5-2003

The Identification and Isolation of a Full Length Beta Carbonic Anhydrase cDNA in Arabidopsis thaliana

Lisa Michelle Ballard

The Identification and Isolation of a Full Length Beta Carbonic Anhydrase cDNA in *Arabidopsis thaliana*

A Thesis

Submitted to the Honors College of the
Louisiana State University and
Agricultural and Mechanical College
In partial fulfillment of the
Requirements for
Upper Level Honors

In

Biochemistry

Supervised by
Dr. Sue Bartlett

By
Lisa Michelle Ballard
May 2003

Abstract

Based on information gathered from the Arabidopsis genome database, two forms of cDNA encoding a β-carbonic anhydrase were isolated. The sequence of each form extended beyond the 5' terminus of the β-CA, AT5G14740, indicating that the cDNA isolated previously was actually truncated at the 5' end. The two cDNA samples that were isolated encoded a 331 and 291 amino acid polypeptide, respectively. The shorter form did not contain the second exon that was found in the larger cDNA, and the reading frame for the first exon was different. This indicates that this β-CA has at least two alternatively spliced forms.

I. Literature Review

Carbonic anhydrase (CA) is a metalloenzyme that catalyzes the reversible hydration of CO_2.

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+ \quad (1)$$

Examples of CA have been found in all phyla, and CA is important in many physiological functions that involve carboxylation or decarboxylation reactions, including both photosynthesis and respiration (Moroney *et al.*, 2001). In addition, CA has also been found to be involved in processes such as pH homeostasis, ion transport, renal tubule acidification, and bone resorption. (Mitsuhashi *et al.*, 2000; Hewett-Emmett and Tashian, 1996)

To date, three distinct families of CA have been identified which are designated as α–, β–, and γ– CA. All three families are zinc enzymes with one Zn^{2+} per CA monomeric unit. Another unifying characteristic among the three families is that they all catalyze the reversible hydration of carbon dioxide. However, this tends to be where the similarities between the families end. There are no significant primary sequence homologies between representatives of the three families. This lack of homology leads to the belief that each of the families is evolutionarily unrelated (Hewett-Emmett and Tashian, 1996). Evidence points to the idea that CA has evolved three different times and may represent an example of convergent evolution of catalytic function.

α–Carbonic Anhydrase

α–CA has been found in animals, plants, eubacteria, viruses, algae, and prokaryotes (Mitsuhashi *et al.*, 2000; Moroney *et al.*, 2001). So far, in vertebrates the α type is the only form of CA that has been found. In humans at least ten different isoforms of α–CA have been identified as well as three distinct CA related proteins (Moroney *et al.*, 2001)

Each α–CA isoform differs in many aspects such as catalytic activity, subcellular localization (cytoplasmic, mitochondrial, membrane bound, or secreted from the cell), tissue distribution and domain structure (Jiang and Gupta, 1999). However, the overall general structure for each of the α– isoforms is the same. They all have three histidines coordinating the zinc atom. These histidines, as well as approximately 17 other residues, are conserved across all of the isoforms (Moroney *et al.*, 2001; Lidskog, 1997). These residues are important for functions such as catalytic activity and stabilization of protein structure (Lidskog, 1997). Overall, there is a significant primary sequence homology within the CA domain which is approximately 260 amino acid residues in length (Jiang and Gupta, 1999).

The secondary structure of α–CA is dominated by the presence of a ten-stranded antiparallel β–sheet forming a spherical molecule with two halves (Lidskog, 1997). Also, there are a few short α– helices present. The active site has been described as a funnel-shaped crater located near the center of the molecule with the zinc atom located near the bottom. It is coordinated to three

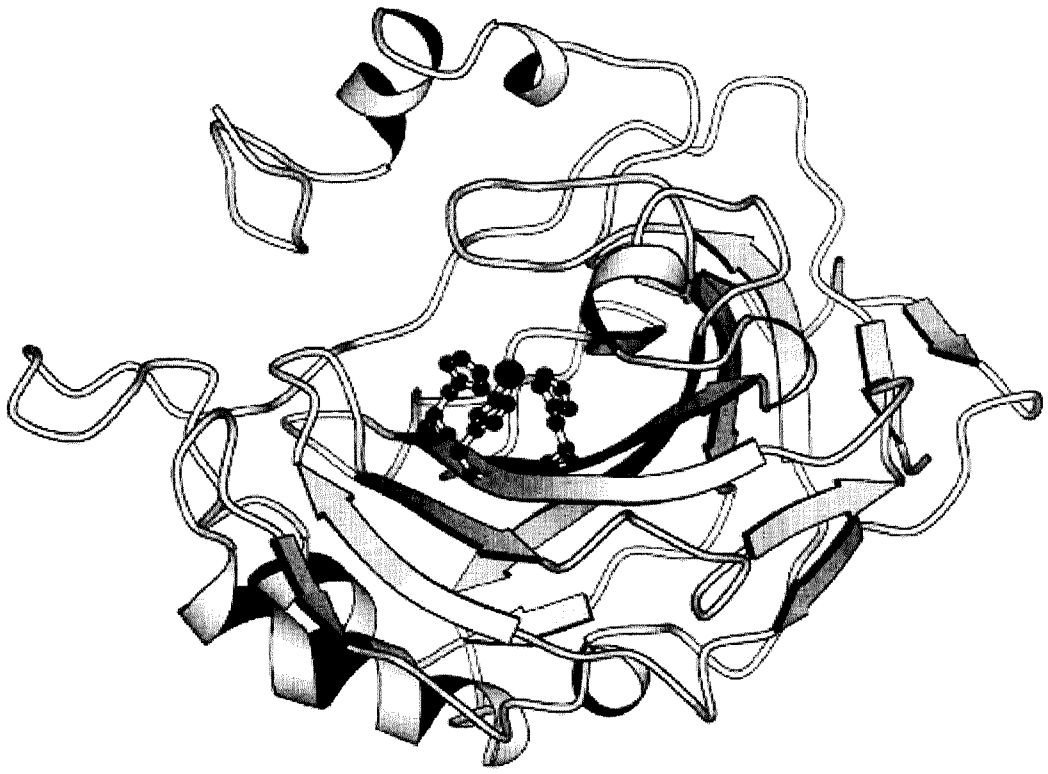

Figure 1.

Schematic drawing of the tertiary structure for human α–CAII. The zinc ion is shown in the center coordinated to the three histidyl ligands. β–strands are indicated as arrows and the α– helices as ribbons (Lidskog, 1997).

nitrogen atoms each from histidine in a tetrahedral geometry with water being the fourth ligand (Lidskog, 1997) In general α–CA is found in the monomeric form.

There are two distinct subsets of α–CA within vertebrates, the soluble isoforms (CA I, CA II, CA III, and CA VII) and the membrane-associated and secreted isoforms (CA IV, CA VI, CA XII, CA XIV, and CA XVI) (Moroney *et al.*, 2001). Each membrane-associated form contains a hydrophobic C-terminal extension. The function of this extension, which is not present in the soluble form, is to localize the particular isoform to the appropriate membrane.

The overall catalytic reaction for each of the three families of CA is the same. However, the exact mechanism of each has been found to be different. The α–CA pathway is described as a two-stage ping pong mechanism (Kimber and Pai, 2000):

$$EZn^{2+}OH^- + CO_2 + H_2O \leftrightarrow EZn^{2+}OH_2 - HCO_3^- \leftrightarrow EZn^{2+}OH_2 + HCO_3^- \quad (2)$$

$$EZn^{2+}H_2O + B \leftrightarrow EZn^{2+}OH^- + BH^+ \quad (3)$$

E represents the enzyme, and B is a buffer molecule. A zinc-bound OH^- ion attacks a CO_2 molecule to form a metal-bound HCO_3^- ion, which is then displaced by a H_2O molecule (eqn. 2). The rate limiting step is the regeneration of the zinc-bound OH^- by the transfer of a water proton from the metal center to His, which in turn delivers the proton to a buffer base (eqn. 3) (Lidskog, 1997). A schematic of the catalytic mechanism in detail is shown in Fig. 2.

