Journal of Plant Biology, September 2000, 43(3) : 162-170

Arabidopsis thaliana Cytidine Deaminase 1 Shows More Similarity to Prokaryotic Enzymes Than to Eukaryotic Enzymes[1]

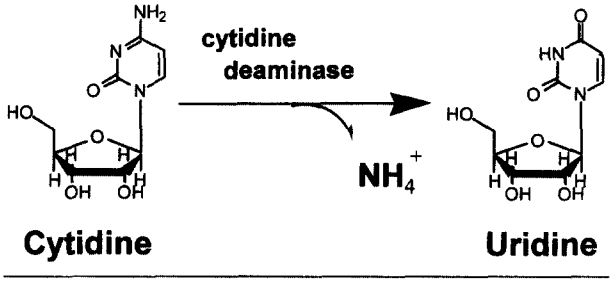
Chris Kafer and Robert W. Thornburg
Department of Biochemistry, Biophysics, and Molecular Biology, 2212 Molecular Biology Building, Iowa State University, Ames, Iowa 50011, USA

Two Expressed Sequence Tagged (EST) clones were identified from the Arabidopsis database as encoding putative cytidine deaminases. Sequence analysis determined that the two clones overlapped and encoded a single cDNA. This cytidine deaminase corresponds to the *Arabidopsis thaliana* gene, *cda1*. The deduced amino acid sequence was more closely related to prokaryotic cytidine deaminases than to eukaryotic enzymes. The cDNA shares 44% amino acid identity with the *Escherichia coli* cytidine deaminase but only 26 and 27% identity with human and yeast enzymes. A unique zinc-binding domain of the *E. coli* enzyme forms the active site. A similar putative zinc-binding domain was identified in the Arabidopsis enzyme based upon primary sequence similarities. These similarities permitted us to model the active site of the Arabidopsis enzyme upon that of the *E. coli* enzyme. In this model, the active site zinc is coordinated by His^{73}, Cys^{103}, Cys^{107}, and an active site hydroxyl. Additional residues that participate in catalysis, Asn^{64}, Glu^{66}, Ala^{78}, Glu^{79}, and Pro^{102}, are conserved between the Arabidopsis and *E. coli* enzymes suggesting that the Arabidopsis enzyme has a catalytic mechanism similar to the *E. coli* enzyme. The two overlapping ESTs were used to prepare a single, full-length clone corresponding to the *A. thaliana cda1* cDNA. This cDNA was subcloned into pProExHtb and expressed as a fusion protein with an N-terminal His_6 tag. Following purification on a Ni-NTA-Agarose column, the protein was analyzed for its kinetic properties. The enzyme utilizes both cytidine (K_m = 226 μM) and 2'-deoxycytidine (K_m = 49 μM) as substrates. The enzyme was unable to deaminate cytosine, CMP or dCMP.

Keywords: cytidine deaminase

Pyrimidines are at the core of cellular metabolism. They are components of DNA and RNA, they activate monosaccharides, they modify regulatory proteins, and they are involved in basic aspects of cellular physiology. Deficiencies in pyrimidine metabolism completely block cellular growth and development in both bacteria (Yamanaka et al., 1992) and yeast (Prakash et al., 1979; Jong and Campbell, 1984).

Cytidine deaminase (EC 3.5.4.5) catalyzes the first step in the salvage of pyrimidine nucleotides by hydrolytic deamination of cytidine into uridine. Cytidine deaminase has been identified and characterized from a wide variety of sources from bacteriophage (Maley, 1978) to mammals (Rothman et al., 1978). Our interest in cytidine deaminase stems from its role in pyrimidine salvage; however, cytidine deaminase also generates significant interest due to a variety of other observations. This makes a comparative analysis of the plant enzyme important in understanding the role of this enzyme in cell biology. First, in humans, cytidine deaminase is responsible for the inactivation of useful chemotherapeutic agents such as cytidine arabinoside that are widely used in the treatment of leukemia (Steuart and Burke, 1971; Meyers et al., 1973). Second, cytidine deaminase catalyzes RNA editing. The best characterized example of RNA editing is the C to U editing of nucleotide 6666 of apolipoprotein B (apo-B) RNA in mammalian intestine catalyzed by a novel cytidine deaminase (Navaratnam et al., 1993; Teng et al., 1993). This modification generates a new stop translation codon (UAA) from a glutamate codon (CAA) producing a truncated protein (Powell et al., 1987).



Abbreviations: EST-Expressed Sequence Tagged.

[1]Journal Paper Number J-18324 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa Project No. 3340.
*Corresponding author; fax +1-515-294-0453
e-mail thorn@iastate.edu

RNA editing also occurs in both the mitochondria and chloroplasts of plants from bryophytes (Malek et al., 1996) to gymnosperms (Wakasugi et al., 1996) and angiosperms including both monocots (Freyer et al., 1997) and dicots (Lippok et al., 1996). RNA editing has been shown to completely modify the coding region of genes by creation of new stop codons (Zanlungo et al., 1995), new start codons (Kadowaki et al., 1995), and even creating entirely new open reading frames (Wakasugi et al., 1996).

MATERIALS AND METHODS

Bacterial Strains

Escherichia coli strains XL1 Blue or DH5α were used for routine subcloning and sequencing DNA preparations. *E. coli* BL21 was used for expression of the recombinant protein due to the lack of *omp* and *lon* proteases.

