The Journal of Experimental Biology 207, 4559-4572
Published by The Company of Biologists 2004
doi:10.1242/jeb.01287

A GPI-linked carbonic anhydrase expressed in the larval mosquito midgut

Terri J. Seron[1,2], Jennifer Hill[1] and Paul J. Linser[1,2,*]

[1]*The Whitney Laboratory* and [2]*Department of Fisheries and Aquatic Sciences, University of Florida, Saint Augustine, FL 32080, USA*

*Author for correspondence (e-mail: pjl@whitney.ufl.edu)

Accepted 20 September 2004

Summary

We have previously described the first cloning and partial characterization of carbonic anhydrase from larval *Aedes aegypti* mosquitoes. Larval mosquitoes utilize an alkaline digestive environment in the lumen of their anterior midgut, and we have also demonstrated a critical link between alkalization of the gut and carbonic anhydrase(s). In this report we further examine the nature of the previously described carbonic anhydrase and test the hypothesis that its pattern of expression is consistent with a role in gut alkalization. Additionally we take advantage of the recently published genome of the mosquito *Anopheles gambiae* to assess the complexity of the carbonic anhydrase gene family in these insects. We report here that the previously described carbonic anhydrase from *Aedes aegypti* is similar to mammalian CA IV in that it is a GPI-linked peripheral membrane protein. *In situ* hybridization analyses identify multiple locations of carbonic anhydrase expression in the larval mosquito. An antibody prepared against a peptide sequence specific to the *Aedes aegypti* GPI-linked carbonic anhydrase labels plasma membranes of a number of cell types including neuronal cells and muscles. A previously undescribed subset of gut muscles is specifically identified by carbonic anhydrase immunohistochemistry. Bioinformatic analyses using the Ensembl automatic analysis pipeline show that there are at least 14 carbonic anhydrase genes in the *Anopheles gambiae* genome, including a homologue to the GPI-linked gene product described herein. Therefore, as in mammals which similarly possess numerous carbonic anhydrase genes, insects require a large family of these genes to handle the complex metabolic pathways influenced by carbonic anhydrases and their substrates.

Key words: GPI-linked carbonic anhydrase, mosquito, larva, *Aedes*, *Anopheles*, muscle, alkaline gut.

Introduction

Some larval insects such as mosquitoes and some caterpillars are known to possess a highly alkaline digestive system (Dadd, 1975). Berenbaum's review (Berenbaum, 1980) of Lepidopteran insects reported gut pH values ranging from 7.0 to 10.3, through the influence of diet. Caterpillars feeding on leaves containing tannins were found to display a more alkaline pH (average pH 8.76) than those feeding on low tannin diets (average pH 8.25; Berenbaum, 1980). The alkaline gut is believed to disrupt and inhibit the formation of insoluble tannin–protein complexes. These complexes could potentially interfere with insect digestion by blocking the active sites of many different digestive enzymes. Therefore, the alkaline gut serves as a hypothetical benefit to the larval insects by solublizing proteins in a tannin-free form. The advantage of the alkaline digestive strategy is clear; however, the mechanism of the high alkalinity is not.

The larval mosquito gut is known to perform both digestive and assimilation functions (Clements, 1992). pH values ranging from 7 to 11 (Dadd, 1975) along the length of the mosquito gut are presumed to support these different functions. Some regions of the mosquito gut are known to be associated with various functions. The gastric caeca are responsible for ion and water transport (Clements, 1992). The anterior midgut is responsible for alkaline digestion, while the posterior midgut absorbs nutrients (Clements, 1992). The Malpighian tubules actively transport potassium and fluid (Clements, 1992). The midgut contents of larval *Aedes aegypti,* the mosquito known to spread yellow fever, have a pH as high as 11 in the anterior portion of the gut while the two adjacent gut regions, the gastric ceaca and posterior midgut, have a pH close to 8 (Zhuang et al., 1999). Carbonic anhydrases (CAs) catalyze the reversible hydration of carbon dioxide (CO_2) to bicarbonate (HCO_3^-) and therefore are predicted to function within the anterior midgut region that surrounds the most alkaline pH. However, it has been shown that a CA activity is present in the gastric caeca and posterior midgut, but not the anterior midgut (Corena et al., 2002). Although a CA enzymatic activity was not localized within the anterior midgut cells, acidification (i.e. inhibition of alkalization) of the anterior midgut lumen was observed upon incubation with a CA-specific sulfonamide inhibitor (Corena et al., 2002). This result indicates that CA activity is indeed involved in the alkalization of the mosquito

gut lumen. Further studies of mosquito CA isoforms are being pursued to better understand the alkaline gut system.

A more detailed characterization of a previously described CA of larval *Aedes aegypti* (Corena et al., 2002) is presented in this study, as well as preliminary analyses of a homologous CA isoform cloned from *Anopheles gambiae*. These CA enzymes share some characteristics of the mammalian CA IV isozyme, including a glycosyl-phosphatidylinositol (GPI) link to the plasma membrane. Mammalian CA IV enzymes have been found in dynamic organs such as kidney, lung, gut, brain, eye and capillary endothelium (Chegwidden and Carter, 2000). The human CA IV isoform was found to be as active as the CA II (the so-called 'high activity' CA) isoform in carbon dioxide hydration and even more active in bicarbonate dehydration (Baird et al., 1997). The anterior mosquito midgut lacks a highly active cytosolic CA II-like isozyme (Corena et al., 2002). Therefore, the presence of a highly active CA IV-like isozyme within the mosquito gut may be able to provide the buffering capacity that is needed within the highly alkaline anterior midgut. Our original hypothesis was that the expression of CA controls the regionalized pH extremes found in the larval mosquito midgut. Our results indicate that the distribution of CA alone cannot fully explain the pH gradients found in the midgut.

Materials and methods

Experimental insects

Aedes aegypti L. eggs were obtained from a colony maintained by the United States Department of Agriculture (USDA) laboratory in Gainesville, Florida. The eggs were allowed to hatch in 20 ml of 2% artificial seawater (ASW; 8.4 mmol l^{-1} NaCl, 1.7 mmol l^{-1} KCl, 0.1 mmol l^{-1} $CaCl_2$, 0.46 mmol l^{-1} $MgCl_2$, 0.51 mmol l^{-1} $MgSO_4$, and 0.04 mmol l^{-1} $NaHCO_3$). The mosquito larvae were reared in 2% ASW at room temperature. The *Aedes* larvae were fed a mixture of yeast and liver powder (1:1.5 g respective dry mass; ICN Biomedicals Inc., Aurora, OH, USA). Eight to ten days were required for this species to reach the early fourth instar.

Anopheles gambiae Giles eggs were obtained from the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. Strict handling guidelines were followed with this particular species, which does not currently inhabit Florida, due to its inherent ability to acquire and transmit the causative agent of malaria, the parasitic protozoan *Plasmodium*. This *Anopheles* species was therefore reared in deionized water inside a locked incubator set at 30°C. A mesh screen served as a second barrier within the incubator while the sealed (but not airtight) containers harboring the *Anopheles* larvae served as the third barrier against escape. The *Anopheles* larvae were fed a Wardley tropical fish flake food (The Hartz Mountain Corp., Secaucus, NJ, USA). Early fourth instar larvae were chosen for all experiments. Ten to twelve days from the hatch day were required for this species to reach the early fourth instar. Late fourth instar larvae that went unused were sacrificed to prevent any chance of adult emergence.

Preparation and fixation of tissue

To dissect out the midgut, the heads of the cold-immobilized larvae were pinned down using fine stainless-steel pins to a Sylgard layer at the bottom of a Petri dish containing hemolymph substitute solution consisting of 42.5 mmol l^{-1} NaCl, 3.0 mmol l^{-1} KCl, 0.6 mmol l^{-1} $MgSO_4$, 5.0 mmol l^{-1} $CaCl_2$, 5.0 mmol l^{-1} $NaHCO_3$, 5.0 mmol l^{-1} L-succinic acid, 5.0 mmol l^{-1} L-malic acid, 5.0 mmol l^{-1} L-proline, 9.1 mmol l^{-1} L-glutamine, 8.7 mmol l^{-1} L-histidine, 3.3 mmol l^{-1} L-arginine, 10.0 mmol l^{-1} dextrose, 25 mmol l^{-1} Hepes, pH 7.0 adjusted with NaOH (Clark et al., 1999). The anal segment and the saddle papillae were removed using ultra-fine scissors and forceps, and an incision was made longitudinally along the thorax. The cuticle was gently pulled apart and the midgut and gastric caeca were removed. In some cases, the gut contents enclosed in the peritrophic membrane slid out, leaving behind the empty midgut. In other cases, it was necessary to remove the peritrophic membrane and its contents manually. For enzyme histochemistry, fixation was in 3% glutaraldehyde in 0.1 mol l^{-1} phosphate buffer, pH 7.3, overnight at 4°C (Ridgway and Moffet, 1986). For *in situ* hybridization and immunohistochemistry, dissected tissues were fixed overnight in 4% paraformaldehyde in 0.1 mol l^{-1} phosphate buffer, pH 7.2, or 4% paraformaldehyde in 0.1 mol l^{-1} cacodylate buffer pH 7.2, respectively. Some digital images were acquired using a Leica DMR microscope equipped with a Hammamatsu CCD camera (Shizouka Pref., Japan). Other images were gathered using a Leica LSCM SP2 laser scanning confocal microscope (Exton, PA, USA). All images were assembled using Corel Draw-11 software.

Bioinformatics

The National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov) was used for the majority of the bioinformatical data presented in this study. The first mosquito genome, *Anopheles gambiae,* was released in 2001 (Holt et al., 2002), and made accessible to the public on the NCBI website. The basic local alignment search tool (BLAST; Altschul et al., 1990) was employed for primer construction as well as analyzing PCR products. The NCBI Blast Flies database (www.ncbi.nlm.nih.gov/BLAST/Genome/FlyBlast.html), together with the Ensembl database (www.ensembl.org/Anopheles gambiae/) were used to predict the number of CA genes in the *Drosophila melanogaster* and *Anopheles gambiae* genomes by inputting the *Aedes aegypti* CA as the search sequence.