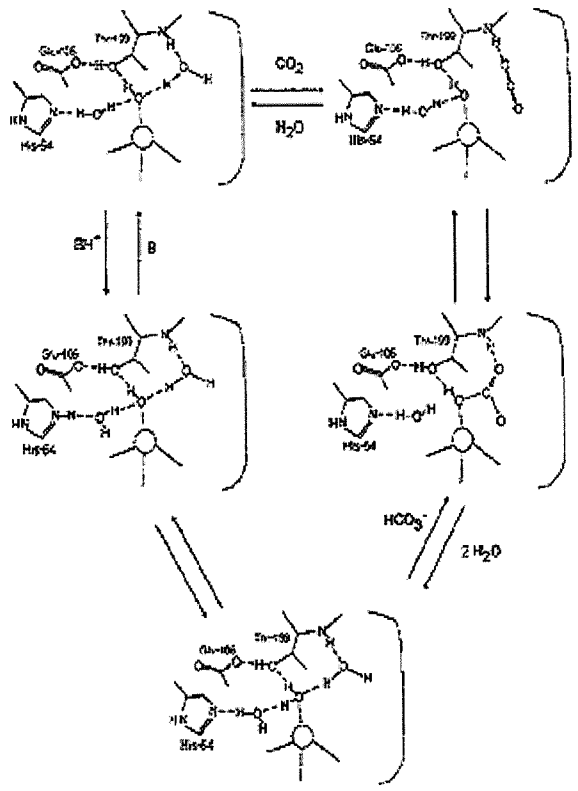

Figure 2

Catalytic mechanism of human α–CA II that includes the substrate coordinating to the Zn^{2+} and active site residues is shown. The reactions correspond to those shown in eqns. (2) and (3). The dashed lines indicate hydrogen bonds. (Lidskog, 1997)

β–Carbonic Anhydrase

Currently, β–CA's have been found in higher plants, algae, eubacteria, prokaryotes, thermophillic archaebacteria, and fungi (Moroney *et al.*, 2001; Kimber and Pai, 2000; Mitsuhashi *et al.*, 2000). Homologous sequences of uncertain function have been identified in the fungus *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans,* and *Drosophila melanogaster* (Kimber and Pai, 2000). So far no examples of β–CA have been found in vertebrates. The function of β–CA has been predominantly studied in higher plants because it is thought to play a significant role in the process of photosynthesis. The primary role in photosynthesis is to minimize the resistance to diffusion of CO_2 from the stomatal air space, where CO_2 is initially absorbed, to the chloroplast stroma where it can be available for carbon fixation by the enzyme ribulose bisphosphate carboxylase/ oxygenase (RuBisCO; Kimber and Pai, 2000).

The common structural component among the isoforms that defines β–CA is that the zinc ligands always include two cysteines and a histidine and the active site cleft is located at the subunit interface. Of the residues within the active site, only the three zinc ligands and a Asp/Arg pair are conserved in every known sequence regardless of the species. However, patterns of sequence conservation can be observed within the active site residues based on species hierarchy. This leads to the conclusion that β–CA's can be divided into two classes: the cab type and the plant type. The plant type includes β–CA's from

all higher plants, whereas the cab type is found in other organisms such as *Methanobacterium thermoautotrophicum, Mycobacterium tuberculosis*, and *Bacillus subtilis* (Kimber and Pai, 2000). The only significant difference between the two classes is in the organization of the hydrophobic pocket and the sequence conservation within the active site. However, the basic fold of β–CA is very similar for both classes.

The secondary structure of β–CA resembles neither that of α– or γ–CA. It is predominantly α– helical and is built around the core of a dinucleotide-binding fold motif with a four-stranded parallel β–sheet ordered 2-1-3-4 with α– helices forming right-handed crossover connections (Kimber and Pai, 2000; Fig. 3, Fig. 4). β–CA is usually found in dimer, tetramer, or octamer form (Strop *et al.*, 2001). The active site cleft is located at the C terminus of the parallel β– sheet and is for the most part isolated from the solvent. In the cab type β–CA, the zinc atom resides in the interface between the two monomer subunits, but all three of the coordinating residues come from the same subunit (Strop *et al.*, 2001). However, for the plant type β–CA, the active site still resides in the interface between the two monomer subunits, but the coordinating residues do not all come from the same subunit. A cystine and histidine residue comes from one monomer, and the other cystine residue originates from the adjacent monomer.

During photosynthetic carbon fixation in plants, CO_2 represents the substrate for RuBisCO, however the much more highly soluble HCO_3^- is the dominant chemical species in the chloroplast. Presumably, plant type β-CA in

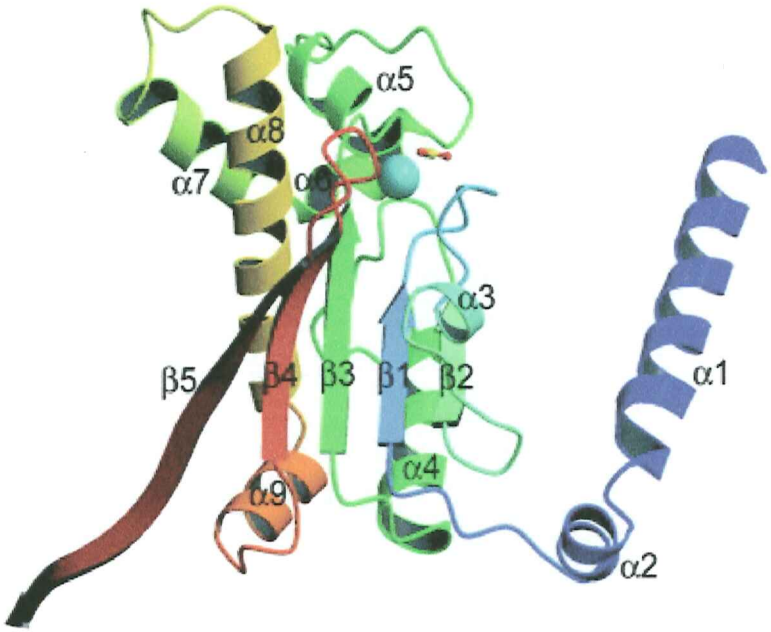

Figure 3

Ribbon diagram of the secondary structure for plant type β–CA monomer. The color is graded from blue to red, N-terminus to C-terminus (Kimber and Pai, 2000).

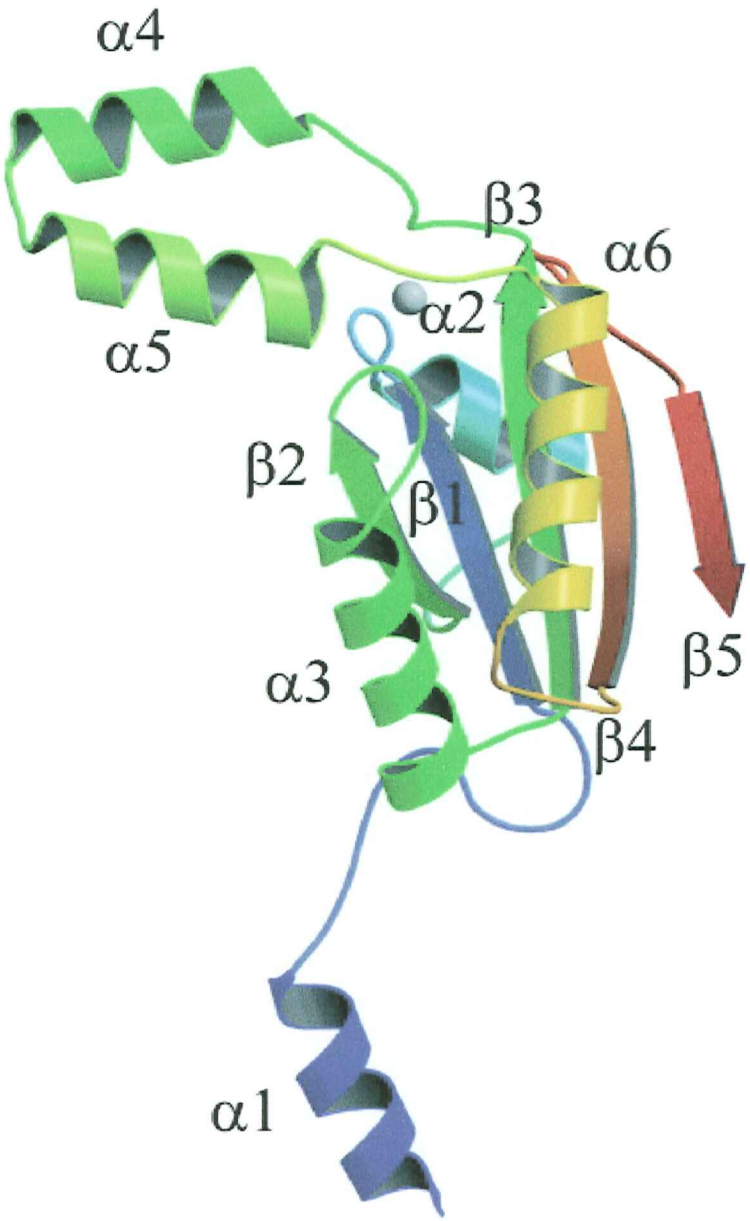

Figure 4

A diagram of the secondary structure for a Cab monomer from *Methanobacterium thermoautotrophicum*. Structurally, it is very similar to the plant type CA. Colors change from blue to red indicating N-terminus to C-terminus (Strop *et al.*, 2001).

