly Val Leu Pro Gly Ile Lys Val Asp Lys Gly Thr Ile Glu Val Val Gly Thr
GAU GGC AAG CCU UUU GUU GAU GUC CUC AAG GAG GGA GGC GUC CUC CCU GGC AUC AAG GUU GAC AAG GGC ACC AUU GAG GUU GUU GGC ACU   (363)

                      120                                     130                                     140
Asp Lys Glu Thr Thr Thr Gln Gly His Asp Asp Leu Gly Lys Arg Cys Ala Lys Tyr Tyr Glu Ala Gly Ala Arg Phe Ala Lys Trp Arg
GAU AAG GAG ACC ACC ACC CAA GGC CAU GAC GAC CUU GGC AAG CGC UGC GCC AAG UAC UAC GAG GCC GGU GCC CGC UUU GCC AAG UGG CGC   (453)

                      150                                     160                                     170
Ala Val Leu Lys Ile Gly Pro Asn Glu Pro Ser Gln Leu Ala Ile Asp Leu Asn Ala Gln Gly Leu Ala Arg Tyr Ala Ile Ile Cys Gln
GCU GUU CUC AAG AUU GGC CCC AAU GAG CCA UCA CAG CUU GCC AUC GAC CUG AAC GCU CAG GGU CUG GCU CGC UAU GCC AUC AUC UGC CAG   (543)
                                                                                                                X71
                      180                                     190                                     200
Glu Asn Gly Leu Val Pro Ile Val Glu Pro Glu Ile Leu Val Asp Gly Pro His Asp Ile Asp Arg Cys Ala Tyr Val Thr Glu Thr Val
GAG AAU GGU CUG GUG CCA AUU GUU GAG CCU GAG AUC CUU GUU GAU GGC CCU CAU GAC AUU GAU CGC UGC GCU UAC GUC ACU GAG ACC GUC   (633)

                      210                                     220                                     230
Leu Ala Ala Cys Tyr Lys Ala Leu Asn Glu His His Val Leu Leu Glu Gly Thr Leu Leu Lys Pro Asn Met Val Thr Pro Gly Ser Asp
CUU GCU GCC UGC UAC AAG GCG CUC AAC GAG CAC CAU GUC CUC CUG GAG GGU ACC CUC CUG AAG CCC AAC AUG GUG ACU CCA GGC UCC GAC   (723)
                                                                                    X71
                      240                                     250                                     260
Ser Lys Lys Val Thr Pro Glu Val Ile Ala Glu Tyr Thr Val Arg Thr Leu Gln Arg Thr Val Pro Ala Ala Val Pro Ala Val Leu Phe
UCC AAG AAG GUG ACU CCU GAG GUG AUU GCU GAG UAC ACC GUC CGU ACC CUC CAG AGG ACC GUC CCU GCU GCU GUG CCU GCU GUU [C]UC UUC   (813)

                      270                                     280                                     290
Leu Ser Gly Gly Gln Ser Glu Glu Glu Ala Thr Arg Asn Leu Asn Ala Met Asn Lys Leu Ser Thr Lys Lys Pro Trp Ser Leu Ser Phe
CUC UCU GGU GGA CAG AGC GAG GAG GAG GCC ACC CGC AAC CUC AAU GCC AUG AAC AAG CUC AGC ACC AAG AAG CCG UGG UCC CUG UCU UUC   (903)

                      300                                     310                                     320
Ser Phe Gly Arg Ala Leu Gln Ala Ser Thr Leu Lys Ala Trp Ala Gly Lys Val Glu Asn Leu Glu Lys Ala Arg Ala Ala Phe Leu Ala
UCC UUC GGC CGU GCC CUC CAG GCG AGC ACC CUC AAG GCC UGG GCU GGC AAG GUG GAG AAC UUG GAG AAG GCU AGA GCU GCC UUC CUC GCC   (993)

                      330                                     340                                     350
Arg Cys Lys Ala Asn Ser Glu Ala Thr Leu Gly Thr Tyr Lys Gly Asp Ala Ala Ala Asp Thr Glu Ser Leu His Val Lys Asp Tyr Lys
AGG UGC AAG GCC AAC UCU GAG GCU ACC CUC GGC ACC UAC AAG GGU GAU GCU GCC GCC GAC ACC GAG AGC CUC CAC GUC AAG GAC UAC AAG  (1083)

355
Tyr OP
UAC UGA GUUGCUUCUCUUUGCGUGCCCGUUUUGGAACUCCUGCCUACGAGCUAGGUUGCGGAGUGUGGUU[U]GCUCGUUAGAGCUUAUUUUCGGACAUCCAAAAAACAUGCUUCCUAU  (1200)
GUGGGAUGAUGGUUAUCUUUAUAUUUGUAUGUUAAUUAGUCUCUUUGCUGUUAAAUUUCGUGUAAGUUGGUCCUGCCGAUGGAGAAUCGAGCAGCCCCCUUUUUUUGUUCUAUCAACUA  (1319)