Ensembl is a joint project between the European Bioinformatics Institute and the Sanger Institute to bring together genome sequences with annotated structural and functional information. The NCBI protein database (pdb) and the BLAST were used in conjunction with the 3-dimensional structure viewer (Cn3D; Hogue, 1997) for the prediction of antibody accessible peptide regions in mosquito proteins. BLAST analyses also confirmed that the chosen antigenic peptides were unique. The conserved domain database (CDD; Marchler-Bauer et al., 2002) and the conserved domain

architecture retrieval tool (CDART; Geer et al., 2002) were used to predict the function of our newly cloned mosquito proteins. Alignments were produced using Clustal W (Thompson et al., 1994), as implemented in DNAman software (Lynnon Biosoft, Vaudreuil, Quebec, Canada).

Cloning of CA from Aedes *and* Anopheles *larval midgut*

The strategies for cloning and sequencing the CA from *Aedes aegypti* have been previously described (Corena et al., 2002). To clone the homologous CA from *Anopheles gambiae*, the *Aedes* CA sequence was BLASTed against the *Anopheles gambiae* genome (Holt et al., 2002). The most similar gene sequence was then used to derive exact primers (5′-AACACTATCTTTTCAGAACCAG [forward primer]; 5′-TAGTAGTACTATCGCTCCCA [reverse primer]) for PCR-based cloning using an optimized protocol for invertebrate tissues (Matez et al., 1999). Amplified cDNA pools from fourth instar *Anopheles gambiae* prepared as described (*ibid*) were used as the basis for the PCR cloning.

In situ *hybridization*

Sense and antisense digoxygenin (DIG)-labeled cRNA probes were generated by *in vitro* transcription using a DIG RNA labeling kit (Roche Molecular Biochemicals, Indianapolis, IN, USA). The full-length *Aedes* CA was subcloned using a PCR manufactured 5′ *Sal*I restriction site and a 3′ *Xho*I site. Full-length sense and antisense DIG cRNA probes were produced according to the manufacturer's instructions.

These *in situ* experiments contained an additional pre-fixation step. A glass electrode fitted to a micromanipulator was used to inject 4% paraformaldehyde into the thoracic cavity, just behind the head. Successful perfusion was easily identified by the cessation of the otherwise constant muscle twitching along the length of the body. Then the exoskeleton was removed by careful dissection. Subsequent steps for *in situ* hybridization methods were adapted from Westerfield (1994). The midguts were washed with PBS at room temperature and then incubated in 100% methanol at –20°C for 30 min to ensure permeabilization of the gut tissue. The tissue was washed (5 min each wash) in 50% methanol in PBST [Dulbecco's phosphate buffered saline (Sigma-Aldrich, St Louis, MO, USA) plus 0.1% Tween-20], followed by 30% methanol in PBST and then PBST alone. The tissue was fixed in 4% paraformaldehyde in 0.1 mol l^{-1} phosphate buffer for 20 min at room temperature and washed with PBST. The larval midguts were digested with proteinase K (10 μg ml^{-1} in PBST) at room temperature for 10 min, washed briefly with PBST and fixed again, as described previously.

Prehybridization of the tissue was accomplished by incubation in HYB solution [50% formamide, 5× SSC (1× SSC=0.15 mol l^{-1} NaCl, 0.015 mol l^{-1} sodium citrate buffer, pH 7.0), 0.1% Tween-20] for 24 h at 55°C. The larval midguts were transferred to HYB+ solution (HYB plus 5 mg ml^{-1} tRNA, 50 μg ml^{-1} heparin) containing 5 ng ml^{-1} DIG-labeled probe and incubated overnight at 55°C. Excess probe was removed by washing at 55°C with 50% formamide in 2× SSCT for 30 min (twice), 2× SSCT for 15 min and 0.2× SSCT for 30 min (twice). For detection, the tissue was incubated in PBST containing 1% blocking solution (Roche Molecular Biochemicals) for 1 h at room temperature. The tissue was incubated with anti-DIG-alkaline phosphatase (Roche Molecular Biochemicals) diluted 1:5000 in blocking solution for 4 h at room temperature. The tissue was washed with PBST and incubated in alkaline phosphatase substrate solution (Bio Rad Laboratories, Hercules, CA, USA) until the desired intensity of staining was achieved (2–3 h) with sense and antisense samples receiving identical incubations.

Real time polymerase chain reaction

Region-specific cDNA was produced from dissected mosquito tissue using the Cells-to-cDNA standard protocol (Ambion INC, Austin, TX, USA). The gut regions used to make the amplified cDNA pools were incubated in 50 μl of hot cell lysis buffer for 10 min at 75°C. The lysed tissues were treated with 2 U of DNase I for 30 min at 37°C. The DNase I was then inactivated by heating to 75°C for 5 min. For the reverse transcription reaction, 10 μl of cell lysate was combined with 4 μl dNTP mix (contains 2.5 mmol l^{-1} each dNTP) and 5 μmol l^{-1} random decamer first strand primer in 16 μl total volume. The mixture was incubated at 70°C for 3 min and then chilled on ice for 1 min. This mixture was then combined with 1× reverse transcription (RT) buffer as supplied with the enzyme, 1 U M-MLV reverse transcriptase, and 10 U RNAse inhibitor, and incubated at 42°C for 1 h. The reverse transcriptase was then inactivated by incubation at 95°C for 10 min. Primers were designed using Primer Express software (Applied Biosystems; Foster City, CA, USA). The SYBR Green PCR Master mix, which includes SYBR Green I dye, Amplitaq Gold DNA Polymerase, dNTPs and buffer, was used for all real-time polymerase chain reaction (PCR) investigations. Each cycle of PCR was detected by measuring the increase in fluorescence caused by the binding of the SYBR Green dye to double-stranded DNA using an ABI Prism 7000 Sequence Detection System (Applied Biosystems). Initially, each primer set, including the control 18s ribosomal RNA (GenBank accession no. M95126), was assessed to determine the optimal concentration of primer to be used. All real-time experiments used the same two-step cycling profile: 50°C for 2 min followed by 95°C for 10 min and 40 cycles of 95°C for 15 s and 60°C for 1 min. Whole gut cDNA (100 μg l^{-1}) was used as template with 500 nmol l^{-1}, 300 nmol l^{-1}, 100 nmol l^{-1}, or 50 nmol l^{-1} of each primer set and 1× SYBR green I master mix in 25 liters total volume. Each reaction was done in triplicate. The optimal concentration was then chosen based on the amplification plots and the dissociation curves generated. Once a concentration was chosen for each primer set, the efficiency of amplification of that set was determined. Serial dilutions of whole gut cDNA were used as template with the appropriate concentration of primers and 1× SYBR green I master mix in 25 μl total volume. The threshold cycle number (Ct) was plotted *versus* the log of the template concentration

and the slope (m) and intercept (b) were determined. These pre-determinations were then used in the standardized comparison of the amount of 18s transcript and CA transcript in each of the cDNA samples tested. For each analysis a control containing all of the necessary PCR components except the cDNA template was run. To determine the relative expression level for each transcript analyzed, the following equation was used: (Ct–b)/m. The average log(ng) for each transcript was then compared to the average log(ng) of 18s RNA transcript to normalize the values. Then the expression levels were determined relative to the transcript with the greatest normalized log(ng) value and expressed in a bar graph using Microsoft Excel software.

Antibody production

An 18 amino acid peptide from the cloned *Aedes* CA sequence was chosen for antibody production. In order to increase the probability that this antibody would be specific for this particular CA sequence (in the event that other CA isoforms were expressed in the larval mosquito gut), attempts were made to synthesize an antigenic peptide that would be specific to this isoform. The well-characterized mammalian CA isoforms served as a model when trying to choose a unique CA peptide sequence. The comparison of the mosquito CA with the mammalian isoforms yielded a peptide sequence from the amino (N) terminus, where CA isoforms showed the most diversity, and least conservation. The N terminus of our mosquito CA was predicted to have an extended loop secondary structure. Unlike an alpha helix, an extended loop is more accessible to antibody probing. Furthermore, three-dimensional analyses (Cn 3D v4.1 NCBI) of predicted CA IV structures (human 1ZNC and mouse 2ZNC) predicted that the N terminus is exposed and accessible. An 18 amino acid peptide was therefore chosen from the N terminus of the *Aedes* CA sequence. This peptide sequence (GVINEPERWGGQCETGRR) was sent to Sigma-Genosys (Woodlands, TX, USA), where it was synthesized and conjugated to bovine serum albumin (BSA). The synthetic peptide–BSA construct and Freund's incomplete adjuvant were injected into two rabbits to elicit an immune response. Prior to injection, a blood sample from each rabbit was collected to serve as the control pre-immune serum. Every 2 weeks a blood sample was collected from the rabbits, the fraction of immunoglobulin G (IgG) pooled, and another dose of the peptide-BSA construct administered. Three months after the initial injections, the final bleeds were collected and used for all immunohistochemical analyses.

Immunohistochemistry

The resultant antisera were used to determine the specificity of the antibodies as well as to determine the localization of the larval mosquito proteins. Dissected and fixed whole-mount mosquito guts were washed 6× in Tris-buffered saline (TBS), placed in pre-incubation medium (pre-inc; TBS with 0.1% TritonX-100 and 2% bovine serum albumin) for a minimum of 1 h, and then incubated in primary antibody (1:1000) overnight at 4°C. The guts were then washed in pre-inc and incubated in FITC-conjugated goat anti-rabbit (GAR) or Alexa-GAR secondary antibody (Jackson ImmunoResearch, West Grove, PA, USA; 1:250 dilution) overnight at 4°C. The whole-mount preparations were rinsed in pre-inc and mounted onto slides using *p*-phenylenediamine (PPD, Sigma-Aldrich) in 60% glycerol. Draq 5 (Biostatus Limited, Shepshed, UK, 1:1000 dilution) was applied before mounting to visualize nuclear DNA. The samples were examined and images captured using the Leica LSCM SP2 laser scanning confocal microscope.

Live preparations were examined, following a similar procedure, to ensure that antibodies were capable of localizing extracellular proteins only. In this case, living larvae were pinned to Silgard dishes and the exoskeleton opened and pinned back. The living gut preparation, which we have shown remains functional for many hours (Boudko et al., 2001), was exposed to antiserum or preimmune serum diluted 1:1000 in hemolymph substitute solution (HSS; Clark et al., 1999). The samples were washed extensively in HSS and then fixed in paraformaldehyde as described above, followed by labeling with fluorescent secondary antibodies. The live gut assays were also performed to determine whether this specific CA is tethered to the cell membrane *via* a GPI linkage. Ten live gut preparations were incubated with phosphoinositol-specific phospholipase C (PI-PLC, 5 units per ml in HSS; Sigma-Aldrich) for 3 h at 37°C. PI-PLC was used as a tool in determining the presence of a GPI link. Controls in which the guts were incubated in HSS alone were also performed. The guts were then washed in HSS, fixed, and treated with primary and secondary antibodies as described above.