the chloroplast stroma catalyzes the conversion of HCO_3^- to CO_2, effectively increasing the concentration of CO_2 near RuBisCO. In the archeon *Methanobacterium thermoautotrophicum* a cab type β–CA converts CO_2 to HCO_3^- which is the opposite direction from that physiologically proposed for the plant-type enzyme. This suggests the role may be to provide HCO_3^- to enzymes in the archael CO_2 fixation pathway.

One study (Kimber and Pai, 2000) found that the substrate binding groups in plant type β–CA have a one to one correspondence with the functional groups in the α–CA active site composing what is essentially a mirror image. Because of this similarity within the active site, it is likely that α– and β– CA share a similar catalytic mechanism (Rowlett *et. al.*, 2002). However, since cab-type β–CA has a completely different active site configuration as compared to the plant-type, it is possible that that the two forms also differ in their mechanisms. The proposed catalytic mechanisms for cab and plant type β–CA are shown in figures 5 and 6.

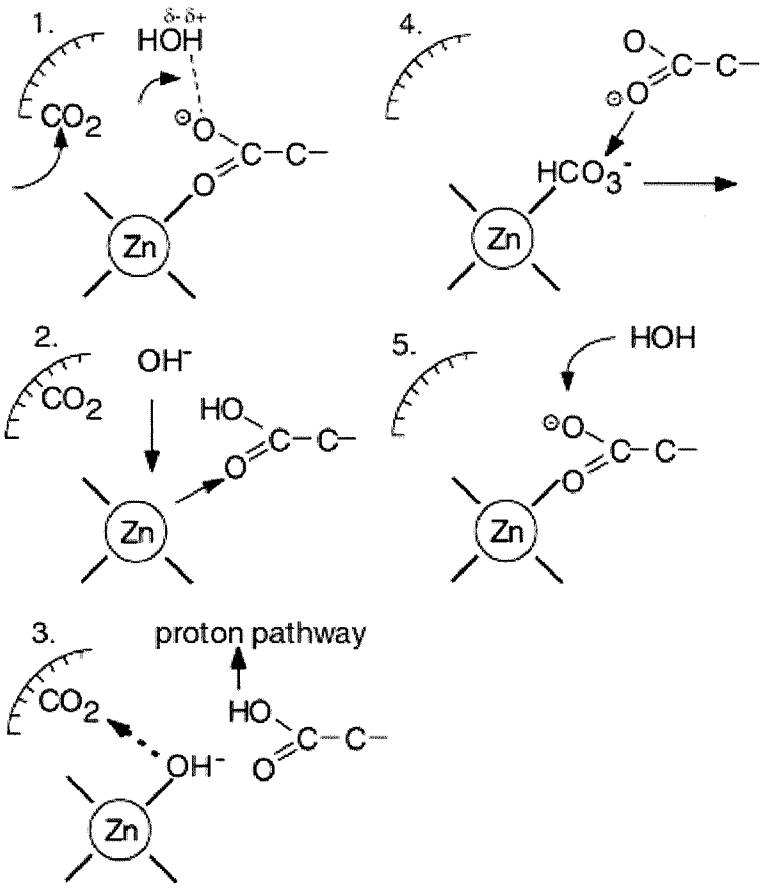

Figure 5

Shown is the proposed catalytic pathway for cab- type β–CA from the red alga, *Porphyridium purpureum* (Mitsuhashi *et al.*, 2000). One noticeable difference within the pathway is the lack of a water molecule within the zinc-ligand binding radius. The CO_2 association into the active site triggers the subsequent reactions (step 1).

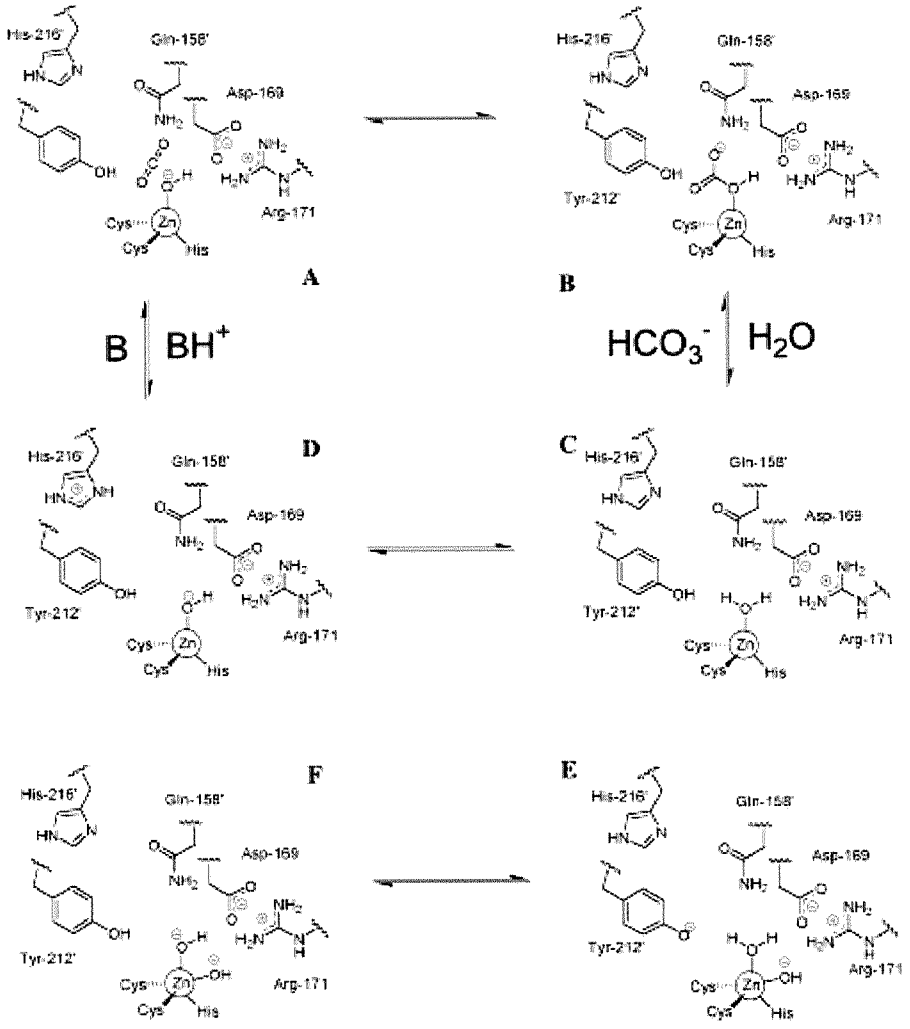

Figure 6

Shown is the summary of the proposed mechanism for *Arabidopsis thalinana* β–CA (Rowlett *et al.* 2002). *A. thalinana* represents a plant-type CA. Characteristics of this mechanism include a zinc-bound hydroxide ion that is the nucleophile for CO_2 hydration. Based on the pH, the enzyme will follow path ABCDA under acidic conditions and path ABEFA under basic conditions.

γ–Carbonic Anhydrase

γ–CA was first discovered in the archaebacterium, *Methanosarcina thermophila* (Alber and Ferry, 1994). Genes encoding putative proteins with similar sequences have been found in eubacteria and in plants. For example, at least three γ–CA cDNA's have been isolated from *A. thaliana* (Nielson III, 2001).