DNA Sequencing

DNA sequence reactions were performed at the Iowa State University Nucleic Acid Facility using the Applied Biosystems Prism Dye-deoxy Cycle Sequencing Kit. The reactions were run on an Applied Biosystems Prism 377 DNA sequencer (Perkin-Elmer Corp., Norwalk, CT). Sequencing was initiated from known vector sequences. On the basis of these runs, primers specific to the cytidine deaminase sequence were constructed. Both strands were sequenced in duplicate or triplicate. RNA structural calculations were conducted with the program RNAdraw (Matzura and Wennborg, 1996).

Construction of Expression Cassette

Two Espressed Sequence Tagged (EST) clones (150P17T7 and 149H23T7) were identified as potential cytidine deaminase cDNAs by Blast searches (Altschul et al., 1990) and obtained from the Arabidopsis Biological Resource Center stock center (ABRC, 1995). The complete sequence of each clone showed that they were partial cDNAs that together encoded a single open reading frame of a putative cytidine deaminase.

A single, full-length cytidine deaminase clone was constructed from these two cDNAs. The 150P17T7 EST was digested with XbaI and BglII and ligated into the XbaI/BglII sites of pRT146 (a pUC derivative in which a BglII linker had been inserted into the HincII site). This resulting plasmid was named pRT428. The EST clone, 149H23T7, was digested with BglII and SphI and ligated into the BglII/SphI sites of pRT428. The resulting clone, pRT429, contained the full-length cytidine deaminase cDNA.

The expression plasmid, pRT439, was constructed by PCR amplification of the coding region with the pair of oligonucleotides CYDP2 and CYDP3; CYDP3 (5'-GCGCGGATCC*ATCGAAGGTCGT***ATG**GATAAGCC-AAGCTTCGTA-3') contained the start codon (bold), a Factor Xa protease site (italicized), and a BamHI site (underlined); CYDP2 (5'-CGGGATCCCTCGAG**CTA**G-CTTCATAGCAATGAAACAC-3') was complementary to the cDNA, included the stop codon (bold), and contained BamHI and XhoI sites (underlined). The resulting PCR product was digested with BamHI and XhoI and ligated into the BamHI/XhoI sites of pProEXHtb (Gibco, Gaithersburg, MD).

Expression of Cytidine Deaminase and Purification of Recombinant Protein

Individual *E. coli* BL21 transformants carrying pRT439 were picked from ampicillin selection plates and used to inoculate liquid media. The culture was grown in 50 to 100 mL NZCYM media (10 g/L casein hydrolysate, 5.0 g/L yeast extract, 1.0 g/L casamino acids, 2.0 g/L $MgSO_4 \cdot 7H_2O$) w/ 75 μg/mL ampicillin at 37°C until approximately 0.6 A_{600}. A 1 mL aliquot was taken to serve as an uninduced control. IPTG was then added to a final concentration of 1 mM, and the cultures were incubated for another 4 h.

Cells were harvested by centrifugation at 5000*g* for 10 min at 4°C. The pellet was resuspended in 10 mL binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-Cl pH 8.0) containing 1 mg/mL lysozyme and sonicated on ice for 30 s. Cellular debris was removed by centrifugation at 10,000*g* for 20 min at 4°C. The supernatant was applied to a 1 mL Ni-NTA-affinity column (Quiagen). The column was washed extensively (20 to 30 column volumes) with wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris-Cl pH 8.0) to remove nonspecifically bound proteins. The bound, recombinant His_6 tagged protein was eluted with 1.5 mL of column-stripping buffer (100 mM EDTA, 20 mM Tris-Cl pH 8.0, 0.5 M NaCl). The protein was stored at −20°C until used and was stable through several freeze/thaw cycles.

Cytidine Deaminase Assay

The cytidine deaminase assay was performed

essentially as described (Wentworth and Wolfenden, 1978). Each reaction included 0.1 mL 1.67 mM cytidine, 0.5 mL 0.1 M Tris-HCl pH 8.0, enzyme, and water to 1 mL final vol for specific activity measurements. The decrease in absorbance at 290 nm was followed with time. As was observed for the *E. coli* enzyme (Wentworth and Wolfenden, 1978) no loss of activity was found when the enzyme was treated with high levels of EDTA. Assays for 2'-deoxycytidine deaminase were performed essentially as described (Ipata and Cercignani, 1978). One unit of activity is defined as the amount of enzyme required to catalyze the deamination of 1 µMol of substrate per min at 27°C in the above assay.

RESULTS

Characterization of the Cytidine Deaminase mRNA

The full sequence of the *Arabidopsis thaliana* cytidine deaminase mRNA was constructed from two independent EST clones. Initially, we determined the entire nucleotide sequence of each clone. The comparison of these two clones, 149H23T7 and 150P17T7, revealed an identical overlapping region of 672 nucleotides (see Fig. 1, Panel A). Clone 150P17T7 is incomplete at the 3' end. It contains the 5' untranslated region, start codon and the majority of the coding region extending to nucleotide 1003 (Ser^{288}). Clone 149H23T7 is incomplete at the 5' end. It begins in the coding region at nucleotide 331 (Val^{65}) and extends through the stop codon and a short 3' untranslated region to the poly A tail of the cytidine deaminase. A full-length clone was prepared from the two incomplete cytidine deaminase clones as outlined in the Materials and Methods section (Fig. 1, Panel B).