CA protein expression

Recombinant *Aedes* CA was produced using the pET100 vector (Invitrogen, Carlsbad, CA, USA). Specific primers were designed to amplify the cDNA. The 3′ primers included the sequence 5′ to the hydrophobic tail region. The 5′ primers contain the sequence CACC preceeding the native start codon for correct frame insertion. PCRs were performed using 1 U of Platinum P*fx* polymerase (Invitrogen), the gastric caeca cDNA collections as template (200 ng), 1× Pfx amplification buffer as supplied with the enzyme, 1.2 mmol l^{-1} dNTP mixture, 1 mmol l^{-1} $MgSO_4$, and 0.3 μmol l^{-1} of each primer in a total volume of 50 μl. A three-step PCR protocol was used consisting of 94°C for 2 min followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 1 min.

The resultant blunt-ended cDNA (4 μl from PCR mix) was ligated with the pET100 directional Topo vector (1 μl and 1 μl salt solution; Invitrogen) for 10 min at room temperature. Top 10 chemically competent *E. coli* (50 μl; Invitrogen) were transformed by incubating 3 μl of ligation mix with the cells for 30 min on ice, followed by a heat shock of 42°C for 30 s. SOC medium (250 μl; Invitrogen) was added to the cells and they were then incubated at 37°C for 30 min with shaking. The transformation mix (100 μl) was then plated on a Luria-Bertoni-carbenicillin (LB-carb) plate (50 μg ml^{-1}) and

incubated overnight at 37°C. Colonies were sequenced using Big Dye version 1.1 as described previously.

The purified plasmids (10 ng each) were transformed into BL21 Star (DE3) cells (Invitrogen) for CA expression. However, after SOC addition and incubation, the culture was transferred to fresh LB-carb (10 ml) and grown overnight at 37°C with shaking. The next day, 1 ml of culture was transferred to 100 ml of fresh LB-carb and was grown at 37°C with shaking. Optimization experiments were performed in order to facilitate the production of the greatest quantity of CA protein. For production of CA protein, isopropylthio-β-galactoside (IPTG, 1 mmol l^{-1} final concentration; Stratagene, La Jolla, CA, USA) was added when the culture had attained an optical density of 0.5 at a 600 nm wavelength. Achieving this density took about 1.5 h of growth at 37°C and 200 revs min^{-1}. Zinc, in the form of zinc sulfate (0.5 mmol l^{-1} final concentration), was added along with the IPTG to facilitate the proper conformation of an active zinc-binding CA protein. In order to optimize the duration of the induced growth phase, samples were collected every hour for 6 h. These samples were analyzed on an SDS-PAGE 4–12% Bis-Tris gel to compare CA protein content. 4 h of growth was determined to be ideal for the production of the truncated *Aedes* CA.

Total protein was collected using the Probond Purification System according to the manufacturer's instructions for soluble proteins (Invitrogen). The cells were harvested by centrifugation, sonicated in native buffer (250 mmol l^{-1} $NaPO_4$, 2.5 mol l^{-1} NaCl; Invitrogen) with lysozyme (1 mg ml^{-1}; Sigma-Aldrich), and centrifuged again to collect a crude protein extract. The supernatant was applied to a Probond nickel column (Invitrogen) and washed free of non-specific binding contaminants. The nickel column binds the CA protein due to the added histidine tag, a repeat of six histidine residues within the pET100 expression vector that is inserted after the carboxy (C) terminus of the CA protein. CA was eluted by adding imidazole (250 mmol l^{-1}; Invitrogen) to the column, which competes and displaces the histidine tag. Eluted fractions were separated on an SDS-PAGE 4–12% Bis-Tris gel (Invitrogen). Separated proteins were electroblotted to nitrocellulose membranes and then analyzed for total protein and then by immunostaining by standard methods.

Results

Bioinformatics of Aedes aegypti *CA*

We have previously cloned a CA cDNA from the *Aedes aegypti* midgut (accession number AF395662; Corena et al., 2002). Our initial structure prediction indicated that the protein is cytosolic. However, further characterization has indicated that this CA is actually membrane associated *via* a GPI-link. We have determined that the CA propeptide sequence encodes an extracellular protein with a hydrophobic tail region. The first 17 amino acids of the propeptide are predicted by the Simple Modular Architecture Research Tool (SMART) program to be the signal sequence (Letunic et al., 2002). This sequence 'flags' the message for transport to the endoplasmic reticulum (ER). Using the PSORT II server, the prediction of membrane topology (MTOP) finds the *Aedes* CA sequence to be GPI-anchored. Amino acid G-276 is predicted by the GPI prediction server to be the site for GPI attachment (Eisenhaber et al., 1999). The hydrophobic tail (L278–A289) allows translocation of the transcript through the ER plasma membrane and is also predicted to stabilize the protein with the membrane until the pre-formed GPI anchor is transferred to the protein. The hydrophobic tail is then cleaved to produce a completely extracellular protein that is tethered to the cell by the GPI link (for a review of GPI-linked proteins, see Brown and Waneck, 1992).

Sequence comparisons of CA IV-like isoforms

Using exact primers deduced from the *Anopheles* gambiae genome, we also cloned a CA IV-like cDNA from the malaria mosquito. This CA isoform (Ensembl gene ID: ENSANGG00000018824, chromosome 2L) is partially predicted by the Ensembl CA protein family (ENSF00000000228) as one of 14 gene family members found in the *Anopheles gambiae* genome. These cloned mosquito cDNAs from *Aedes aegypti* and *Anopheles gambiae* are 61% identical in amino acid residues and show the greatest likeness to the mammalian CA IV isozyme. In contrast to the mammalian CA IV, which is encoded by 7 exons (Sly and Hu, 1995), only three exons make up the *Anopheles* CA isoform. Alignment of the mosquito CA IV-like isoforms from *Aedes* and *Anopheles* with various mammalian CA IV isozymes shows amino acid similarities between these CA isoforms (Fig. 1). The multiple leucine residues within the N terminus of the mammalian CA IV propeptides that comprise the signal sequence are also found in the *Aedes* and *Anopheles* CA IV-like isoforms. One important feature of the mosquito CA IV-like sequences is the conserved alignment of G-69 (human CA IV numbering) with the human, bovine and rabbit CA IV sequences. This particular amino acid residue has been changed to glutamine (Q) in rat and mouse CA IV, which results in reduced enzyme activity (Tamai et al., 1996a,b). Additionally, all of the CA IV sequences, including the mosquito isoforms, display a hydrophobic tail region. In addition to the conserved CA IV-like features of GPI-linked proteins, there are also conserved cysteine residues (C28 and C211, human CA IV numbering) between all of these CAs (Fig. 1). It has been determined *via* cysteine labeling, proteolytic cleavage and sequencing that these two cysteine residues, in the human CA IV, form a disulfide bond (Waheed et al., 1996). A second disulfide bond is present in the mammalian CA IVs between residues C6 and C18 (human CA IV numbering; Waheed et al., 1996). This second pair of cysteine residues, and hence the resultant disulfide bond, is not present in either of the mosquito isoforms.

In situ *hybridization for CA localization*

In situ hybridization analyses indicate that the *Aedes aegypti* CA message is expressed most heavily within the epithelial cells of the gastric caeca and posterior midgut

(Fig. 2). An antisense cRNA probe corresponding to the entire cDNA sequence generated strong cytoplasmic staining of the proximal gastric caeca, while the distal cap cells (*) were void of label consistent with previously published histochemical staining (Corena et al., 2002) (Fig. 2B). Anterior to the gastric caeca, a strong localization was evident in a small subset of cardia cells that encircle the tissue, forming a collar (Fig. 2B). These 'collar cells' are clearly different from the surrounding cells in this same area. This technique also highlighted a set of specific epithelial cells that are found only in a subset of the posterior midgut. These CA-positive cells form a ring of about five cells in width that circumscribe the lower-posterior gut region (Fig. 2A,C). CA message was also localized to longitudinal and circular muscle fibers of the anterior and posterior midgut (Fig. 3). Following the longitudinal muscle fibers, in close association, are distinct nerve fibers that also display strong CA labeling (Fig. 3). Epithelial cells of the anterior midgut were clearly void of signal beneath the labeled muscle and nerve cells. Specific staining was also evident however within the abdominal ganglia of the nervous system (CNS) and peripheral nerve tissue (Fig. 4). No labeling was seen in the Malpighian tubules. Sense probes showed no labeling (not shown).