When the γ–CA was first isolated from *M. thermophila*, comparision of the deduced primary structure with protein databases and alignment to known CA's indicated that this enzyme is distinct from all previously described CA's, suggesting another class. γ–CA is found as a trimer of identical subunits. The secondary structure of each subunit is dominated by seven turns of a left-handed β– helix. The zinc ions are located between the subunits, and are coordinated to two histidines from one subunit and one from another. Even though the active site is shared by two different subunits, it still has architecture similar to that of α–CA.

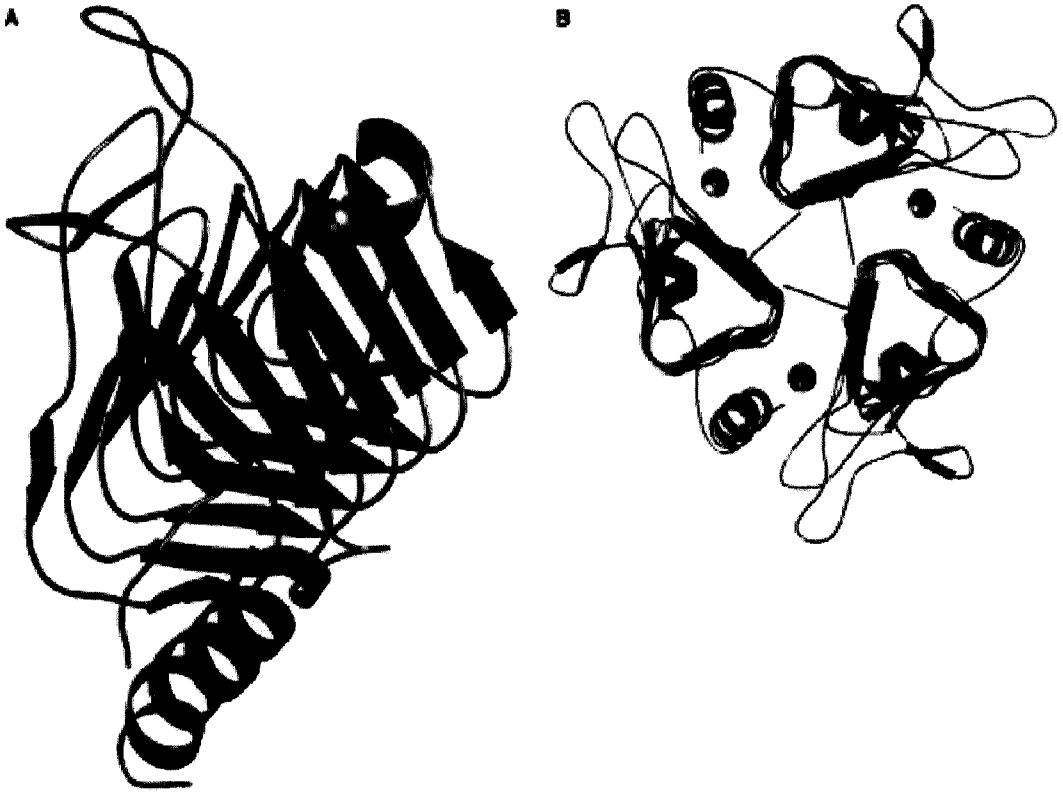

Figure 7

Shown is the drawing of the structure of γ–CA from *M. thermophila*. A). A side view of the monomer. B). A cross section of the trimer along the molecular 3-fold axis. The zinc ions are visible at the interfaces between each of the monomers. (Lidskog, 1997)

The exact catalytic mechanism for γ–CA has not been thoroughly investigated. However, since the active site is similar to that of α–CA, one could assume that the mechanism would also be similar. The exact function of γ–CA in *M. thermophila* has not been determined. However, a metabolic switch from methanol to acetate results in the elevation of γ–CA activity suggesting that this enzyme is important for growth on acetate (Alber and Ferry, 1994).

Uncharacterized Carbonic Anhydrase's

Recent studies have shown that there may be a potential for additional classes of CA. One study using *Thalassiosira weissflogii* claimed to have isolated a cDNA that encoded a CA, but its sequence did not match any known CA from the three gene families (Moroney *et al.*, 2001). However, the deduced protein sequence did not appear to have any zinc binding residues which is a common feature in all of the confirmed CA families. In addition, they have not yet demonstrated that the expressed protein has any CA activity. These factors lead to questioning of the validity of their results and whether or not there really is an additional CA family.

Identifying novel isoforms within the α–, β–, and γ– CA families

New studies are regularly being released identifying new isoforms of CA within various organisms. There are definitely numerous potential

isoforms of the three defined groups that have yet to be defined in *Arabidopsis*. In *Arabidopsis*, a few of these putative CA genes are known only from genomic sequences, but most of them are clearly expressed and have associated expressed sequence tags. So far based on the genomic sequence 7 α–CAs, 6 β–CAs, and 3 γ–CAs have been identified. However, ESTs have only been found for the β, γ, and half of the α sequences. This suggests that not all of these isoforms are actively expressed. Also, in only a few cases has the enzymatic activity actually been measured. The goal of this project was to identify and isolate the cDNA for a specific β–CA from *Arabidopsis* based on the sequence projections obtained from the genomic sequence database.

II. Materials and Methods

Isolation of RNA from *Arabadopsis thaliana*

Isolation of RNA from young *Arabadopsis thaliana* whole plants was performed using the RNAqueous small scale phenol-free total RNA isolation kit (Ambion). Approximately 75 mg of plant tissue was used. Complete disruption of the tissue was obtained by grinding the plant tissue in liquid nitrogen until homogenized. Once the RNA had been isolated with the kit, it was precipitated with lithium chloride to remove carbohydrate and DNA contaminants. A final purification step was performed on the RNA sample using the RNeasy Mini RNA Clean-up protocol (Qiagen). Final concentration of the RNA was 392.7 ng/μl.

Amplification of β-CA3 using Reverse Transcription Polymerase Chain Reaction (RT-PCR)

The total RNA was used as a template source to amplify the β-CA3 cDNA. The 5' primer (CTTATGGTCCCCTTTTGGACTACAG) and the 3' primer (CTTATATACTTGATTGTGTAGTGGTAG) were obtained from Qiagen. The 5' primer contained the start codon (ATG) at the beginning of the β-CA3 open reading frame. The 3' primer was actually designed in the reverse direction. However, the TATATA portion of the sequence did match part of the 3' end of the β-CA3 sequence and was enough to obtain the correct product. The RT-PCR reaction was set up using the ProSTAR HF Single-Tube RT-PCR System (Stratagene) as follows: 10x HF RT-PCR buffer, 100ng of 5' primer, 100ng 3'

primer, 40 mM dNTP mix, 4μl of isolated RNA (approximately 1500ng), 1.25 units Stratascript reverse transcriptase, and 2.5 units *TaqPlus Precision* DNA polymerase. The final volume was brought to 50 μl using RNase-free water.

The product was amplified using the PTC-100 Programmable Thermal Controller (MJ Research) with the following protocol:

1. 90 mins at 42 °C
2. 1 min at 95 °C
3. 30 secs at 95 °C*
4. 30 secs at 59 °C*
5. 3 mins at 68 °C*
6. 10 mins at 68 °C
7. HOLD at 4 °C

*Cycles 3-5 are repeated 40 times

The DNA yield from the reaction was determined by electrophoresis using a 1% agarose gel in 1X TBE (90 mM Tris-borate, 2mM EDTA, pH 8.0).

PCR amplification of β-CA3

A PCR amplification of β-CA3 was performed using the RT-PCR product obtained previously as the template. The reaction was set up as follows: 10x cloned Pfu reaction buffer, 25mM each dNTP, 100ng 5' primer, 100ng 3' primer, 2.5 units *PfuTurbo* DNA polymerase, and 1 μl RT-PCR product. The final volume for the reaction mixture was brought up to 50 μl using double distilled water. The product was also amplified using the PTC-100. The protocol was set as follows:

1. 2 mins at 95 °C
2. 30 secs at 95 °C*
3. 30 secs at 59 °C*
4. 2 mins at 72 °C*
5. 10 mins at 72 °C

6. HOLD at 4 °C
*Steps 2-4 were repeated 30 times.

The DNA yield from the reaction was determined by electrophoresis using a 1% agarose gel in 1X TBE (90 mM Tris-borate, 2mM EDTA, pH 8.0).