mRNA Features

The mRNA encoding the putative cytidine deaminase deduced from the overlapping clones is 1161 nucleotides long. Based upon DNA matrix analysis, there were no significant internal duplications within the mRNA. A 141-nt 5' untranslated region was identified. There was no nucleotide bias of either AT (52%) or GC (48%) within the 5' untranslated region. The first ATG was found at nucleotides 142 to 144 of the mRNA. The identification of a purine nucleotide at nucleotide 139 and a G at nucleotide 145 (–3 and +4 relative to the ATG start codon) indicate that this

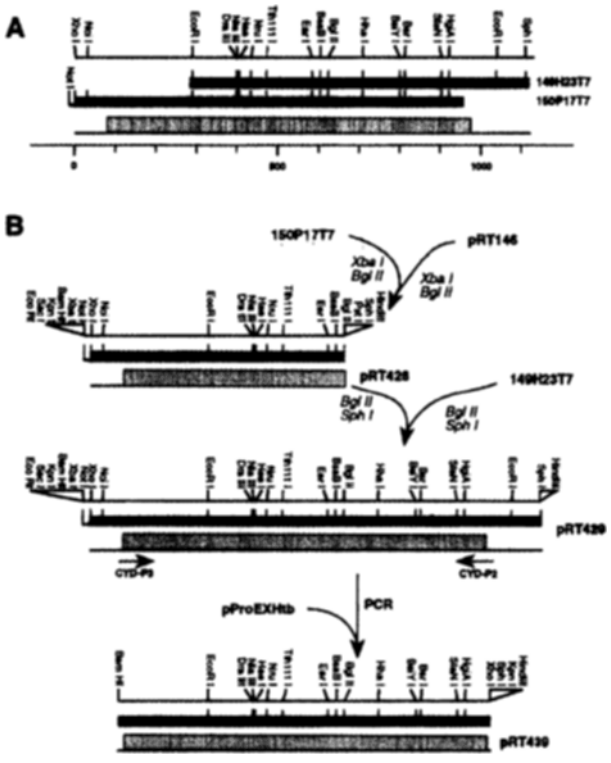


Figure 1. *A. thaliana* cytidine deaminase cDNAs. **Panel A.** Physical structure of the cytidine deaminase EST clones 150P17T7 and 149H23T7 are presented as black lines. Conserved restriction enzyme sites between the two clones are indicated. The stippled box below the clones shows the structural features (5' UTR, coding region, and 3' UTR) of the cDNA. A scale showing the size of the clones in nucleotides is presented below. **Panel B.** Construction of the full-length cytidine deaminase cDNA. Features of this panel are the same as in Panel A.

mRNA is likely to be efficiently translated (Kozak, 1986).

The mRNA contains a single, 301 codon-long open reading frame that encodes a protein with high identity to the *E. coli* cytidine deaminase (see below). An amber stop codon (TAG) is located at nucleotides 1045 to 1047 of the mRNA and a 113 bp 3' untranslated sequence extends beyond. The polyadenylation site was found at nucleotide 1136. A consensus polyadenylation signal (AATAAA) was observed 46 nucleotides upstream from the polyadenylation site at nucleotides 1085 to 1090 of the mRNA. Further analysis of this region of the mRNA revealed that an unusual hairpin structure overlapped this polyadenylation signal (Fig. 2). The ΔG of formation of this hairpin structure was –38.64 kJ/mol at 20°C, indicating that this structure is likely to form spontaneously within the mRNA. Thus, this polyadenylation signal is

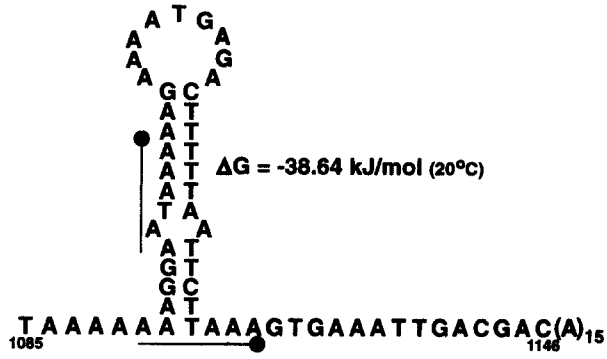


Figure 2. Stem-loop structure forms a novel polyadenylation signal. The sequence of the *A. thaliana* cytidine deaminase mRNA from 1029 to 1090 is presented. This structure was calculated to form with a ΔG of -38.64 kJ/mol at 20°C.

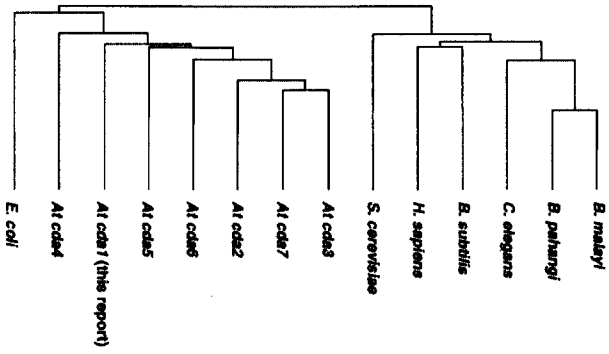


Figure 3. Phylogenetic analysis of cytidine deaminases. GenBank Accessions for the individual sequences used in this analysis are: *E. coli* GenBank Accession M60916; *S. cerevisiae* U20865; *H. sapiens* L27943; *Bacillus subtilis* P19079; *Cenorhabditis elegans* U61949; *Brugia pahangi* X91065; *B. malayi* U80980, *A. thaliana* CYD1 AF134487; *A. thaliana* CYD2 through CYD7, AF080676.