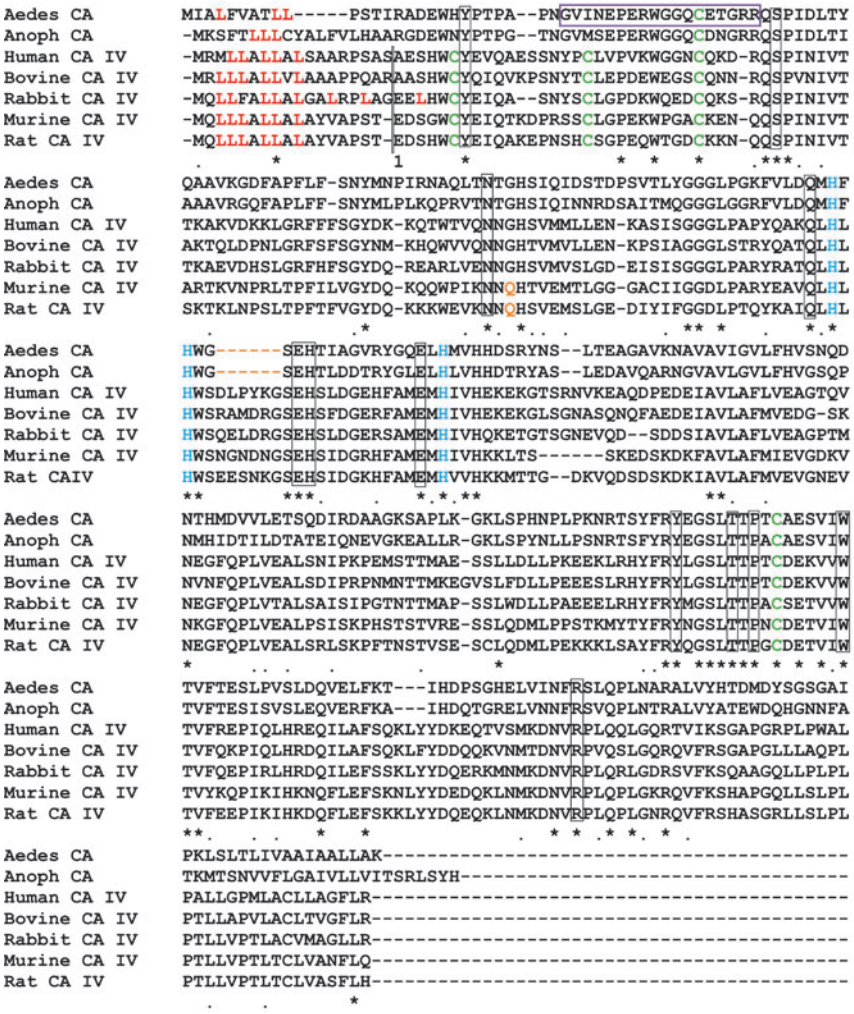

Fig. 1. Alignment of several mammalian CA IV enzymes with two mosquito CA IV-like isoforms. The leucine-rich signal sequences are displayed in all aligned isoforms (red), along with the 3 essential zinc-binding histidines (blue), and cysteine residues (green) that form disulfide bonds. The reduced activity in rodent CA IVs is caused by the glycine-69 mutation to glutamine (orange; Tamai et al., 1996a,b), which the mosquitoes do not display. Important conserved residues are boxed. The position of mammalian signal sequence cleavage is shown (vertical line) and therefore the following amino acid is residue #1 in the functional protein. The peptide sequence used for antibody generation is also displayed (violet horizontal box). Asterisks, identical residues; dots, conserved residues. Broken orange lines mark the shortened active site region within the two mosquito sequences when compared to mammalian CA IV enzymes.

Real-time PCR analysis of Aedes aegypti *CA IV-like transcripts*

Real-time PCR was used to compare the levels of *Aedes aegypti* CA mRNA within specific tissue regions of the larvae. 20 fourth instar *Aedes aegypti* larvae were dissected and the head, gastric caeca (GC), anterior midgut (AMG), posterior midgut (PMG), and Malpighian tubules (MT) were pooled. RNA was isolated from each tissue sample for subsequent real time PCR analysis. *Aedes aegypti* ribosomal RNA (GenBank accession number M95126) was used to normalize the quantity of transcript from each sample. The results are presented in graph format in Fig. 5. This technique found the gastric caeca to contain the greatest quantity of CA message within the gut sections (Fig. 5). The head section contained roughly half as much message as the gastric caeca (Fig. 5). The localization of CA IV-like message within the larval head supports the *in situ* hybridization finding of CA message within CNS tissue. The anterior midgut, posterior midgut and Malpighian tubule collections showed much lower levels of CA message (Fig. 5).

Immunolocalization of CA IV-like protein in the mosquito gut

The N-terminal peptide sequence (GVINEPERWGGQCETGRR, see Fig. 1) was chosen from the *Aedes aegypti* CA sequence as an antigen for antibody production. The resultant antiserum was used to analyze recombinant protein expressed in bacteria and to immunolocalize the CA IV-like isoform within the mosquito gut. The pre-immune serum was used as a control for all experiments. Fig. 6 shows a western immunoblot analysis of the cloned CA expressed in *E. coli*. XPress epitope antibody (Invitrogen Inc.) identifies the expressed

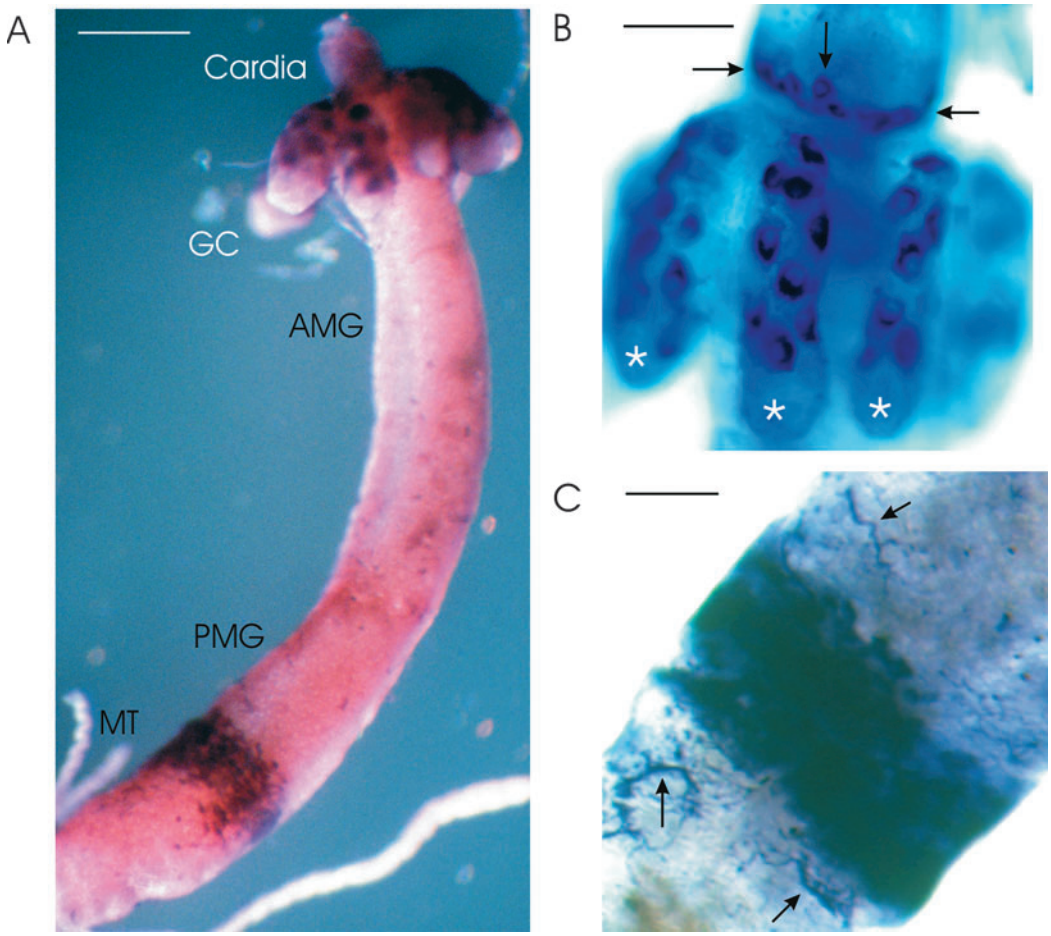

Fig. 2. Localization of CA mRNA in a whole-mount preparation of early fourth instar *Ae. aegypti* midgut. (A) The whole-mount gut preparation localizes CA message to specific cells of the gastric caeca (GC) and posterior midgut (PMG).The anterior midgut (AMG) epithelial cells and the malpighian tubules (MT) showed little or no labeling. (B) A subset of cardia (arrows) and GC cells display the CA message. The distal lobes of the GC, called Cap cells, display no staining (asterisks). (C) There is a distinctive labeling pattern of CA message within a specific band of PMG epithelial cells. In addition, numerous trachea are heavily labeled along the length of the midgut (arrows). Scale bars, 300 μm (A), 150 μm (B), 75 μm (C).

recombinant protein band (Fig. 6B) and the same band then labels intensely with the rabbit anti-CA peptide serum (Fig. 6B).

Whole-mount preparations of larval mosquito guts were immunostained with the anti-CA serum diluted 1:1000. These preparations were counterlabeled with TRITC-conjugated Phalloidin and DRAQ-5 to label muscle (actin) and nuclei (DNA), respectively. Fig. 7 shows laser scanning confocal images of whole gut preparations at two magnifications. The most prominent staining was of a specific subset of the muscles, which encircle the gut epithelial tube. It has long been known that larval mosquito gut is contained within a tightly associated tubular meshwork of muscles (e.g. see chapter 5 in Clemens, 1992). Immunostaining with the antibodies to the CA IV-like CA from *Aedes aegypti* defines a subdivision in the muscle basket: CA positive and CA negative muscle fiber bundles (Fig. 7). The labeling of the muscle fibers is on the plasma membrane surface of the muscles, and in direct contact with the hemolymph and the basal side of the anterior midgut epithelium. The conservation of the peptide epitope (used in generating this antibody) between different mosquito species (see Fig. 1; 14 of 18 amino acids conserved between *Aedes aegypti* and *Anopheles gambiae*), led us to test the immunostaining capacity in a number of larval mosquito species. In each of five species that we tested (*Aedes aegypti, Aedes albopictus, Anopheles gambiae, Anopheles quadramaculatus, Ochlerotatus taeniorhynchus*), similar discrimination of a subset of gut muscles was seen (not all shown). The overall pattern of CA-positive muscles in *Aedes aegypti* was such that in the anterior half of the gut, the lateral quadrants of the gut tube were bounded by a meshwork of CA-positive muscles (both circular and longitudinal) that covered approximately one quarter of the circumference of the gut each. This regularly arranged mesh of muscle ran approximately two thirds of the length of the midgut from caecum to the pylorus. In the midst of the posterior midgut region, this grouping of muscles dissipates. On the dorsal and ventral sides of the gut tube, only a few longitudinal muscles labeled for the CA and this labeling proceeds throughout the full length of the midgut. In all cases, the CA-positive muscles were accompanied by CA-negative muscles. Once the presence of the lateral arrangement of CA-positive muscles in the anterior gut is recognized, it then becomes a simple matter to see these distinct but overlapping muscles relative to the remainder of the muscle basket even in the absence of CA staining. That is to say, that the CA-positive muscles create a higher density of basket muscles on the lateral aspects of the anterior gut tube that can be recognized with actin staining alone in *Aedes aegypti* (Fig. 7B). In other