                                                                                                              1154
UGCUGUAAUGCUGGCUAUGUAUCGGCAAAAACAAUUCUAUAUGCUGAGUUGGAGUCGGCAAAAAUUCAUAAAAAAAAAAAAAAAAAA                                 (1406)
                1085                                                             205
```

FIG. 2. Nucleotide and amino acid sequence of maize cytoplasmic aldolase. A composite nucleotide sequence assembled from the sequence of pZM1154 (from the 5′-end to position 31) and pZM205. The derived amino acid squence is shown above each triplet. Boxed nucleotides indicate positions where polymorphism was found. Sequence region corresponding to the oligonucleotide primers is underlined with arrows. The beginning and end of each cDNA clone is denoted below the nucleotide sequence by vertical arrows and the clone number.

Table II. *Amino Acid Composition of Maize Cytoplasmic Aldolase*

The residues found are expressed as number of residues per molecular weight of 36,592 D. The number of residues expected is derived from the sequence of the cDNAs.

	Found	Expected
Alanine	37	39
Asp + Asn	30	32
Glu + Gln	32	35
Phenylalanine	8	8
Glycine	33	29
Histidine	4	5
Isoleucine	16	17
Lysine	27	30
Leucine	34	36
Methionine	2	3
Proline	15	15
Arginine	18	15
Serine	20	17
Threonine	24	24
Half Cystine	ND	8
Valine	25	25
Tryptophan	ND	3
Tyrosine	15	14

considerable homology to other known aldolases (3, 21, 29). Furthermore, the amino acid sequence that was derived from the cDNA sequence successfully predicted the amino acid composition of the maize cytoplasmic aldolase (Table II).

Plants have two forms of aldolase, cytoplasmic, and plastid. The sequence presented in the paper is the cytoplasmic form from several criteria. Antisera specific for maize cytoplasmic aldolase (13) specifically recognize the hybrid release translation product of pZMX71 (9) which is 100% homologous in nucleic acid sequences. In both maize and spinach, the cytoplasmic form of aldolase has a blocked N-terminus (this paper and Ref. 16), whereas the plastid form of aldolase in spinach has a free N-terminus. The N-terminal protein sequence of the plastid form has been determined (16) and the sequence presented here is distinct from this sequence. The induction by anaerobiosis of mRNA which hybridizes specifically to pZMX71 has been demonstrated by Hake *et al.* (9). The complete homology of the cDNA clones pZM1085, pZM1154, and pZM205 to pZMX71 further indicate that they also correspond to the anaerobically induced form of maize aldolase. There may be differences between this mRNA and the aerobic form to account for the transcriptional variation during anaerobiosis (9, 13) and some form of transcriptional control which may also be operating (24). One such difference has been described in the *Drosophila* heat shock system which implicates the 5′-untranslated region of heat shock mRNA as responsible for translational control (14).

Structure of mRNA. Polymorphism was found at two bases in the coding region and at one base in the 3′-untranslated region (Fig. 2). This polymorphism results in the change of a Leu to a Val at amino acid position 263 and of an Asn to Lys at amino acid position 47. One explanation for these polymorphisms is to attribute them to cloning artifacts *in vitro* (that is, mistakes made by the reverse transcriptase) or transcriptional error *in vivo* (mistakes made by the RNA polymerase). Alternatively, this polymorphism could reflect allelic differences.