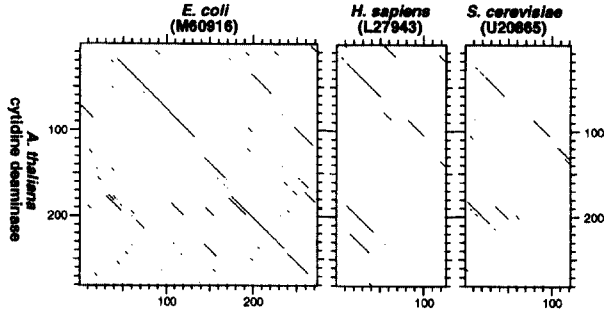


Figure 4. Protein identity matrix comparison of the *A. thaliana* cytidine deaminase (GenBank Accession AF1334487) with cytidine deaminase from *E. coli* (M60916), *H. sapiens* (L27943), and *S. cerevisiae* (U20865). The analysis was performed with 30% identity in a window of 23 amino acids.

likely to be located within a hairpin structure. Uniquely, this hairpin structure results in the formation of a new consensus polyadenylation site that spans the base of the hairpin. This bipartite polyadenylation signal is located only 14 nucleotides from the polyadenylation site (see Fig. 2).

Protein Features

A single open reading frame was identified within the putative cytidine deaminase mRNA that encodes a 301 amino acid protein. The full-length protein has a calculated molecular weight of 32,582 and a pI of 5.52. To identify this protein, BLAST searches (Altschul et al., 1990) were performed using the translated amino acid sequence. This analysis identified the most closely related sequences in the databases as cytidine deaminases. Interestingly, the most closely related sequence was the *E. coli* cytidine deaminase rather than other eukaryotic homologues (Fig. 3). Recently a family of seven cytidine deaminase cDNAs from Arabidopsis was deposited in the GenBank. All of these proteins show greater similarity to the *E. coli* enzyme than to other eukaryotic enzymes. The cytidine deaminase cDNA identified in this report corresponds to the cytidine deaminase *cda1* gene.

To further verify the relationship between the Arabidopsis and other enzymes, we performed a protein matrix analysis comparing the Arabidopsis cDNA with sequences from *E. coli* (GenBank Accession Number M60916), *Homo sapiens* (L27943), and *Saccharomyces cerevisiae* (U20865). This analysis is shown in Figure 4. While identity is clearly present between the Arabidopsis enzyme and all other cytidine deaminases, the most extensive identity is with the *E. coli* enzyme, and this identity extends throughout the full length of the enzyme. Further, the length of these enzymes differs significantly. Other eukaryotic cytidine deaminases are approximately half the size of the *E. coli* and Arabidopsis enzymes.

There are seven cysteine residues in the Arabidopsis cytidine deaminase, so there must be at least one unpaired cysteine residue in the cytidine deaminase protein as is found in other cytidine deaminases (Ipata and Cercignani, 1978). Both acidic and basic residues are scattered throughout the protein; however, the distribution of the negatively charged residues is a bit unusual. Both the N-terminal third (amino acids 1 to 120) and the C-terminal third of the protein (amino acids 225 to 301) are rich in glutamate (7/9 and 8/10 residues respectively). In contrast the middle third (amino acids 121 to 224)

contains 16 negatively charged amino acids of which 15 are aspartates.

This protein does not contain readily discernible N-terminal targeting sequences that would direct the cytidine deaminase towards the secretory pathway, nucleus, or mitochondria. However, this protein does contain some features of a chloroplast protein including a high percentage of serine and alanine (27%) in the N-terminal 30 amino acids (Nakai and Kanehisa, 1992).

It has previously been suggested that the mRNA editing cytidine deaminase in mammalian cells forms a heterodimer with the 60 kDa RNA binding component by virtue of a leucine zipper (Navaratnam et al., 1993). Numerous leucine residues are present throughout the middle portion of the Arabidopsis protein (amino acids 143 to 179). The spacing; however, of these residues in the Arabidopsis cytidine deaminase does not form a perfect heptad repeat, suggesting that this portion of the protein may not form a leucine zipper. Further, structural predictions (Rost et al., 1994) on this portion of the cytidine deaminase indicate that this region is not likely to form an α-helix.

Model of the Active Site

Other cytidine deaminase enzymes contain a zinc-binding, active site that is required for catalysis (Driscoll and Zhang, 1994). This active site has been best studied in the *E. coli* enzyme (Betts et al., 1994; Xiang et al., 1995; Carlow et al., 1996; Xiang et al., 1996; Xiang et al., 1997; Carlow et al., 1998). In the *E. coli* enzyme a zinc-binding pocket exists between two alpha helices. In the *E. coli* enzyme the zinc is coordinated by His^{102}, Cys^{129}, Cys^{132}, and an active site water molecule (Betts et al., 1994; Xiang et al., 1995; Carlow et al., 1996; Xiang et al., 1996; Xiang et al., 1997; Carlow and Wolfenden, 1998). An alignment of the Arabidopsis sequence with that of the *E. coli* enzyme and other cytidine deaminases is presented in Figure 5, Panel A. The zinc-binding amino acids are indicated with arrows. These zinc-binding residues are conserved in the Arabidopsis enzyme as residues His^{77}, Cys^{103}, and Cys^{107}.