A-Tailing Procedure

The A-tailing procedure was used to add a deoxyadenosine nucleoside tail, in a template independent fashion, to the 3' ends of the PCR product. This A- overhang facilitated cloning into a pGEM-T Easy vector which contains a 5' T-overhang (Promega). First, 30 μL of the β-CA3 product was purified by adding ½ volume of 7.5 M ammonium acetate. The sample was centrifuged at 14,000 rpm for 5 minutes and the supernatant was retained. Then, 2 volumes of 100% ethanol were added, and the sample was set on ice for 10 minutes. After centrifugation at 14,000 rpm for 10 minutes, the supernatant was removed. The pellet was washed with 50 μL of 70% ethanol, and recentrifuged for 5 minutes. After removal of the supernatant, the pellet was air dried.

Once the pellet had dried, it was resuspended in 15.4 μL of double distilled H_2O. The A-tailing reaction was set up as follows: 10x thermophilic DNA polymerase buffer, 1 unit Taq DNA Polymerase, 2.5 mM $MgCl_2$, 0.2 mM dATP. The final volume of the reaction was 20 μL. The reaction mixture was incubated at 70 °C for 30 minutes. The product was run out on a 1% low melt agarose gel. The bands representing the product were excised from the gel.

Ligation of β-CA3 into pGEM-T- easy Vector

The low melt gel was heated at 65 °C for 10 minutes to melt it. The pGEM-T Easy Vector System I (Promega) was used for the ligation reaction. Based on the concentration of the band on the low melt gel, the ligation reaction was set up with a ratio of approximately 2:1 for the insert and vector. The reaction was prepared as follows: 50ng pGEM-T Easy Vector, β-CA3 gel insert, 1x rapid ligation buffer, and 3 units T4 DNA ligase. The final volume of 20 μL was reached using double distilled H_2O. The reaction was incubated overnight at 16 °C.

Transformation into *Escherichia coli* DH5α compentent cells

The β-CA3 gene in the pGEM-T Easy vector was transformed into *Escherichia coli* DH5α competent cells. The ligation reaction was heated to 65 °C for 5 minutes to make sure that the agarose was melted. It was then diluted 10-fold with double distilled H_2O. 50 μL of competent cells plus 5 μL of the diluted ligation reaction were added to a cold 0.2 cm electroporation cuvette. The cells were shocked at 2.5 kV using the Bio-Rad Gene Pulser II. Cells were plated on LB agar containing ampicillin (30 ug/ml) and X-gal (40mg/ml) for selection of colonies containing the insert.

Plasmid DNA isolation

Eight colonies were chosen from the agar plates and grown overnight in 3mL LB media containing 30μg/mL ampicillin. The alkaline lysis method (Maniatis *et al.*, 1982) was used as a rapid, small scale isolation of the plasmid DNA from *E.coli*. The contents of the miniprep samples were checked to

see if they contained the PCR insert by using both an EcoR I restriction digest for 1 hour and a Sac I restriction digest overnight.

Preparing samples for sequencing

Using samples 1 and 2, six 3mL LB media overnight tubes were inoculated to culture additional *E. coli* cells. The plasmid DNA was isolated using the alkaline lysis method as before. However, for the final step the DNA pellet was reconstituted in half of the specified amount of TE. The Plasmid Preparation Protocol (Promega, 1989, pg 14) was used to purify the samples. This protocol allows for rapid isolation of plasmid DNA without needing to use a column or CsCl gradients. The samples were sent to the Pennington Biomedical Research Center for sequencing.

III. Results and Discussion

Identification of potential CAs

The characterization of a cDNA sequence for a β–CA from *Arabidopsis* was reported by Fett and Coleman in 1994. This sequence is located on chromosome 5 with accession number AT5G1470. However, based upon the *Arabidopsis* genomic database, it was predicted that this sequence was actually a truncated form. It appeared that additional sequences upstream of the 5' end should have been included within the cDNA. In addition, ESTs were identified that supported the inclusion of these upstream portions as part of the sequence.

Based upon this information, the purpose of this study was to identify and characterize the full length sequence for the β–CA. Specific primers were designed that incorporated the upstream sequence. If this sequence is indeed transcribed as part of the mRNA, a product should be able to be obtained which will be referred to as β–CA3.

Amplification of the βCA3 cDNA

Based on the proposed CA sequence, the next step was to determine if the specific gene sequence was being actively transcribed and expressed in *Arabidopsis*. This was accomplished by using RT-PCR to amplify the cDNA. RT-PCR is the PCR amplification of a reverse transcription product. To begin the procedure a reverse transcriptase is used to copy the RNA to DNA. Then, the DNA polymerase can make numerous copies of the target sequence.

This allows for amplification of very small amounts of RNA to thousands or millions of copies allowing it to be plentiful enough for sequencing. Therefore, if the gene is being expressed even in small amounts, it should still be able to be amplified.

RNA used as a template source in the RT-PCR reactions was extracted from Arabidopsis plants. Young plants were chosen because they were more likely to have a higher expression level of CA. The Ambion RNAqueous kit was used to isolate the RNA from the plants. Table 1 shows that this procedure resulted in a very high yield of RNA. The A_{260}/A_{280} ratio gives a good estimate of the purity of the sample with an expected range between 1.8 and 2.0. The Ambion procedure yielded a ratio of 1.18 indicating that there were some contaminants present. Further purification of the RNA was carried out using LiCl precipitation and the RNeasy clean-up kit. After the final step the final concentration of the sample was 392.7 μg/mL indicating some product loss. However, the A_{260}/A_{280} ratio was 2.18 showing a significant reduction in the contaminant level, and the sample was used in the RT-PCR reactions.

Figure 8 shows the results of the RT-PCR reaction using primers based on the β–CA3 sequence in the database. The expected size for the β–CA3 was 996 base pairs. On the gel, a doublet band was visualized at approximately 1Kb. The presence of a double band was unexpected. However, because both bands were very similar to the expected size, both were chosen for cloning.

Procedure	A_{260} (nm)	A280 (nm)	A_{260}/A_{280}	Conc. (μg/mL)
Ambion Kit	1.080	0.918	1.18	2590.9
LiCl clean-up	0.225	0.248	0.91	540.6
Rneasy clean-up	0.164	0.117	2.18	392.7

Table 1. Absorbance readings and calculation of concentration for *Arabidopsis* RNA after each isolation and purification procedure.

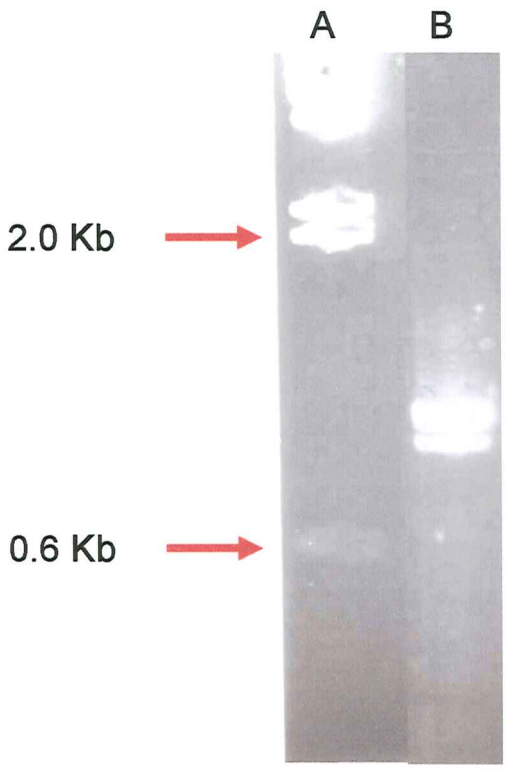

Figure 8. Agarose gel electrophoresis analysis of the RT-PCR product. A) Lamba Hind III DNA marker. B) RT-PCR product using primers specific to βCA3.

The DNA product obtained from the RT-PCR was used as a template for PCR in order to obtain enough DNA to use for cloning. The PCR should show the same pattern of bands as the RT-PCR. Figure 9 shows that once again a doublet of bands is visible in the 1kb range.

The reaction catalyzed by *Pfu Turbo* DNA polymerase in the previous PCR made DNA copies that were blunt ended. In order to ligate the PCR product into a pGEM-T Easy vector, "sticky ends" were added by A-tailing the products. The A-tailing procedure allows for the addition of an adenosine nucleotide to only the 3' end of the PCR product which creates a "sticky end". The addition of just a single nucleotide is sufficient to allow for ligation to the T-tailed plasmid.