Another form of polymorphism detected was the different sites of poly(A) addition. pZM1154 differed from pZM205 in that poly(A) addition occurred 78 bp upstream (Fig. 2). The sequence AAUAAA has been described as the poly(A) addition signal (8) but we found no such sequence in the 3′-untranslated region of either pZM1154 or pZM205. pZM205 has the sequence AAAAACAAU 34 bp upstream from the poly(A) addition site. pZM1154 has no recognizable poly(A) addition signal upstream from its presumptive poly(A) addition site.

Sequence analysis suggests a preferred but variable utilization of polyadenylation signals in plants (6). The canonical sequence AAUAAA is found 31 nucleotides in front of the poly(A) addition site in maize sucrose synthase (30). This poly(A) addition signal is also found 50 to 100 nucleotides upstream from poly(A) addition sites in three of five zein mRNAs analyzed by Marks *et al.* (19). They also found zein mRNA with the signal variants AAUAAG and AAUAAU. In alcohol dehydrogenase 2, a single polyadenylation site has been described preceded 15 bp upstream by the sequence AAUAAU. In contrast, like maize aldolase, the small subunit of ribulose 1,5-bisP carboxylase has no recognizable poly(A) addition signal sequence (1).

There is precedence for polymorphism in the sites of poly(A) addition in plants. The poly(A) addition site for maize alcohol dehydrogenase 1 differs in different alleles (22). The *Adh1-1S* allele has four closely spaced poly(A) sites. The *Adh1-1F* allele has seven, including an additional major poly(A) site approximately 120 bases downstream from the most distal *Adh1-1S* poly(A) site (22). Recent work, reviewed by Birnstiel *et al.* (2), indicates that polyadenylation may be part of a concerted 3′-cleavage/polyadenylation reaction that requires recognition sequences before and after the site of poly(A) addition.

Conservation of the Aldolase Sequence. The amino acid sequence of aldolase, like other glycolytic enzymes, has been highly conserved (3, 21, 28). The primary structure of aldolase has been determined for several organisms, allowing a comparison to be made of conserved features important in catalysis, regulation, and evolution. Figure 3 shows an alignment of the amino acid sequence of maize aldolase with the sequences derived from trypanosome (5), *Drosophila* (17), and two mammalian aldolase isozymes (21, 29). This alignment of the maize amino acid sequence with other aldolase sequences suggests the existence of conserved domains in the protein. These conserved domains presumably define regions of the protein that are essential for catalysis, substrate binding, subunit interaction, folding, and regulation.

Chemical modification studies have identified specific amino acids involved in catalysis (11). The active site of rabbit aldolase A lies in a highly conserved region of the protein that includes the amino acid, Lys-229, which forms a Shiff base with the substrate. This lysine group is conserved in maize cytoplasmic aldolase and is found at position 225 (Fig. 3). Arg-55 and Lys-146 have been identified in the rabbit aldolase A as the residues interacting with the C-1 phosphate group of fructose 1,6-diphosphate. Lys-108 has been similarly identified as interacting with the C-6 phosphate group. The residues are also found in maize cytoplasmic aldolase (Arg-52, Lys-142, and Lys-103) where they are located in highly conserved domains of the protein (see the boxed amino acids in Fig. 3). The C-terminal tyrosine residue is also found to be highly conserved.

In addition to conservation of catalytic sites, subunit association sites have also been conserved. Heil and Lebherz (10) have shown that plant and mammalian aldolases are able to form active hetero-tetramers. Domains involved in subunit association have not yet been identified for aldolase. Comparisons of such distantly related aldolases will aid in the characterization of domains essential to specific functions of the protein.