Structural predictions on the Arabidopsis cytidine deaminase were performed using the PredictProtein server at EMBL-Heidelberg. The region from amino acids 77 to 110 was predicted to contain two α-helices separated by a β-sheet. This same pattern is observed in the 3D structure of the *E. coli* cytidine deaminase [5]. Therefore, a 24 amino acid sequence corresponding to the zinc-binding domain was modeled using the *E. coli* cytidine deaminase crystal structure (PDB accession # 1AF2) using a "first approach" method with the Swiss Model program at http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html. The results of this are shown in Figure 5 panel B. In this model, the sidechains of selected amino acids are shown. As indicated above, His^{77}, Cys^{103}, and Cys^{107} are involved in the coordination of the active site zinc atom. In addition, a hydrogen bond between the carboxylate group of the conserved Glu^{79} and zinc-bound hydroxyl group apparently is critical in the stabilization of the transition state intermediate. It also serves as a proton shuttle from the zinc-bound

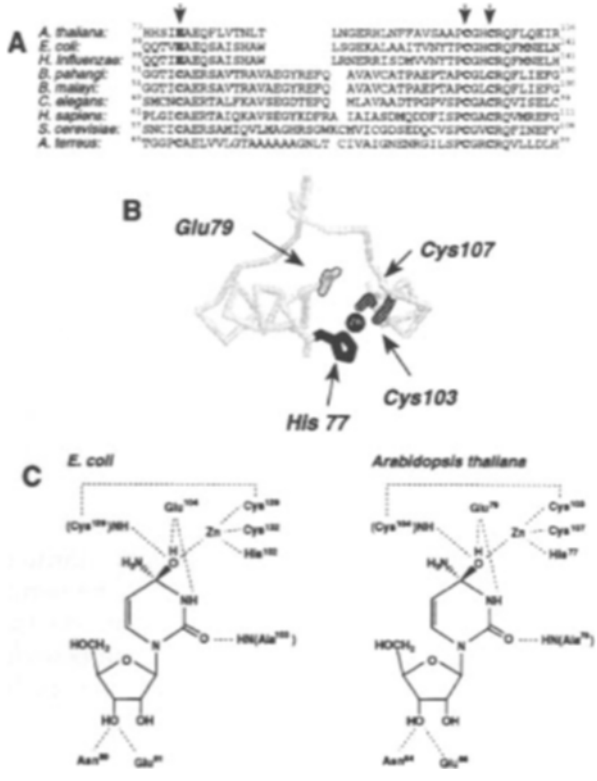


Figure 5. *A. thaliana* cytidine deaminase active site. **Panel A.** Alignment of zinc-finger domains from various cytidine deaminases. GenBank Accessions for the individual sequences used in this analysis are presented in the legend to Figure 3. **Panel B.** Model of the *A. thaliana* cytidine deaminase zinc-finger domain from amino acids 77 to 111. The alpha carbon chain for the peptide is presented. The side chains of the amino acids that participate in the enzyme mechanism are presented. The zinc-coordinating His^{77} is shown in black, and Cys^{103} and Cys^{107} are stippled. The active site Glu^{79} is shown in hatched. **Panel C.** Structural elements at the cytidine deaminase active site. The active site of the *E. coli* is taken from [5]. The conserved active site residues of the Arabidopsis enzyme are shown interacting with the tetrahedral active site reaction intermediate.

hydroxyl group to the leaving amino group (Carlow et al., 1996; Xiang et al., 1997). As can be seen by comparing the model with the *E. coli* enzyme structure, each of these functional sidechains in the Arabidopsis enzyme is in the appropriate location to mediate the hydroxylation/deamination reaction. In addition, several other amino acids that are known to participate in substrate binding in the *E. coli* enzyme such as the Ala^{103} backbone amide that forms a hydrogen bond with the 2-keto group of cytidine, Asp^{89} and Glu^{91} that form hydrogen bonds with the ribosyl 3' hydroxyl, and Pro^{128} that orients the amino leaving group are all conserved in the Arabidopsis enzyme as Ala^{78}, Asn^{64}, Glu^{66}, and Pro^{102}. Thus, based upon the high degree of conservation of structure and amino acids that are involved in the active site chemistry, we postulate that the mechanism of cytidine deamination in the Arabidopsis enzyme is similar to that of the *E. coli* enzyme. Figure 5, Panel C shows the conservation of interacting residues between the *E. coli* and the Arabidopsis enzyme.