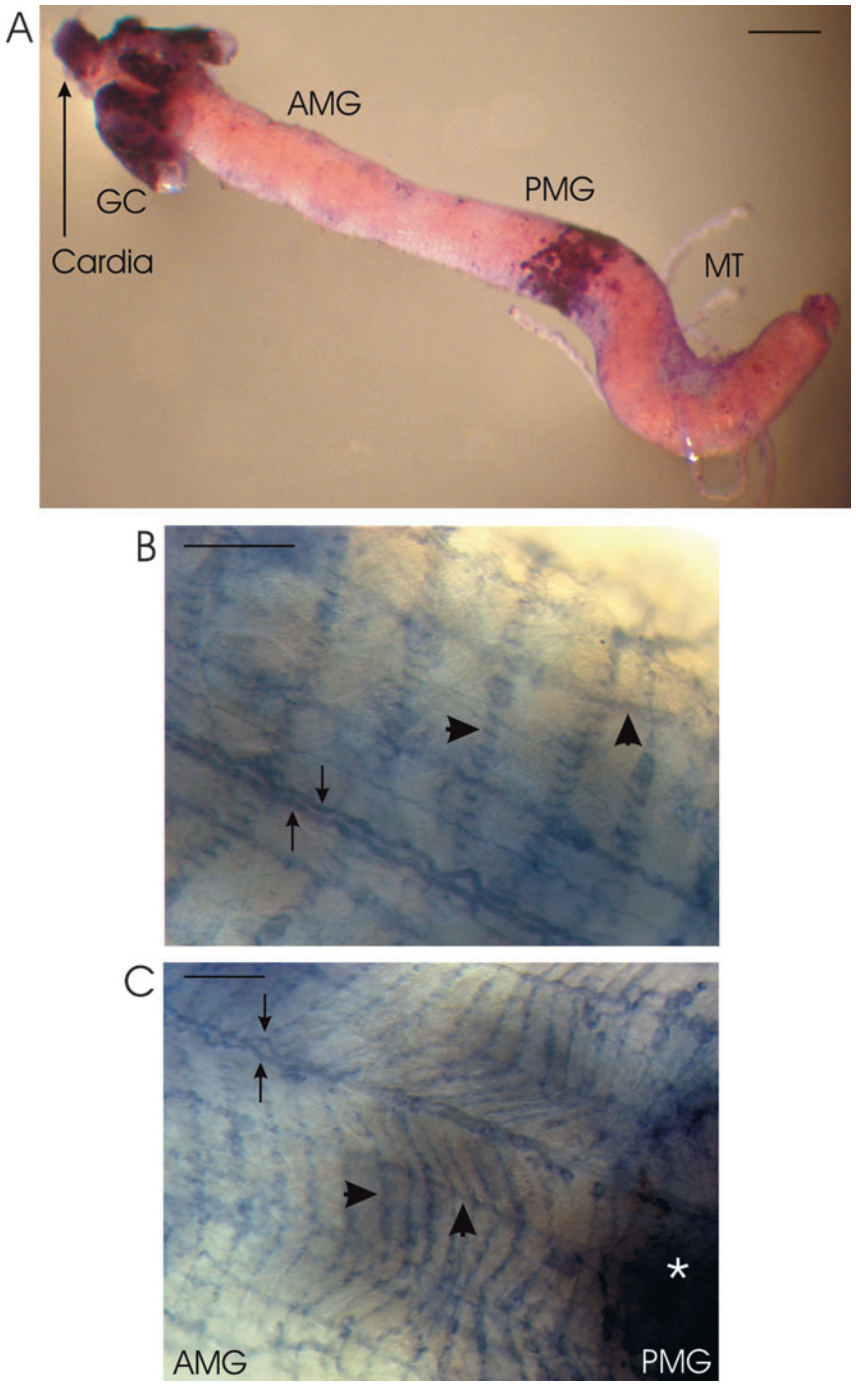

Fig. 3. *Ae. aegypti* anterior midgut *in situ* hybridization CA labeling. While the GC and PMG display heavy epithelial labeling for the CA message (A), there is also specific labeling seen in muscle (B,C, large arrowheads) and nerve cells (B, small arrows). (A) A representative whole-mount larva displaying the strong epithelial label in GC and PMG along with muscle fiber staining in the AMG that can be overlooked at low magnification. (B) The beginning of the anterior midgut displays both muscle and nerve fiber labeling. The labeled fibers reveal striated muscle running longitudinally down the length of the anterior gut and circularly around the girth of the gut (B,C, large arrowheads). The nerve fibers can be distinguished by their non-striated wavy appearance (B,C, small arrows). (C) The AMG (left) displayed strong labeling in muscle and nerve fibers while displaying no epithelial cell labeling. The PMG (right, asterisk) shows fiber labeling as well as intense epithelial cell labeling. Scale bars, 300 μm (A), 25 μm (B), 50 μm (C). Abbreviations as in Fig. 2.

larval mosquitoes, a similar distinction between CA-positive and CA-negative muscles exits. In *Anopheles gambiae*, the lateral CA-positive muscles have a rounded posterior extreme and hence appear somewhat wing-shaped (Fig. 7D–F). In the posterior midgut of *Anopheles gambiae*, very little labeling of any muscle fibers occurs with the CA antibody. All species examined showed lateral groupings of CA-positive muscles intermixed with CA-negative muscles in the anterior midgut. The posterior midgut musculature varied in CA-labeling from very few positive muscles to mostly positive muscles between species. It is impossible to state the functional distinction between the two classes of gut muscles at this point. However, both peristaltic and antistaltic contractions are known to occur in larval mosquito gut and perhaps the two muscle types contribute differentially to these gut movements. Immunolabeling of the gastric caeca and posterior midgut was also seen, although frequently obscured by the labeling of the basket muscles surrounding the gut epithelial tube. Immunoreactivity was also found within the neural ganglia and immunoreactive nerve fibers that traverse the ventral gut in punctate clusters (Fig. 8). There was no immunoreactivity in the Malpighian tubules.

Phospholipase-C treatment

In order to validate that the CA IV-like isoform cloned from *Aedes aegypti* is indeed GPI linked to the membrane, live fourth instar *Aedes aegypti* and *Anopheles gambiae* larvae were subjected to phosphoinositol-specific phospholipase C (PI-PLC) treatment and subsequent immunohistochemistry. This enzyme specifically cleaves the GPI-anchor and therefore severs GPI-linked proteins from the plasma membrane. Larvae subjected to PI-PLC treatment showed a dramatic decrease in CA antibody immunoreactivity along the midgut muscle and nerve fibers, as compared to the non PI-PLC treated controls (not shown). This evidence supports the bioinformatical finding that the mosquito CA IV-like isoform is in fact GPI-linked to the outer plasma membrane.

To further substantiate the cell-surface localization of the muscle CA compartment, living larvae were dissected in HBSS and laid open. The living gut tissue was then exposed

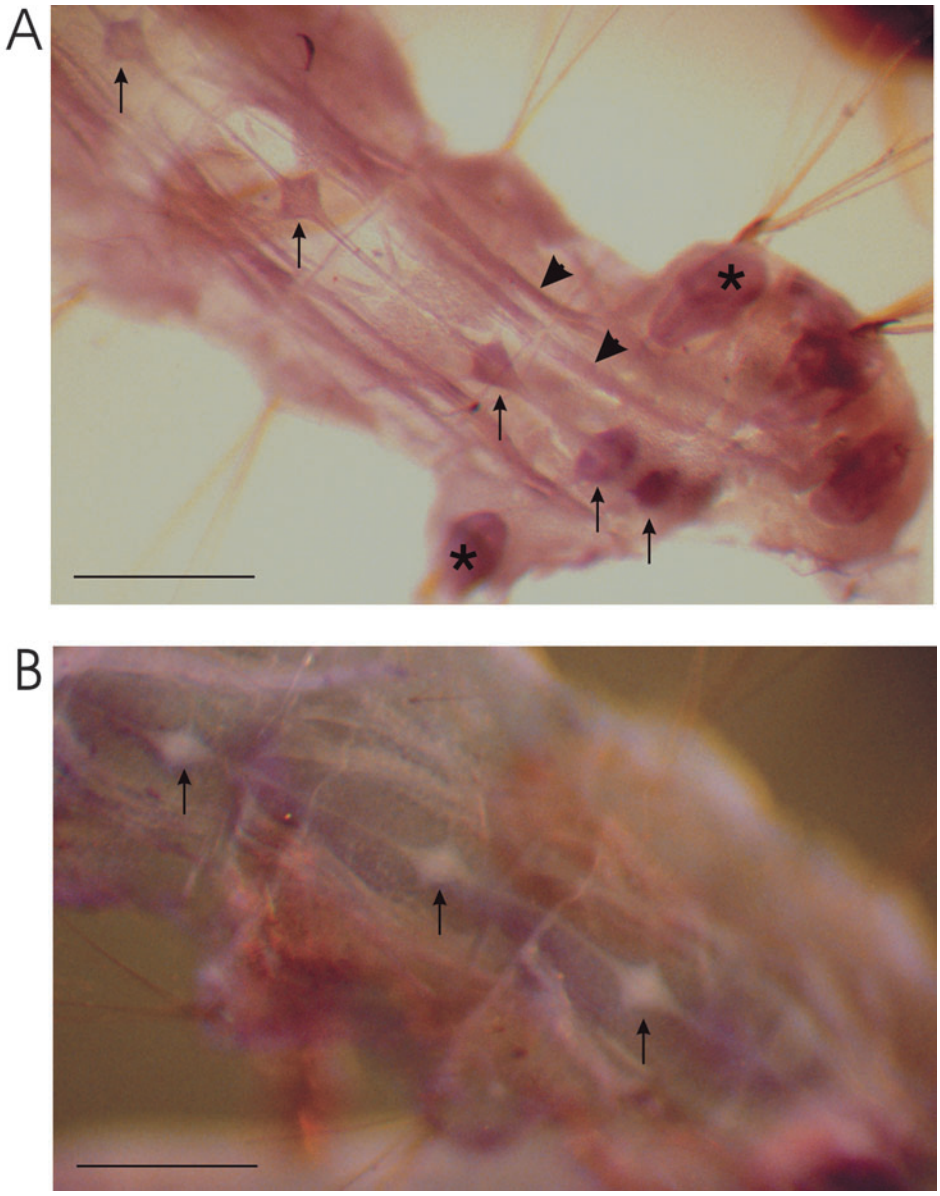

Fig. 4. Localization of CA message within *Ae. aegypti* CNS tissue. *In situ* hybridization localized the CA message within all ventral ganglia CNS clusters (A, arrows) as well as hair sensory cells (A, asterisk) and longitudinal nerve fibers (A, arrowheads). (B) Sense control showed no labeling in ganglia (arrows) or other neural structures. Scale bars, 300 μm.

to antibodies followed by washing, fixation and subsequent localization of antibody binding with secondary antibodies. As with pre-fixed tissue, the antibodies specifically labeled the specific gut muscles described previously (not shown). High magnification confocal microscopy also shows the muscle-surface labeling clearly when viewed as a single *z*-plane in cross section (Fig. 9).