Ligation into the pGEM-T- easy vector

Before the A-tailed PCR products were used in a ligation reaction, they were purified by electrophoresis on a low melt agarose gel. The two bands together were excised from the gel. DNA ligase was used to catalyze the reaction to join the PCR product and the plasmid together. This enzyme catalyzes the formation of phosphodiester bonds between the adjacent nucleotides (Chang SH, *et al.* 1996).

Ligation reactions were set up with at least a 2:1 or 3:1 molar ratio of insert to vector to help increase the efficiency of the reaction. A higher concentration of insert is used to push the reaction in favor of binding the insert to the vector as opposed to any side reaction such as binding the vector to

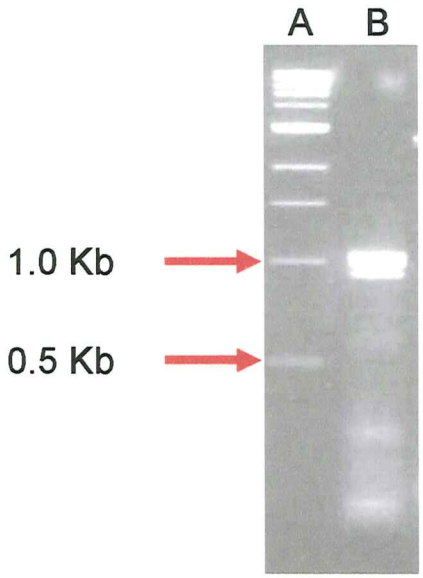

Figure 9. Agarose gel electrophoresis analysis of the PCR product. A)1Kb DNA marker. B) PCR product using primers specific to βCA3.

another vector. The concentration of the PCR product was estimated based on the band intensity from the low melt gel.

Transformation into DH5α cells

The newly formed plasmid that contains the PCR insert was transformed into *E. coli* competent cells. The transformation process was carried out using electroporation which causes the competent cells to take up the foreign plasmids. Electroporation causes the lipid membrane matrix to be disrupted by using a strong external electric field which leads to an increase in transmembrane conductivity and diffusive permeability (Tsong, 1991).

The cells were plated on LB agar with ampicillin. The pGEMT-Easy vector has a gene that encodes resistance to ampicillin (Fig. 10). This helps the screening process because only cells that take up the vector should be able to survive in the presence of the antibiotic.

The pGEMT-Easy vector also contains an additional screening method because it contains the *lacZ* gene which encodes β-galactosidase (Fig 10). This is an enzyme that breaks down lactose into glucose and galactose. However, X-GAL can be used as a substrate, and its breakdown product is blue. The lacZ gene is located within the multiple cloning site. Disruptions at this site due to insertion of the PCR product would prevent the production of β-galactosidase. The β–galactosidase substrate, X-GAL was included in the media.

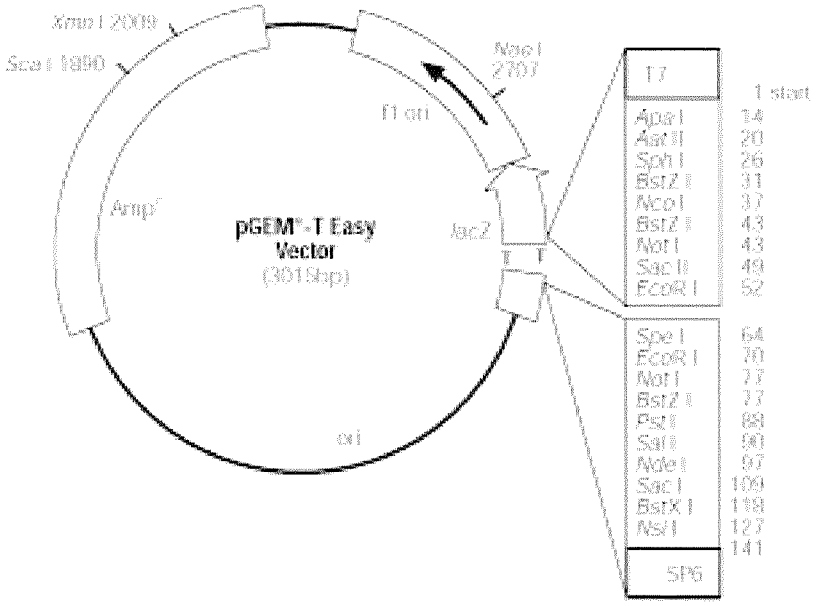

Figure 10. pGEM-T Easy Vector map and sequence reference points (Promega).

Plasmids lacking inserts were blue, while those containing inserts were white. Therefore, a screen based on the color of the colonies can be performed to determine which colonies contain the plasmid that has an insertion in the cloning site.

Based on these screening methods discussed, eight colonies were chosen. Cultures of each were grown overnight, and the alkaline lysis method was used to isolate the plasmid DNA. The pGEM-T Easy vector has an EcoR I site on each side of the cloning region. Based on the sequence for β-CA3 in the database, it does not contain any EcoR I restriction sites. Therefore, a restriction digest using EcoR I should excise the PCR insert in its entirety from the plasmid. Figure 11 shows the results of the digest. Two different sized bands were observed. This corresponds to the doublet of bands from the PCR, and apparently, a representation of each has been recovered. Samples 1, 3, 4, and 6 contain the larger band, and samples 2, 5, and 8 contain the lower band. No plasmid product was observed from sample 7.

To confirm that the insert is correct, a restriction digest using Sac I was performed. Based on the restriction site map, this enzyme should produce a 500 bp fragment if in the reverse orientation within the plasmid. It is unlikely to find CA in the forward orientation because any potential expression of the enzyme would be toxic to *E. coli*. Figure 12 shows the result of the digest. All of the samples have a 500 bp fragment, indicating that the cDNAs likely represent the β–CA3.

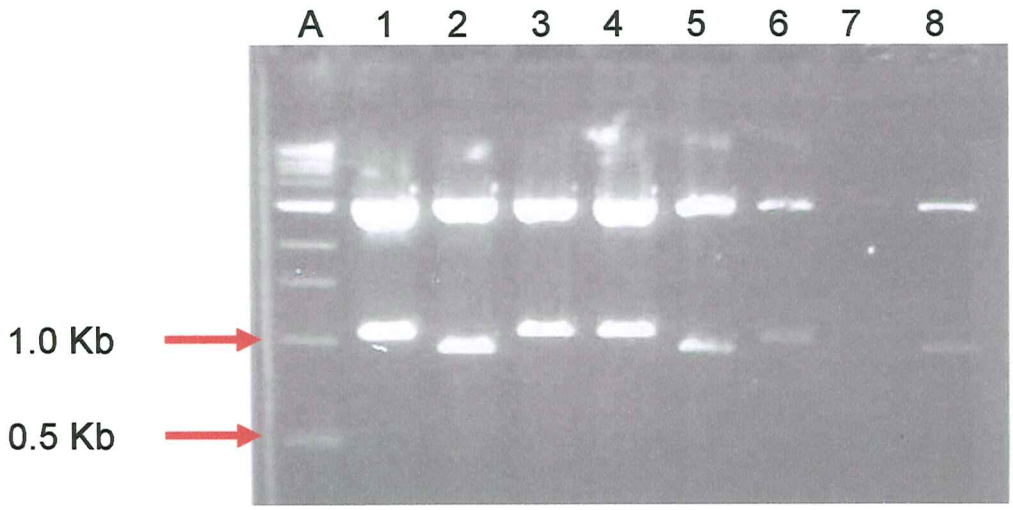

Figure 11. Agarose gel electrophoresis of the plasmids digested with EcoR I. A) 1 Kb DNA marker. 1-8) Digestion products from each of the eight colonies.