Expression, Purification, and Analysis of Recombinant Cytidine Deaminase

To conclusively demonstrate that this cDNA encodes a cytidine deaminase, we expressed a recombinant form of the protein in *E. coli*, purified it, and evaluated its enzymatic activity. A full-length cytidine deaminase clone was constructed as described in Materials and Methods. This construct contained a 32 amino acid extension at the N-terminus of the recombinant protein that contains a His_6 affinity tag to facilitate purification. A secondary structural analysis of the recombinant protein with this N-terminal extension was performed. This extension was not expected to significantly alter the protein structure or its localization in *E. coli*. This clone was placed under the control of the powerful *Trc* promoter. This plasmid, pRT439, was then transformed into *E. coli* BL21 cells, and the recombinant cytidine deaminase was expressed. After an initial growth to log phase, the cells were induced with 1 mM IPTG for 4 h. The resulting cells were concentrated by centrifugation, resuspended, and broken by sonication. After another centrifugation to pellet cell debris, the crude lysate was applied to a Ni-NTA column and allowed to flow through under gravity. The column was washed and subsequently eluted with 100 mM EDTA. The purity of the protein preparation was evaluated by SDS PAGE as shown in Figure 6. This figure shows the specific induction of a 36 kDa protein following IPTG treatment (compare lanes 1 and 2). The recombinant protein expressed from pRT439 is predicted to have a molecular mass of 36.4 kDa. Lane 3 shows the protein fraction bound to the Ni-NTA affinity column and specifically eluted with imidizole. This purified enzyme fraction was utilized for all subsequent protein methods.

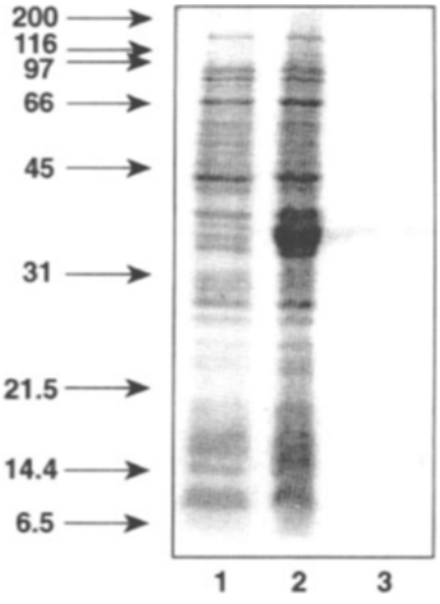


Figure 6. Purification of the recombinant cytidine deaminase on the Ni-NTA-affinity column. Lane 1, proteins from uninduced *E. coli* BL21 cells containing pRT439; Lane 2, proteins from IPTG-induced BL21 cells containing pRT439; Lane 3, purified recombinant cytidine deaminase. Molecular weight marker sizes in kDa are indicated to the left of the figure. Molecular weight markers used were, Myosin, 200 kDa; β-galactosidase, 116 kDa; Phosphorylase B, 97 kDa; BSA, 66 kDa; Ovalbumin, 45 kDa; Carbonic Anhydrase, 31 kDa; Soybean Trypsin Inhibitor, 21.5 kDa; Lysozyme, 14.4 kDa; Aprofinin, 6.5 kDa.

Initially we evaluated the deamination of the cytosine-free base, cytidine nucleosides, and cytidine nucleotides. This analysis, shown in Table 1, demonstrates that the recombinant enzyme shows no activity against either the free base, cytosine, or the cytidine nucleotides, CMP or dCMP. The lack of activity in deaminating cytosine has been previously reported for the *E. coli* enzyme where the removal of the ribose reduces catalytic activity by 10^8 (Carlow et al., 1998). These authors indicated that the electron-withdrawing nature of the substituent ribose activates the cytosine ring for nucleophilic attack. The lack of activity for CMP and dCMP also can be explained by the enzyme structure. An important hydrogen bond is formed between the ribose 5'-hydroxyl and the back-

Table 1. Substrate specificity: The standard enzyme reaction contained 167 μM cytidine analog, 0.05 M Tris-HCl pH 8.0, enzyme, and water in a final volume of 1 mL. The decrease in absorbance at 290 nm was followed with time. Experiments were performed in triplicate or quadruplicate. For determination of the K_m values the standard enzyme reaction contained variable amounts of cytidine or 2'-deoxycytidine ranging from 50 to 500 μM. One unit of cytidine deaminase is defined as the amount of enzyme required to deaminate 1 μMole of cytidine per minute in the above standard reaction mixture at 25°C. Specific activity is units per milligram of protein as determined by the method of Bradford (Bradford, 1976).

substrate	specific activity[a]	K_m[b]	V_{max}[c]
cytidine	12.0 ± 1.6	226.1 ± 45	39.7 ± 3.8
2'-deoxycytidine	20.3 ± 3.6	49.3 ± 6.7	24.4 ± 1.3
CMP	nd[d]		
dCMP	nd		
cytosine	nd		

[a] units per mg protein, [b] μMolar, [c] μMoles/min
[d] no deaminase activity was detected

bone carbonyl oxygen atom Ala631 of the other monomer [5]. The presence of a phosphate group on the ribose 5'-hydroxyl would therefore not permit this substrate to fit the binding pocket.

In contrast, the Arabidopsis enzyme showed high activity in the deamination of cytidine and an almost twofold higher activity in the deamination of 2'-deoxycytidine (see Table 1). This also was observed with the *E. coli* enzyme (Ashley and Bartlett, 1984) and can be explained from the crystal structure. The 2'-hydroxyl does not make hydrogen bonds within the active site. In order for cytidine to bind, the 2'-hydroxyl must be desolvated, and this desolvation of the 2'-hydroxyl destabilizes the cytidine protein complex. Consequently, the deoxycytidine is more strongly bound [5]. To assess this with the Arabidopsis enzyme, we also determined kinetic constants for the recombinant enzyme with both of the active substrates. As shown in Table 1, the K_m values for cytidine and deoxycytidine indicate that deoxycytidine binds approximately fivefold more tightly to the Arabidopsis enzyme than does cytidine (49 μM vs 226 μM). Again, this is similar to the *E. coli* enzyme.