Fig. 10 shows a Clustal alignment of the two CA IV-like mosquito CA sequences plus the putative homologue from *Drosophila* aligned with all known human CA isoforms. Conservation of critical amino acids such as histidines known to be involved in coordination of zinc in the active site are all present in the insect CAs but vary in the human genes, which produce inactive CA-related proteins (Fig. 10).

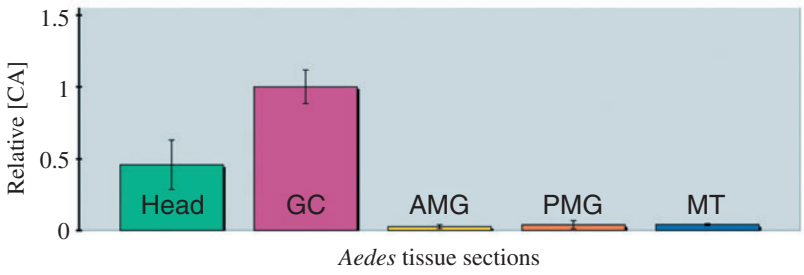

Fig. 5. Real-time PCR analysis of relative concentrations of CA message in *Ae. aegypti* larvae. The GC tissue displays the greatest amount of CA message. This value was arbitrarily set to 1 so that the other tissue sections could be relatively compared. The AMG and PMG along with the MT display very little CA message. The head section displays roughly half the amount of message found in the gastric caeca. All samples were normalized to 18S RNA. Values are means ± 1 S.E.M. Abbreviations as in Fig. 2.

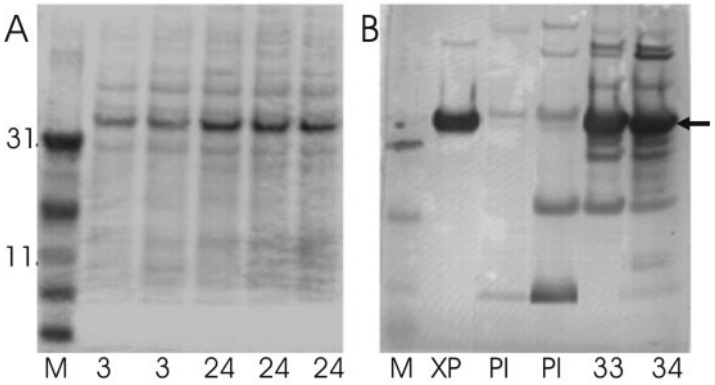

Fig. 6. Immunoblot analysis of bacterial expressed *Aedes* CA. (A) Protein staining (Fast Green) of an electroblot from an SDS-PAGE analysis of extracts from cultures of bacteria that had been transformed with the CA expression vector. Labels at the bottom of the panel show the molecular mass marker lane (M) and extracts from 3 or 24 h of culture time. Labels on the left indicate molecular masses of the markers. A protein of approximately 35 kDa displays the highest level of expression. In (B) the same blot was subsequently immunostained with the XPress epitope antibody (Invitrogen, lane XP) preimmune serum from rabbit 33 (left lane labeled PI) or rabbit 34 (right lane labeled PI) or the antisera from the rabbits (lanes 33 and 34). Note intense labeling of the same band identified by the XPress antibody as detected by the two rabbit antisera to the CA peptide at approximately 35 kDa, the expected mass of the expressed recombinant CA protein (arrow).

Also, it is very interesting to note that the insect CAs have a shortened active site sequence relative to human forms (Fig. 10, broken red line). Thus it is possible that the active site in the Dipteran CAs may be sufficiently different to provide an avenue for the development of very specific CA inhibitors that might be used in mosquito control strategies.

Discussion

In this study, we present sequence analyses of two homologous GPI-linked CA isoforms that are expressed in the midgut of two different mosquito species that rely on an alkaline digestive strategy. These mosquito CA isoforms share characteristics with the mammalian CA IV isozyme, including the GPI link to the membrane. *In situ* hybridization localized CA message predominantly to the gastric caeca and a subset of posterior midgut epithelial cells, along with specific muscle and neural tissue associated with the midgut. RT-PCR analysis confirmed the presence of CA message within the *Aedes aegypti* gut and CNS. The gastric caeca were found to contain the greatest amount of CA message in relation to the other gut samples while the head sample contained roughly half of the gastric caeca concentration. Immunolocalization of the CA IV-like isozyme within the mosquito gut and CNS demonstrated that the CA message is being translated into protein. Immunoreactivity was most striking on specific muscle fibers of the anterior midgut, along with labeling of the gastric caeca and CNS ganglia. The *in situ* hybridization analyses qualitatively coincide with the immunolocalization. However, the intensity of immunolabeling for this CA on muscles when compared to the gastric caeca seems to contradict the apparent

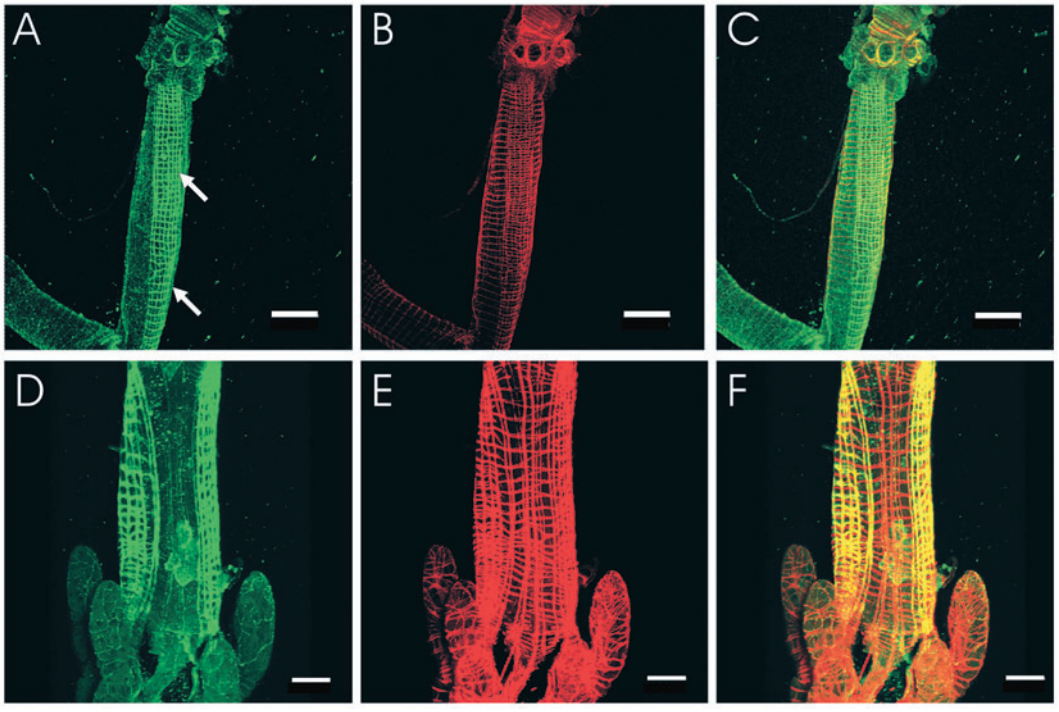

Fig. 7. Immunofluorescence and confocal imaging of CA antibody labeling. (A–C) An isolated midgut from a fourth instar *Ae. aegypti* larva immunostained for CA (A, green) and labeled with TRITC-conjugated Phalloidin (B, red) to visualize actin. (C) Merge of the two colored images. Note the specific checkerboard arrangement of CA labeling in AMG (A, arrows), which corresponds to a subset of the phalloidin labeled muscles seen in B and C. (D–F) A similar subset of the gut musculature at higher magnification, labeled for CA from a fourth instar larva of *An. gambiae*. Scale bars 200 μm (A–C); 80 μm (D–F).

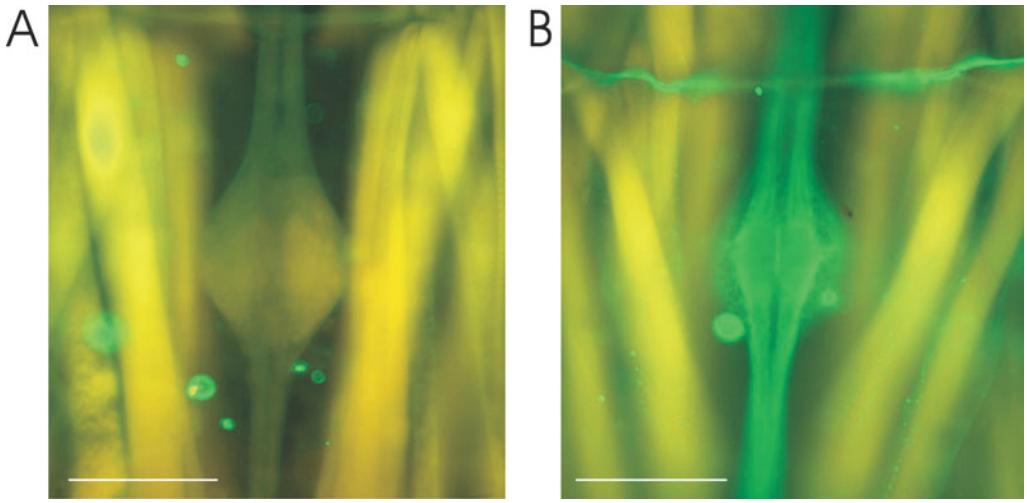

Fig. 8. The *Ae. aegypti* CNS ganglia express this CA IV-like isoform. (A) Pre-immune serum does not show any immunoreactivity for the CNS tissue. (B) Strong immunolabeling for the mosquito CA is displayed in the ventral ganglion clusters, as displayed by the fluorescent green coloring as compared to the yellow control (pre-immune) ganglia. Scale bars, 100 μm.

levels of mRNA expression, which may be explained by differential translation of the message into protein. It could also be due to cross-hybridization between our DIG-labeled cRNA probes and more than one of the 14 different CA gene products that may exist as indicated by genomics. As all of the 14 putative CA genes have regions of high homology, it is possible that our full-length cRNA probe may have hybridized to more than one specific mRNA. A final possible explanation for the seeming difference between mRNA expression and protein levels could simply be antigenic accessibility in the whole-mount method for immunostaining employed herein.