Plant and animal kingdoms diverged at least 1.5 billion years ago and yet maize and human aldolases have a significant degree of sequence homology. Comparison of the maize and mammalian aldolases suggests that the primordial enzyme might have been more like the A isozyme since the maize enzyme is most like rabbit aldolase A (55% homology, see Fig. 3). However, both enzymes have important functions in cells where oxygen is limited and function primarily as a glycolytic enzyme. They may

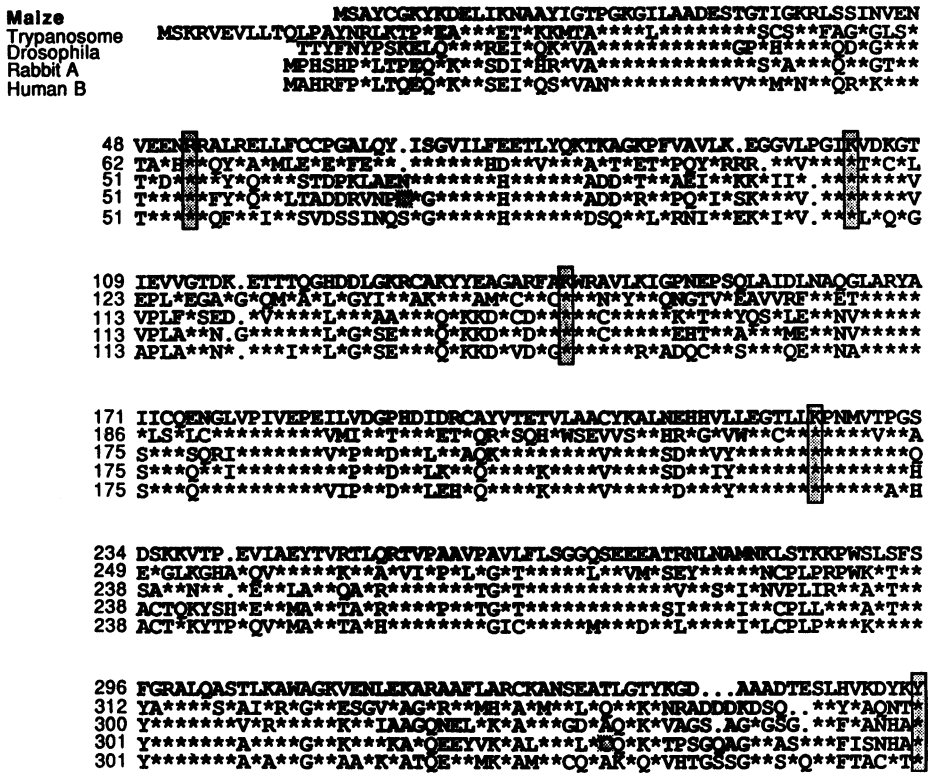

FIG. 3. Comparison of the amino acid sequence of aldolase. The amino acid sequence of maize aldolase is given at the top followed by the amino acid sequence of Trypanosome (5), *Drosophila* (17), rabbit aldolase A (28), and human aldolase B (21). Amino acids are represented by the one letter code. "*" indicates the same amino acid appears as in the corresponding position of the maize enzyme. "." indicates a gap to allow optimal alignment. Conserved amino acids indicated in the text are enclosed in solid shaded boxes. Shaded boxes enclose amino acids thought to be important in catalytic mechanisms of rabbit aldolase but are not conserved in other aldolase sequences (21).

have evolved to fulfill this role.

Acknowledgments—We wish to thank Ed Penhoet in whose laboratory this work was done, for his support and encouragement. Thanks to Sara Hake and Mike Freeling for the gift of pZMX71, Linda Johnson and George Tarr of the University of Michigan Protein Sequencing Facility for their protein analysis, and Mickey S. Urdea of the Chiron Corporation, Emeryville, CA, for the gift of the oligonucleotide primers. We wish to thank Christine Clayton for sharing her results prior to publication. Thanks to Mike Nonet for his invaluable assistance with sequencing and subcloning, and Ed Penhoet and Diane Burgess for critical reading of this manuscript.