Based upon these analyses, we conclude that this cDNA does indeed encode a cytidine deaminase with similar structure and enzymatic activity to the *E. coli* cytidine deaminase. Further, this cDNA appears to be a member of a closely related gene family with at least seven members that are all more similar to the prokaryotic cytidine deaminase than to the eukaryotic enzymes.

DISCUSSION

We have identified two overlapping EST clones as Arabidopsis cytidine deaminase clones. From these two clones we have reconstructed the full-length cDNA and analyzed the mRNA and protein. We have expressed the cDNA in *E. coli* and verified that the expressed protein has cytidine deaminase activity. The mRNA encoding the cytidine deaminase has an unusual feature in its 3' end that may function as a novel polyadenylation site. Such structural features are not common, and to our knowledge this unusual hairpin appears to be unique.

Structurally, the Arabidopsis protein is more similar to the *E. coli* than it is to either the yeast or the human cytidine deaminase. The *E. coli* enzyme is known to be a dimer of 31 kDa subunits [5], while the human enzyme is a tetramer of 13 kDa subunits (Cacciamani et al., 1991). In our case, the Arabidopsis protein is translated with a molecular weight of 32.5 kDa and has identity with the *E. coli* enzyme throughout its sequence.

The substrate specificity also is similar to that found in the *E. coli* enzyme. Both enzymes are unable to use cytosine as a substrate. It has been previously demonstrated that plants, like animals and unlike microorganisms, lack a cytosine deaminase activity (Stougaard, 1993). This lack of cytosine deaminase points to the inability of plants to salvage the cytosine-free base. Indeed, because of this inability to salvage the free base, fluorinated cytosine cannot be metabolized in higher eukaryotes and is an effective antimicrobial compound. This has been very effectively utilized in animal systems, but has been widely overlooked in plant systems, especially plant tissue cultures.

The inability of this enzyme to deaminate CMP or dCMP is also noteworthy and leads to several unanswered metabolic questions. The Arabidopsis UMP/CMP kinase is very effective in converting CMP to CDP (Zhou et al., 1998). Thus, it appears that cytidine nucleotides are recycled to CDP and subsequently CTP rather than to uridine nucleotides. This, however, does not appear to be the case for dCMP. The Arabidopsis UMP/CMP kinase is 30-fold less effective in the conversion of the dCMP into dCDP than in the conversion of CMP into CDP (Zhou et al., 1998). Whether dCMP is metabolized into dUMP or dCDP remains unresolved. The cytidine deaminase, however, is twice as active in the conversion of deoxycytidine into deoxyuridine than in the conversion of cytidine into uridine. It also is unclear whether deox-

yuridine is metabolized to uracil and deoxyribose, to dUMP, or both, as is the case in *E. coli* (Neuhard and Nygaard, 1987). Conversion to dUMP via phosphorylation with thymidine kinase would provide an alternative route to TMP biosynthesis. Deribosylation to uracil would recycle the ring system to either UMP by way of uracil phosphoribosyltransferase (Weers and Thornburg, 1999) or to degradation by dihydrouracil dehydrogenase. In bacteria, deoxyribose-5-phosphate is metabolized to glyceraldehyde-3-phosphate and acetaldehyde (Ackermann et al., 1974). Whether eoxyribose metabolism occurs in plants has not, to our knowledge, been evaluated. Resolution of these metabolic questions will require further investigation and isolation of additional salvage enzymes.

That all of the Arabidopsis cytidine deaminases are much more similar to the *E. coli* enzyme than to the eukaryotic enzymes from either humans or *Saccharomyces* is surprising. This similarity is both structural and functional. The structural similarities have permitted us to model the active site of the Arabidopsis enzyme based on the active site of the *E. coli* enzyme. Based upon this analysis, all features of the *E. coli* enzyme that have been identified to participate in the bacterial enzymatic mechanism are conserved in the Arabidopsis enzyme. The functional similarities between the bacterial and plant enzyme also are quite remarkable and lead us to the conclusion that the bacterial form of the cytidine deaminase has been conserved in plants.

ACKNOWLEDGMENTS

This work was sponsored by funds from the Carver Trust, the Hatch Act, and State of Iowa funds.

Received September 1, 2000; accepted September 19, 2000.

LITERATURE CITED

ABRC *Arabidopsis* Biological Resource Center (1995) Seed and DNA Stock List. Library information for ESTs 239

Ackermann RS, Cozzarelli NR, Epstein W (1974) Accumulation of toxic concentration of methylglyoxal by wild-type *Escherichia coli* K-12. J Bacteriol 119: 357-362

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Ashley GW, Bartlett PA (1984) Inhibition of *Escherichia coli* cytidine deaminase by a phosphapyrimidine nucleoside. J Biol Chem 259: 13621-13627

Betts L, Xiang S, Short SA, Wolfenden R, Carter C (1994) Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. J Mol Biol 235: 635-656

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein-dye binding. Anal Biochem 72: 248-254

Cacciamani T, Vita A, Cristalli G, Vincenzetti S, Natalini P, Ruggieri S, Amici A, Magni G (1991) Purification of human cytidine deaminase: molecular and enzymatic characterization and inhibition by synthetic pyrimidine analogs. Arch Biochem Biophys 290: 285-292