In this study we have shown that the basket of muscles surrounding the larval mosquito midgut is complex and contains at least two distinguishable populations of muscle fibers: CA-positive and CA-negative. Prominent expression of CA on the surface of the muscles may have a role in midgut alkalization. As noted before, the alkaline region of the larval mosquito midgut is restricted to the anterior half of the gut tube. In this region the gut pH can be as high as 11 (Zhuang et al., 1999). It is widely thought that the alkaline buffer is most likely to be carbonate (perhaps potassium carbonate; Boudko et al., 2001), so it stands to reason that a CA activity should be involved in alkalization. Furthermore, we have previously shown that inhibition of CA activity blocks anterior midgut alkalization (Corena et al., 2002). Nevertheless, enzyme histochemistry (*ibid*), real-time PCR, *in situ* hybridization and immunolocalization studies all show there to be little or no CA in the anterior midgut epithelial cells. This very strongly suggests that the bicarbonate source of the anterior midgut carbonate buffer, originates by the action of CA in cells other than the anterior midgut cells themselves. This leaves at least two possibilities: the bicarbonate may be produced and secreted into the gut luminal fluid by the gastric caeca cells and is then stripped of its extra proton once it reaches the anterior midgut, or it may be transported from the hemolymph into the lumen by the anterior midgut cells. The CA IV-like enzyme that we have localized to a specific subset of muscles specifically associated with the anterior region of the midgut could possibly contribute to anterior gut alkalization by maintaining the highest possible concentrations of bicarbonate in the hemolymph in the immediate vicinity of the anterior midgut (thus supporting the local epithelial transport hypothesis). Previous findings from this laboratory have shown a strong efflux of chloride from the AMG epithelium (Boudko et al., 2001). Since chloride transport is frequently part of an exchange with bicarbonate, a net influx of bicarbonate may indeed be characteristic of the AMG epithelium. In the absence of a CA specifically expressed in the AMG cells, transported bicarbonate may simply be shuttled to the gut lumen where it could then be deprotonated to the double anion carbonate. Carbonate has a pKa in excess of 10 and is likely to be a major contributor to the alkaline luminal pH (Boudko et al., 2001).

Our results also show that the CA IV-like CA in mosquito larvae is expressed in the tracheal system. Human CA IV was first purified to homogeneity from lung tissue (Zhu and Sly, 1990), where this cell-surface form of CA contributes to the elimination of gaseous CO_2 from the bicarbonate form transported by red blood cells. It is quite feasible that the GPI-linked CA of the mosquito larva expressed in the trachea performs a similar function. That is, tracheal GPI-linked CA may act to convert ionic and aqueous forms of the ubiquitous aerobic waste product into a gas for elimination by diffusion through the tubule system (e.g. Clements, 1992).

Although the mosquito CA isoforms display similar features to mammalian CA IV enzymes, such as a 5′ signal sequence, a hydrophobic 3′ tail and extracellular GPI expression, there is one striking difference in the amino acid composition of mosquito CA isoform active sites. The active site within all of the 14 characterized mammalian CA isoforms is tightly conserved. Three histidine residues (His-94, His-96 and His-119) are essential for CA activity through their coordinated binding of a required zinc molecule. The absence of one or more of these histidine residues results in inactive proteins called CA-related proteins (CA-RPs), as found in mammalian CA isoforms VIII, X and XI (Tashian et al., 2000). The mosquito CA IV-like isoforms contain all three of the required

histidine residues, along with all of the other 13 highly conserved residues found in most other CAs (refer to figs 1 and 10, Tashian, 1992; Sly et al., 1995; Tamai et al., 1996a,b). However, as the alignment shows in Fig. 1, there is a conserved gap within the mosquito isoform active sites that is not present in any of the mammalian active sites. Because this shortened active site was found in mosquito but was not found in any mammalian CA isoform, we expanded our bioinformatics analyses. The *Drosophila melanogaster* genome was found to contain 14 putative CA genes (ENSF00000000228), the same number found in *Anopheles gambiae*. Only one out of the 14 CA isoforms was discovered to contain the identical number of deleted amino acids within the same active site region (refer to Fig. 10). This *Drosophila* CA sequence (accession number CG3940-PA) may also be a GPI-linked isoform, due to the presence of a leucine-rich 5′ signal sequence and hydrophobic tail region.

We have previously shown that the application of CA-specific inhibitors dramatically decreases the alkalinity of the gut (i.e. pH), and in fact is lethal to the larval mosquitoes (Corena et al., 2002). We now present evidence that a CA found in the mosquito gut is most similar to the mammalian CA IV isozyme but contains a novel active site motif unlike any of the mammalian CA isoforms (Fig. 10). The finding of a novel CA active site within the mosquito may facilitate the construction of a mosquito-specific CA inhibitor for use in larval mosquito control. We are hopeful that our ongoing mosquito CA crystallization project will yield further

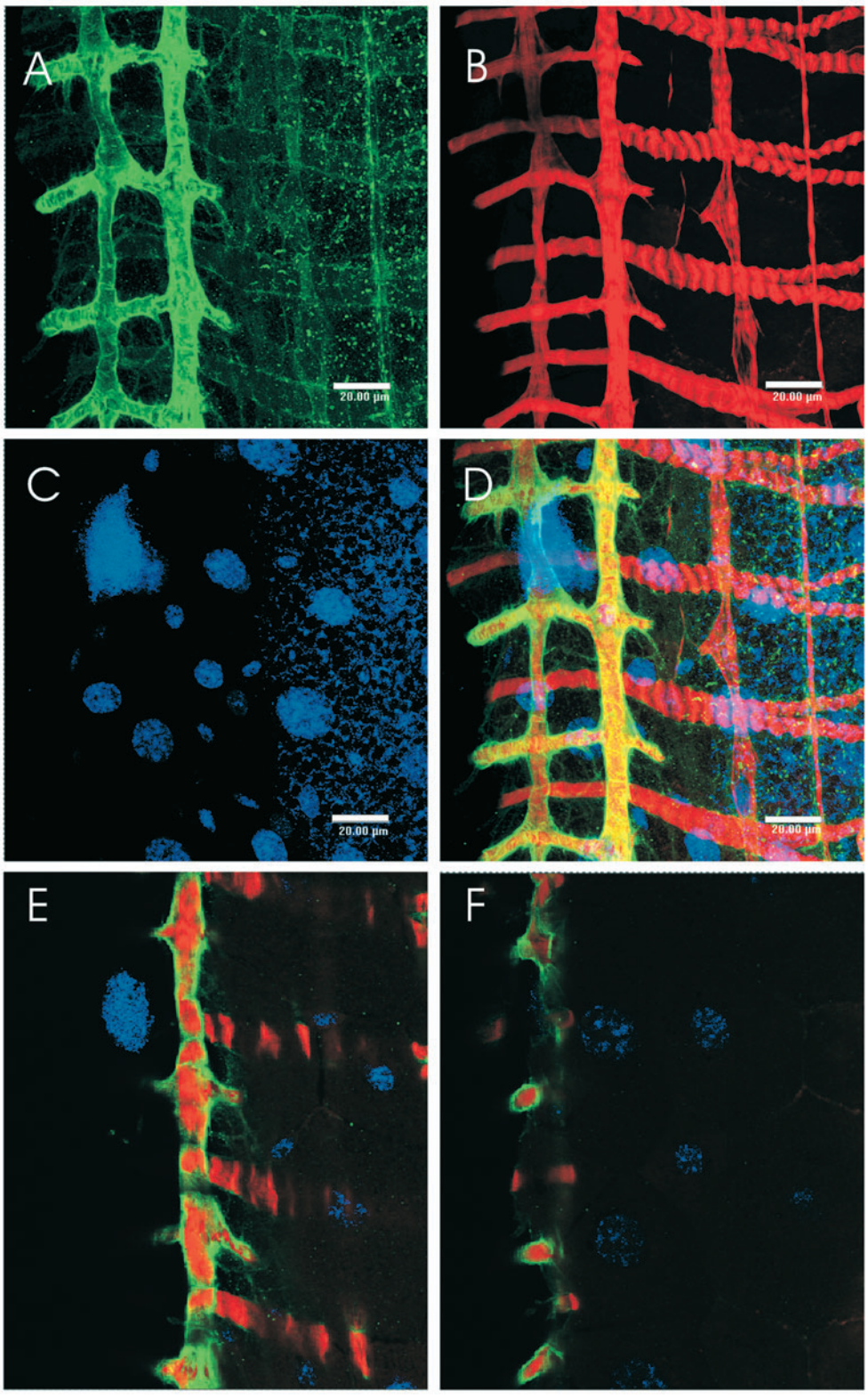

Fig. 9. High magnification confocal imaging of CA (green) actin (red) and DNA (blue) in the AMG of a fourth instar *Ae. albopictus* larva. (A–D) A maximum projection of a *z*-stack of images showing muscle fibers which are clearly labeled for CA (green in A and yellow in the overlay D) and ones that do not label for CA (only red in D). (E,F) Selected planes of focus from the same *z*-stack in three-color overlay. Note CA labeling (green) outlines the CA-positive muscles which are internally red, supporting a cell surface CA localization. All images are the same magnification so the magnification bars shown in A–D (20 µm) apply to all images.

```
AAL72625 Aedes CA       FVLDQMHFHWG--------SEHTIAGVRYGQELHMVHHDS
AAQ21365 Anoph CA       FVLDQMHFHWG--------SEHTLDDTRYGLELHLVHHDT
CG3940-PA Dros CA       FVVEQIHMHWW--------SEHTINDIRYPLEVHIVHRNT
P00915 Human CA I       YRLFQFHFHWG--STNEHGSEHTVDGVKYSAELHVAHWNS
P00918 Human CA II      YRLIQFHFHWG--SLDGQGSEHTVDKKKYAAELHLVHWNT
P07451 Human CA III     YRLRQFHLHWG--SSDDHGSEHTVDGVKYAAELHLVHWNP
P22748 Human CA IV      YQAKQLHLHWS--DLPYKGSEHSLDGEHFAMEMHIVHEKE
AAB47048 Human CA V     YRLKQFHFHWG--AVNEGGSEHTVDGHAYPAELHLVHWNS
CAC42429 Human CA VI    YIAQQMHFHWGGASSEISGSEHTVDGIRHVIEIHIVHYNS
P43166 Human CA VII     YRLKQFHFHWG--KKHDVGSEHTVDGKSFPSELHLVHWNA
JN0576 Human CA VIII    FELYEVRFHWG--RENQRGSEHTVNFKAFPMELHLIHWNS
AAH14950 Human CA IX    YRALQLHLHWG--AAGRPGSEHTVEGHRFPAEIHVVHLST
Q9NS85 Human CA X       HRLEEIRLHFG--SEDSQGSEHLLNGQAFSGEVQLIHYNH
AAH02662 Human CA XI    HRLSELRLLFG--ARDGAGSEHQINHQGFSAEVQLIHFNQ
AAH23981 Human CA XII   YSATQLHLHWG-NPNDPHGSEHTVSGQHFAAELHIVHYNS
BAA85002 Human CA XIV   YVAAQLHLHWG-QKGSPGGSEHQINSEATFAELHIVHYDS
                           . .          *** .       *... *
```

Fig. 10. Clustal alignment of CA protein sequences. All characterized human CA isoforms are presented along with predicted GPI-linked CA isoforms from *Aedes aegypti*, *Anopheles gambiae* and *Drosophila melanogaster*. Histidine residues that are required for the essential binding of zinc are shaded in blue. Note that one or more of these histidine residues are missing from the inactive human CA-related proteins VIII, X and XI, while all three histidines are present within the Dipteran sequences. These three species of Dipteran CAs contain a shortened active site region (marked by red dashes) when compared to any of the human (or other mammalian) CA sequences. This difference may provide a potential target for mosquito-specific CA inhibitors. Asterisks, identical residues; dots, conserved residues.

significant structural differences from the mammalian CA IV structure. These differences could then be utilized in the formulation of a mosquito-specific CA inhibitor.

