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

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

PHILIP M. KELLEY 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 cDNA, pZMX71 was found to hybrid select mRNA encoding a protein of approximately 40,000 mol wt that was selectively precipitated by an antisempecific 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 H2O at 23°C, then germinated on moist paper towels at 27°C for 4 d. Seedlings were collected and submerged within L of distilled H2O. Primary roots were cut off and frozen in liquid N2. 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 Madrid preparation: 6 mM 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 Blender. 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 300 μL distilled H2O. 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 H2O 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 cDNA library was fractionated over a Sepharose CL-4B column (0.5 x 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× Denhardt's, 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...
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 cloned 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 intranarrow base pair interaction (26). The cDNA insert from pZMX71 was subcloned into Pst I cut pUC8, then cut with EcoRI 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 M HCl for 1 h at 150°C, then derivatized with 3-phenyl-2-thiodyantoin 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 dimethylglycinediamine, 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 aldolase 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 aldolase 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, A10TTA19. 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.

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/mmol 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 hybridized 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

| Table I. Synthesis of Maize Anaerobic cDNA Library |
|-----------------|-----------------|
|                 | Yield | Yield % |
| Input maize anaerobic root poly(A) mRNA | 10.0 | 100 |
| First strand synthesis | 4.2 | 42 |
| Second strand synthesis | 4.8 | 24 |
| S1-nuclease treatment | 2.4 | 12 |
| C-tailing | 2.2 | 11 |
| Sepharose CL-4B | 0.5 | 2.5 |
| High mol wt cDNA annealed with G-tailed pUC8 | 0.5, dsDNA |
| Transformants obtained | 80,000 colonies |
| Efficiency | 16,000 colonies/μg |

Average insert size 1,100 bp
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).

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 sequence 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.
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. Antiserum 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 AAAAACAUAU 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 AAUAU. 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 predecence 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 Schiff 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 Leiberher (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
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 Freeing 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

2. BURNSTEIN ML, M BUSELINGER, K STRUB 1985 Transcription termination and 3‘ processing: the end is in Cell! Cell 41: 349-359
4. CHIRGWIN JM, AE PRZYBYLA, RJ MACDONALD, WJ RUTTER 1979 Isolation of biologically active ribonuclease from sources enriched in ribonuclease. Biochemistry 18: 5294-5299
5. CLAYTON CE 1985 Structure and regulated expression of genes encoding fructose bisphosphate aldolase. EMBO J 4: 2997-3003
8. FITZGERALD M, T SHIEN 1981 The sequence 5‘-AAAAA-3‘ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24: 251-260
12. KELLEY PM, M FREEING 1984 Anaerobic expression of maize glucose phosphatase isomerase I. J Biol Chem 259: 673-677
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 DENVIS, EL GERLACH, WJ PEACOCK 1986 Two alleles of maize alcohol dehydrogenase 1 have 3‘ structural and poly(A) addition polymorphisms. Genetics 113: 449-467
28. TOLAN DR, AB AMDSN, SD PUTNEY, MS URDEA, EE PENHOET 1984 The complete nucleotide sequence for rabbit muscle aldolase A messenger RNA. J Biol Chem 259: 1127-1131