Carlow D, Wolfenden R (1998) Substrate connectivity effects in the transition state for cytidine deaminase. Biochemistry 37: 11873-11878

Carlow DC, Short SA, Wolfenden R (1996) Role of glutamate-104 in generating a transition state analogue inhibitor at the active site of cytidine deaminase. Biochemistry 35: 948-954

Carlow DC, Short SA, Wolfenden R (1998) Complementary truncations of a hydrogen bond to ribose involved in transition-state stabilization by cytidine deaminase. Biochemistry 37: 1199-1203

Driscoll DM, Zhang Q (1994) Expression and characterization of p27, the catalytic subunit of apolipoprotein B mRNA editing enzyme. J Biol Chem 269: 19843-19847

Freyer R, Kiefer-Meyer MC, Kossel H (1997) Occurrence of plastid RNA editing in all major lineages of land plants. Proc Natl Acad Sci USA 94: 6285-6290

Ipata P, Cercignani G (1978) Cytosine and cytidine deaminase from yeast. Methods-Enzymol 51: 394-401

Jong AY, Campbell JL (1984) Characterization of *Saccharomyces cerevisiae* thymidylate kinase, the CDC8 gene product. General properties, kinetic analysis, and subcellular localization. J Bio Chem 259: 14394-14398

Kadowaki K, Ozawa K, Kazama S, Kubo N, Akihama T (1995) Creation of an initiation codon by RNA editing in the *coxI* transcript from tomato mitochondria. Curr Genet 28: 415-422

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283-292

Lippok B, Brennicke A, Unseld M (1996) The *rps4*-gene is encoded upstream of the *nad2*-gene in *Arabidopsis* mitochondria. Biol Chem Hoppe Seyler 377: 251-257

Malek O, Lattig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15: 1403-1411

Maley GF (1978) Deoxycitidylate deaminase from T2-infected *Escherichia coli*. Methods Enzymol 51: 412-418

Matzura O, Wennborg A (1996) RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Computer Applications in the Biosciences (CABIOS) 12: 247-249

Meyers R, Malathi VG, Cox RP, Silber R (1973) Studies on nucleoside deaminase. Increase in activity in HeLa cell cultures caused by cytosine arabinoside. J Biol Chem 248: 5909-5913

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14: 897-911

Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannoni F, Teng B, Davidson NO, Scott J (1993) The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem 268: 20709-20712

Neuhard J, Nygaard P (1987) Purines and Pyrimidines, *In* FC Neidhardt, JL Ingraham, BK Low, B Magasanik, M Schaechter, HE Umbarger, eds, *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, American Society for Microbiology, Washington, DC

Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J (1987) A novel form of tissue-specific RNA processing produces Apolipoprotein-B48 in intestine. Cell 50: 831-840

Prakash L, Hinkle D, Prakash S (1979) Decreased UV mutagenesis in *cdc8*, a DNA replication mutant of *Saccharomyces cerevisiae*. Mol Gen Genet 172: 249-258

Rost B, Sander C, Schneider R (1994) PHD - an automatic mail server for protein secondary structure prediction. CABIOS 10: 53-60

Rothman IK, Malathi VG, Silber R (1978) Cytidine deaminase from leukemic mouse spleen. Methods Enzymol 51: 408-412

Steuart CD, Burke PJ (1971) Cytidine deaminase and the development of resistance to arabinosyl cytosine. Nature New Biol 233: 109

Stougaard J (1993) Substrate-dependent negative selection in plants using a bacterial cytosine deaminase gene. Plant J 3: 755-761

Teng B, Burant CF, Davidson NO (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science 260: 1816-1819

Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kossel H, Sugiura M (1996) Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. Proc Natl Acad Sci USA 93: 8766-8770

Weers B, Thornburg RW (1999) Characterization of the cDNA and gene encoding the *Arabidopsis thaliana*. Uracil Phosphoribosyltransferase (AF116860) Plant Physiol 119: 1567

Wentworth DF, Wolfenden R (1999) Cytidine deaminases (from *Escherichia coli* and human liver). Meth Enzymol 51: 401-407

Xiang S, Short SA, Wolfenden R, Carter CW (1995) Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. Biochemistry 34: 4516-4523

Xiang S, Short SA, Wolfenden R, Carter CW (1996) Cytidine deaminase complexed to deazacytidine: A "valence buffer" in zinc enzyme catalysis. Biochemistry 35: 1335-1341

Xiang S, Short SA, Wolfenden R, Carter CW (1997) The structure of the Cytidine deaminase - Product complex provides evidence for efficient proton transfer and ground-state destabilization. Biochemistry 36: 4768-4774

Yamanaka K, Ogura T, Niki H, Hiraga S (1992) Identification and characterization of the *smbA* gene, a suppressor of the *mukB* null mutant of *Escherichia coli*. J Bacteriol 174: 7517-7526

Zanlungo S, Quinones V, Moenne A, Holuigue L, Jordana X (1995) Splicing and editing of *rps10* transcripts in potato mitochondria. Curr Genet 27: 565-571

Zhou L, Lacroute F, Thornburg R (1998) Cloning, expression in *Escherichia coli* and characterization of *Arabidopsis thaliana* uridine 5'-monophosphate/cytidine 5'-monophosphate kinase. Plant Physiol 117: 245-253