Out of the 14 mammalian CA isoforms identified thus far as cytosolic, membrane-bound, secreted and mitochondrial, only CA IV has a GPI link to the cell membrane. The localization of this highly active mammalian isozyme to dynamic tissues such as the gut, brain, kidney and lung supports the important catalyst role of CA for the reversible hydration of CO_2. It should not be surprising that the gut of a mosquito, a highly alkaline and fluctuating system, has been found to contain a presumably active CA IV-like isoform as well. The single amino acid substitution of glycine-69 to glutamine is unique to rodent (rat and mouse) CA IV, and was found to be responsible for their reduced activity rate of only 10–20% of the human CA IV enzyme (Tamai et al., 1996). Mutating glutamine-69 to glycine within the rodent sequence resulted in almost three times greater CA activity (Tamai et al., 1996). Unlike the rodent sequences, both of the mosquito CA IV-like sequences display the high-activity glycine residue (Human CA IV numbering, refer to Fig. 1).

The task ahead is to decipher if a GPI-linked CA is better equipped to function in a highly dynamic system than other CA isoforms. Perhaps the GPI link affords the mosquito CA enzyme a characteristic advantage in buffering such an alkaline pH through its exclusively extracellular expression. Residing at the plasma membrane intrinsically affords this isozyme the best location for monitoring CO_2 and HCO_3^- concentration and flux in the hemolymph in the insect open circulatory system. Indeed, mammalian CA IV isoforms are expressed on membrane surfaces where large fluxes of CO_2 and/or HCO_3^- are expected (Sly, 2000). The most compelling ability of GPI-linked proteins is that they are known to elicit second messengers for signal transduction (Brown and Waneck, 1992). The alkaline pH of the larval mosquito gut was found to drop within 2–3 min after being narcotized or just simply handled (Dadd, 1975). This ‘handling effect’ lends itself to the prediction that larval mosquitoes exert nervous control over the generation of the gut lumen’s pH. Since a GPI-linked CA was localized within the mosquito gut and CNS tissue it seems possible that a GPI-linked CA may regulate the pH of mosquito guts through nervous control and a connection to a signal cascade.

This work was supported by NIH grant R01 AI45098 (to P.J.L.), and an Alumni Fellowship from the University of Florida (to T.J.S.).

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol* **215**, 403-410.

Baird, T. T., Jr, Waheed, A., Okuyama, T., Sly, W. S. and Fierke, C. A. (1997). Catalysis and inhibition of human carbonic anhydrase IV. *Biochemistry* **36**, 2669-2678.

Berenbaum, M. (1980). Adaptive significance of midgut pH in larval Lepidoptera. *Am. Nat.* **115**, 138-146.

Boudko, D. Y., Moroz, L. L., Harvey, W. R., and Linser, P. J. (2001). Alkalinization by chloride/bicarbonate pathway in larval mosquito midgut. *Proc. Natl. Acad. Sci. USA* **98**, 15354-15359.

Brown, D. and Waneck, G. L. (1992). Glycosyl-phosphatidylinositol-anchored membrane proteins. *J. Am. Soc. Nephrol*. **3**, 895-906.

Chegwidden, W. R. and Carter, N. D. (2000). Introduction to the carbonic anhydrases. In *The Carbonic Anhydrases – New Horizons* (ed. W. R. Chegwidden, N. D. Carter and Y. H. Edwards), pp. 13-28. Basel, Switzerland: Birkhauser Verlag.

Clark, T. M., Koch, A. and Moffett, D. F. (1999). The anterior and posterior ‘stomach’ regions of larval *Aedes aegypti* midgut: regional specialization of ion transport and stimulation by 5-hydroxytryptamine. *J. Exp. Biol.* **202**, 247-252.

Clements, A. N. (1992). *The Biology of Mosquitoes*. London, New York: Chapman and Hall.

Corena, M. P., Seron, T. J., Lehman, H. K., Ochrietor, J. D., Kohn, A., Tu, C. and Linser, P. J. (2002). Carbonic anhydrase in the midgut of larval *Aedes aegypti*: cloning, localization and inhibition. *J. Exp. Biol.* **205**, 591-602.

Dadd, R. H. (1975). Alkalinity within the midgut of mosquito larvae with alkaline-active digestive enzymes. *J. Insect Physiol.* **21**, 1847-1853.

Eisenhaber, B., Bork, P. and Eisenhaber, F. (1999). Prediction of potential GPI-modification sites in proprotein sequences. *J. Mol. Biol.* **292**, 741-758.

Geer, L. Y., Domrachev, M., Lipman, D. J. and Bryant, S. H. (2002). CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619-1623.

Hogue, C. W. (1997). Cn3D: a new generation of three-dimensional molecular structure viewer. *Trends Biochem. Sci.* **22**, 314-316.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Rineiro, J. M., Wides, R. et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129-149.

Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R. R., Ponting, C. P. and Bork, P. (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**, 242-244.

Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., Geer, L. Y. and Bryant, S. H. (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**, 281-283.

Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A. (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* **27**, 1558-1560.

Okuyama, T., Waheed, A., Kusumoto, W., Zhu, X. L. and Sly, W. S. (1995). Carbonic anhydrase IV: role of removal of C-terminal domain in glycosylphosphatidylinositol anchoring and realization of enzyme activity. *Arch. Biochem. Biophys*. **320**, 315-22.

Ridgway, R. L. and Moffet, D. F. (1986). Regional differences in the histochemical localization of carbonic anhydrase in the midgut of tobacco hornworm (*Manduca sexta*). *J. Exp. Zool*. **237**, 407-412.

Sly, W. S. (2000). The membrane carbonic anhydrases: from CO_2 transport to tumor markers. In *The Carbonic Anhydrases – New Horizons* (ed. W. R. Chegwidden, N. D. Carter and Y. H. Edwards), pp. 95-104. Basel, Switzerland: Birkhauser Verlag.

Sly, W. S. and Hu, P. Y. (1995). Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* **64**, 375-401.

Stams, T., Nair, S. K., Okuyama, T., Waheed, A., Sly, W. S. and Christianson, D. W. (1996). Crystal structure of the secretory form of membrane-associated human carbonic anhydrase IV at 2.8-A resolution. *Proc. Natl. Acad. Sci. USA* **93**, 13589-94.

Sterling, D., Reithmeier, R. A. and Casey, J. R. (2001). Carbonic anhydrase: in the driver's seat for bicarbonate transport. *Jop.* **2**, 165-70.

Tamai, S., Cody, L. B. and Sly, W. S. (1996a). Molecular cloning of the mouse gene coding for carbonic anhydrase IV. *Biochem. Genet*. **34**, 31-43.

Tamai, S., Waheed, A., Cody, L. B. and Sly, W. S. (1996b). Gly-63–>Gln substitution adjacent to His-64 in rodent carbonic anhydrase IVs largely explains their reduced activity. *Proc. Natl. Acad. Sci. USA* **93**, 13647-52.

Tashian, R. E. (1992). Genetics of the mammalian carbonic anhydrases. *Adv. Genet*. **30**, 321-56.

Tashian, R. E., Hewett-Emmett, D., Carter, N. and Bergenhem, N. C. (2000). Carbonic anhydrase (CA)-related proteins (CA-RPs), and transmembrane proteins with CA or CA-RP domains. In *The Carbonic Anhydrases – New Horizons* (ed. W. R. Chegwidden, N. D. Carter and Y. H. Edwards), pp. 105-120. Basel, Switzerland: Birkhauser Verlag.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-80.

Waheed, A., Okuyama, T., Heyduk, T. and Sly, W. S. (1996). Carbonic anhydrase IV: purification of a secretory form of the recombinant human enzyme and identification of the positions and importance of its disulfide bonds. *Arch. Biochem. Biophys.* **333**, 432-438.

Westerfield, M. (1994). *The Zebrafish Book: A guide for the laboratory use of zebrafish (*Brachydanio rerio*)*, pp. 9.16-9.21. Eugene, OR, USA: University of Oregon Press.

Wetzel, P. and Gros, G. (2000). Carbonic anhydrases in striated muscles. In *The Carbonic Anhydrases, New Horizons* (ed. W. R. Chegwidden, N. D. Carter and Y. H. Edwards), pp 375-399. Basel, Boston, Berlin: Birkhauser Verlag.

Wistrand, P. J. (1984). Properties of membrane-bound carbonic anhydrase. *Ann. NY Acad. Sci.* **429**, 195-206.

Zhang, Y. and Frohman, M. A. (1997). Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs. *Methods Mol. Biol.* **69**, 61-87.

Zhu, X. L. and Sly, W. S. (1990). Carbonic anhydrase IV from human lung. Purification, characterization, and comparison with membrane carbonic anhydrase from human kidney. *J. Biol. Chem.* **265**, 8795-801.

Zhuang, Z., Linser, P. J. and Harvey, W. R. (1999). Antibody to H(+) V-ATPase subunit E colocalizes with portasomes in alkaline larval midgut of a freshwater mosquito (*Aedes aegypti*). *J. Exp. Biol*. **202**, 2449-2460.