LITERATURE CITED

1. BEDBROOK JR, SM SMITH, RJ ELLIS 1980 Molecular cloning and sequencing of cDNA encoding the precursor to the small subunit of chloroplast ribulose-1,5-bisphosphate carboxylase. Nature 287: 692–697
2. BIRNSTIEL ML, M BUSSLINGER, K STRUB 1985 Transcription termination and 3′ processing: the end is in site! Cell 41: 349–359
3. BURGESS DG, EE PENHOET 1985 Characterization of the chicken aldolase B gene. J Biol Chem 260: 4604–4614
4. CHIRGWIN JM, AE PRZYBYLA, RJ MACDONALD, WJ RUTTER 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294–5299
5. CLAYTON CE 1985 Structure and regulated expression of genes encoding fructose biphosphate aldolase. EMBO J 4: 2997–3003
6. DHAESE P, H DE GREVE, J GIELEN, J SEURINCK, M VAN MONTAGU, J SCHELL 1983 Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using agrobacterium T-DNA genes as models. EMBO J 2: 419–426
7. DAVIS RW, D BOTSTEIN, JR ROTH 1980 Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 106–107
8. FITZGERALD M, T SHENK 1981 The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24: 251–260
9. HAKE S, PM KELLEY, WC TAYLOR, M FREELING 1985 Coordinate induction of alcohol dehydrogenase 1, aldolase, and other anaerobic RNAs in maize. J Biol Chem 260: 5050–5054
10. HEIL JA, HG LEBHERZ 1978 "Hybridization" between aldolase subunits derived from mammalian and plant origin. J Biol Chem 253: 6599–6605
11. HORECKER BL, O TSOLAS, CY LAI 1972 Aldolases. *In* PD Boyer, ed, The Enzymes, Vol 7. Academic Press, New York, pp 213–269
12. KELLEY PM, M FREELING 1984 Anaerobic expression of maize glucose phosphate isomerase I. J Biol Chem 259: 673–677
13. KELLEY PM, M FREELING 1984 Anaerobic expression of maize fructose-1,6-diphosphate aldolase. J Biol Chem 259: 14180–14183
14. KLEMENZ R, D HULPMARK, WJ GEHRING 1985 Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. EMBO J 4: 2053–2050
15. LASZLO A, P ST LAWRENCE 1983 Parallel induction and synthesis of PDC and ADH in anoxic maize roots. Molec Gen Genet 192: 110–117
16. LABHERZ HG, MM LEADBETTER, RA BRADSHAW 1984 Isolation and characterization of the cytosolic and chloroplast forms of spinach leaf fructose diphosphate aldolase. J Biol Chem 259: 1011–1017
17. MALEK AA, FX SUTER, G FRANK, O BRENNER-HOLZACH 1985 Amino acid sequence of an invertebrate FBP aldolase (from *Drosophila melanogaster*). Biochem Biophys Res Commun 126: 199–205
18. MANIATIS T, EF FRITSCH, J SAMBROOK 1982 *In* Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 213–246
19. MARKS MD, JS LINDELL, BA LARKINS 1985 Nucleotide sequence analysis of zein mRNAs from maize endosperm. J Biol Chem 260: 16451–16459
20. MAXAM AM, W GILBERT 1977 A new method for sequencing DNA. Proc Natl Acad Sci USA 74: 560–564
21. ROTTMANN WH, DR TOLAN, EE PENHOET 1984 Complete amino acid sequence for human aldolase B derived from cDNA and genomic clones. Proc Natl Acad Sci USA 81: 2738–2742
22. SACHS MM, ES DENNIS, EL GERLACH, WJ PEACOCK 1986 Two alleles of maize *alcohol dehydrogenase 1* have 3′structural and poly(A) addition polymorphisms. Genetics 113: 449–467
23. SACHS MM, M FREELING 1978 Selective synthesis of alcohol dehydrogenase during anaerobic treatment of maize. Molec Gen Genet 161: 111–115
24. SACHS MM, M FREELING, R OKIMOTO 1980 The anaerobic proteins of maize. Cell 20: 761–767
25. SANGER F, S NICKLEN, AR COULSON 1977 DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
26. SEELA F, Q-H TRAN-THI, D FRANZEN 1982 Poly(7-deazaguanylic acid), the homopolynucleotide of the parent nucleoside of queuosine. Biochemistry 21: 4338–4343
27. TARR GE 86 *In* JE Shively, ed, Microcharacterization of Polypeptide: A Practical Manual. Humana Press, Inc, Clifton, NJ
28. TOLAN DR, AB AMSDEN, SD PUTNEY, MS URDEA, EE PENHOET 1984 The complete nucleotide sequence for rabbit muscle aldolase A messenger RNA. J Biol Chem 259: 1127–1131
29. VIEIRA J, J MESSING 1982 The pUC plasmids, and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259–268
30. WERR W, W-B FROMMER, C MAAS, P STARLINGER 1985 Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. EMBO J 4: 1370–1380