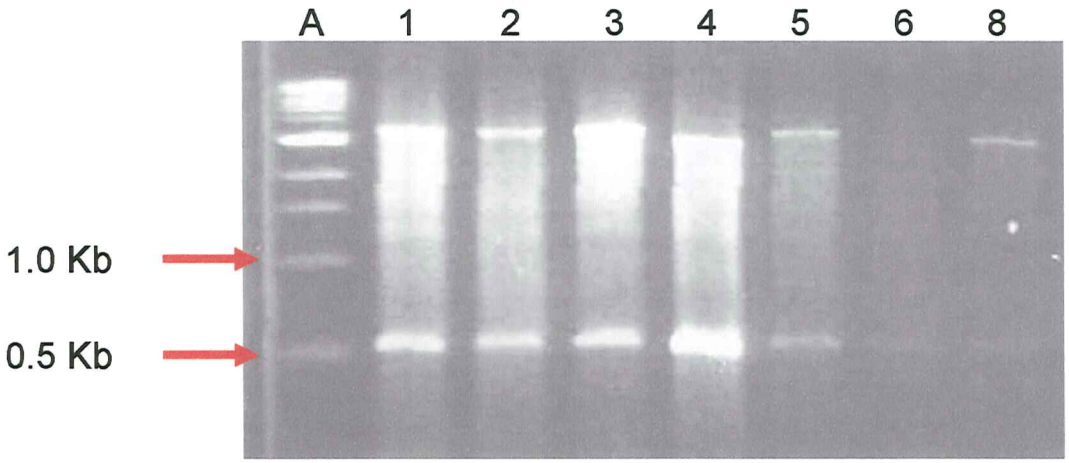

Figure 12. Agarose gel electrophoresis of the plasmids digested with Sac I. A) 1 Kb DNA marker. 1-6,8) Digestion products from each of the eight colonies.

Sequencing

Samples 1 and 2 were chosen as representatives of the larger and smaller sized bands. Additional plasmids containing each insert were prepared, purified, and concentrated for DNA sequencing. The results from the sequencing are shown in Figure 13 and 14. These figures show the location of the sequence for each of the samples compared to the genome database. The orange labeled nucleotides in the figures represent the previously determined β–CA sequence. Based on the results I obtained, the sequence does extend upstream of the reported 5' terminus. This confirms that the cDNA characterized by Fett and Coleman was indeed a truncated version.

There were some observed differences between the sequences in the two samples. Sample 1 contains 10 exons and 9 introns (Fig 13). The sizes of the exons are: 87 bp, 118 bp, 64 bp, 150 bp, 49 bp, 117 bp, 54 bp, 86 bp, 106 bp, and 165 bp which encode a protein containing 331 amino acids. Sample 2 contains only 9 exons and 8 introns (Fig 14). The sizes of the exons are: 87 bp, 64 bp, 150 bp, 49 bp, 117 bp, 54 bp, 86 bp, 106 bp, and 165 bp which encodes a protein containing 291 amino acids. Sample 2 does not contain the second exon that is found in Sample 1. The lack of this exon could indicate an alternatively spliced form of the β–CA. The function of alternative splicing could involve different targeting locations for the β–CA. In addition, the lack of the second exon in Sample 2 causes the reading frame to shift so that the initial ATG does

Figure 13. The location of the nucleotide sequence of Sample 1 within the genomic sequence on chromosome 5. ATG = Translational Start/Stop, ATGC = Exon, atgc = UTR, atgc = Intron, atgc=sequenced sample.

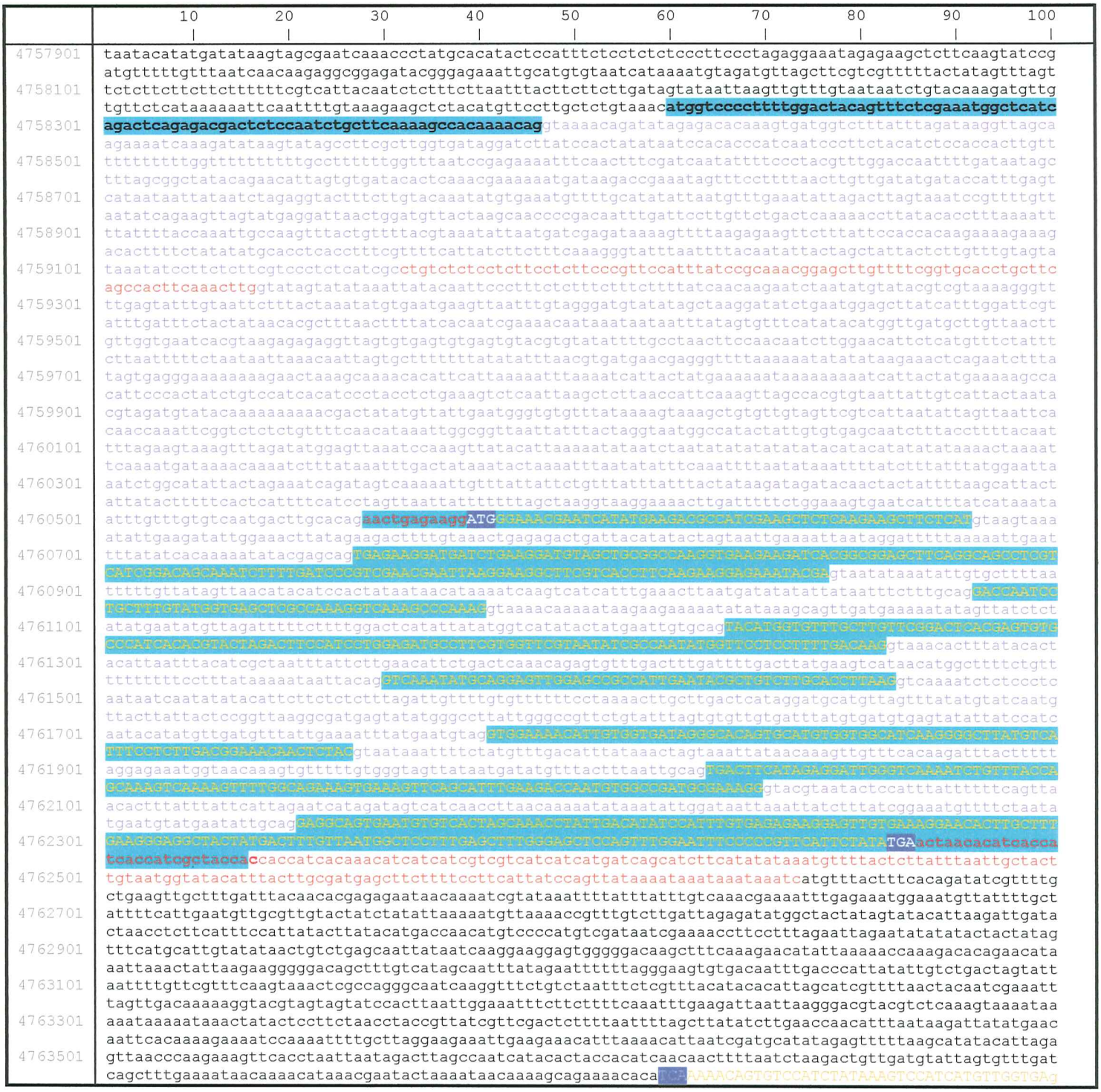

Figure 14. Shown is the location of the nucleotide sequence of Sample 2 within the genomic sequence on chromosome 5. ATG = Translational Start/Stop, ATGC = Exon, atgc = UTR, atgc = Intron, atgc=sequenced sample

not encode a methionine. This indicates that the sequence may extend even further upstream than what is reported here.

The first intron in sample 2 (second in sample 1) is considerably larger than the typical intron and is approximately 1Kb in size for both samples. When such a large intron is present, it is usually an indication that the leading exon is a targeting sequence. In *Arabidopsis* there is usually approximately only a 1Kb separation between adjacent genes on the genomic DNA. The presence of such a large intron perhaps contributed to the fact that the potential for a longer transcript was not identified in the analysis of the Arabidopsis DNA sequence.

The amino acid sequences of Sample 1 and 2 were aligned to that of the previously characterized β–CA (AT5G1470) as well as 5 other β–CA sequences that have been identified in Arabidopsis. Figure 15 shows the result of the alignment using ClustalW (1.8). Using the Genestream align tool, the percent identity of the amino acid sequence for Sample 1 and Sample 2 was calculated as compared to that of the other β–CAs (Table 2). The identity was the highest for At3g01500 with 68.4 and 66.9% respectively.

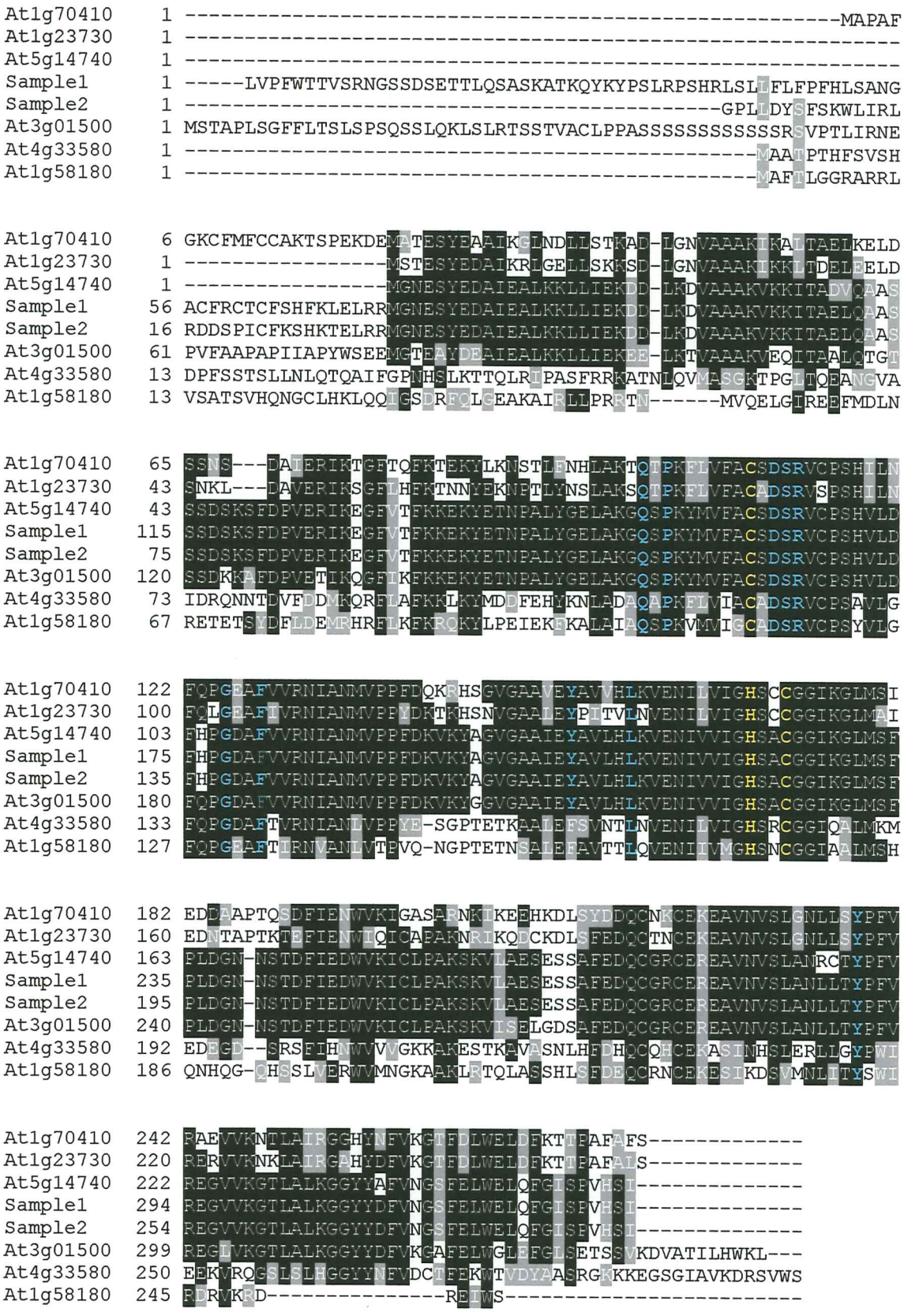

Figure 15. Alignment of Sample 1 and 2 to other β-CAs in Arabidopsis. Black shading indicates the residue is identical to the column-consensus. Gray shading indicates the residue is not identical but at least similar to the column-consensus. The zinc coordinating residues are shown in yellow and the remaining active site residues are shown in blue (Mitsuashi, *et al.* 2000).

β–CA Identification #	Sample 1 % Alignment	Sample 2 % Alignment
AT1G70410	49.8	55.6
AT1G23730	45.6	51.9
AT1G58180	28.4	29.6
AT5G14740	76.7	87.3
AT3G01500	68.4	66.9
AT4633580	34.7	35.9

Table 2. Percent Alignment between Sample 1, 2 and other β–CAs identified in *Arabidopsis*.

IV. Future Directions and Conclusion

Further studies with β–CA3 are planned. First, the activity of both of the samples should be assayed to determine if the cDNA produces functional protein and also to determine if there are any differences in the activity level between the two samples. In order to assay the the CA activity, the samples must be cloned into an expression vector such as the plasmid pET28b. This plasmid has a T7 promoter that allows for overproduction of the protein. The promoter can be controlled due to the presence of the Lac I repressor. When it is the appropriate time for protein expression the repressor can be released by adding isopropyl-beta-D-thiogalactopyranoside (IPTG) allowing initiation of the transcription of the gene.

In addition, studies can be performed to determine the expression patterns of the two alternate forms of β–CA3. This can be accomplished by GFP (Green Fluorescent Protein) fusions to the cDNA. The expression pattern and subcellular localization can be visualized due to the strong fluorescence of GFP. By using the GFP, it allows the identification of the location of the specific β–CA3 isoform distinct from the other CA isoforms.

This study outlines a procedure that can be applied to finding other CAs in *Arabidopsis* and other organisms. Using the genomic database, potential CAs can be identified and primers designed based upon the proposed sequence. Already to date, I have also successfully amplified an α-CA by RT-PCR. In addition two other β–CA cDNAs as well as 3 γ–CA cDNAs have been isolated by members of this laboratory. There are potentially many CAs that could be

successfully identified and characterized based upon this method, and I expect in the near future that cDNA for new isoforms will be reported.

References

Alber BE, Ferry JG. (1994) A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. Proc. Natl. Acad. Sci. **91**, 6909-6913.

Chang SH, Ding H, Klotz AV. (1996) Biological Sciences 4385 Biochemistry Laboratory Manual. Louisiana State University.

Fett JP, Coleman JR. (1994) Characterization and expression of two cDNAs encoding carbonic anhydrase in Arabidopsis thaliana. Plant Physiol. **105**, 707-713.

Hewett-Emmett D, Tashian RE. (1996) Functional diversity, conservation, and convergence in the evolutionary of the α–, β, and γ–carbonic anhydrase gene families. Mol. Phylogen. and Evol. **5,** 50-77.

Jiang W, Gupta D. (1999) Structure of the carbonic anhydrase VI (CA6) gene: evidence for two distinct groups within the α–CA gene family. Biochem. J. **344,** 385-390.

Kimber MS, Pai, EP. (2000) The active site architecture of *Pisum sativum* β–carbonic anhydrase is a mirror image of that of α–carbonic anhydrase. EMBO. **19,** 1407-1418.

Lindskog S. (1997) Structure and Mechanism of Carbonic Anhydrase. Pharmacol. Ther. **74,** 1-20.

Maniatis T, Fritsch EF, Sambrook J. (1982) Molecular Cloning, A laboratory manual. Cold Springs Harbor Laboratory. pgs. 368-369.

Mitsuhashi S, Mizushima T, Yamashita E, Yamamoto M, Kumasaka T, Moriyama H, Ueki T, Miyachi S, Tsukihara T. (2000) X-ray structure of β–carbonic anhydrase from the Red Alga, *Porphyridium purpureum*, reveals a novel catalytic site for CO_2 hydration. J. Biol. Chem. **275,** 5521-5526.

Moroney JV, Bartlett SG, Samuelsson G. (2001) Carbonic anhydrase in plants and algae. Plant Cell and Environ. **24**, 141-153.

Nielsen III ETD. (2001) The examination of the expression patterns of a gamma carbonic anhydrase homologue in *Arabidopsis thaliana*. Honors College Thesis, Louisiana State University, Baton Rouge, LA.

Rowlett RS, Tu C, McKay MM, Preiss JR, Loomis RJ, Hicks KA, Marchione RJ, Strong JA, Donovan, Jr. GS, Chamberlin JE. (2002) Kinetic characterization of wild-type and proton transfer-impaired variants of β–carbonic anhydrase from *Arabidopsis thaliana*. Archives of Biochem. and Biophys. **404,** 197-209.

Strop P, Smith KS, Iverson TM, Ferry JG, Rees DC. (2001). Crystal structure of the "cab"-type β class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*. J. Biol. Chem. **276,** 10299-10305.

Tsong, TY (1991) Electroporation of cell membranes. Biophys J. **60**, 297-306.