

RESEARCH ARTICLE

Open Access

# Characterization of the first beta-class carbonic anhydrase from an arthropod (*Drosophila melanogaster*) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates

Leo Syrjänen<sup>1,2,3\*</sup>, Martti Tolvanen<sup>1</sup>, Mika Hilvo<sup>1</sup>, Ayodeji Olatubosun<sup>1</sup>, Alessio Innocenti<sup>4</sup>, Andrea Scozzafava<sup>4</sup>, Jenni Leppiniemi<sup>1</sup>, Barbara Niederhauser<sup>1</sup>, Vesa P Hytönen<sup>1</sup>, Thomas A Gorr<sup>5</sup>, Seppo Parkkila<sup>1,2,3</sup>, Claudiu T Supuran<sup>4</sup>

## Abstract

**Background:** The  $\beta$ -carbonic anhydrase (CA, EC 4.2.1.1) enzymes have been reported in a variety of organisms, but their existence in animals has been unclear. The purpose of the present study was to perform extensive sequence analysis to show that the  $\beta$ -CAs are present in invertebrates and to clone and characterize a member of this enzyme family from a representative model organism of the animal kingdom, e.g., *Drosophila melanogaster*.

**Results:** The novel  $\beta$ -CA gene, here named *DmBCA*, was identified from FlyBase, and its orthologs were searched and reconstructed from sequence databases, confirming the presence of  $\beta$ -CA sequences in 55 metazoan species. The corresponding recombinant enzyme was produced in Sf9 insect cells, purified, kinetically characterized, and its inhibition was investigated with a series of simple, inorganic anions. Holoenzyme molecular mass was defined by dynamic light scattering analysis and gel filtration, and the results suggested that the holoenzyme is a dimer. Double immunostaining confirmed predictions based on sequence analysis and localized *DmBCA* protein to mitochondria. The enzyme showed high CO<sub>2</sub> hydratase activity, with a  $k_{cat}$  of  $9.5 \times 10^5 \text{ s}^{-1}$  and a  $k_{cat}/K_M$  of  $1.1 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$ . *DmBCA* was appreciably inhibited by the clinically-used sulfonamide acetazolamide, with an inhibition constant of 49 nM. It was moderately inhibited by halides, pseudohalides, hydrogen sulfide, bisulfite and sulfate ( $K_i$  values of 0.67 - 1.36 mM) and more potently by sulfamide ( $K_i$  of 0.15 mM). Bicarbonate, nitrate, nitrite and phenylarsonic/boronic acids were much weaker inhibitors ( $K_i$ s of 26.9 - 43.7 mM).

**Conclusions:** The *Drosophila*  $\beta$ -CA represents a highly active mitochondrial enzyme that is a potential model enzyme for anti-parasitic drug development.

## Background

Carbonic anhydrases (CAs, EC 4.2.1.1) catalyze the reversible hydration of carbon dioxide according to the following reaction:  $\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{HCO}_3^- + \text{H}^+$  [1]. CAs are zinc-containing metalloenzymes, except for the  $\zeta$  form, which uses cadmium [2,3] as alternative metal cofactor. Additionally  $\gamma$ -CAs contain iron *in vivo*, at least in anaerobic *Archaea* [4,5]. The reaction catalyzed by CAs is crucial in the regulation of acid-base balance

in organisms. In addition, CAs participate in many other physiological processes such as bone resorption in vertebrates, gluconeogenesis, production of body fluids, and transport of CO<sub>2</sub> and HCO<sub>3</sub><sup>-</sup> to name but a few [1].

To date, five different classes of CAs have been identified:  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  and  $\zeta$  [6]. A previously known  $\epsilon$ -CA class [7] has been reclassified as a new type of  $\beta$ -CA based on its crystallographic structure [8], which shows a fold nearly identical to those of the archaeal cab-type [9] and plant-type [10]  $\beta$ -CAs. In  $\zeta$ -CA, the geometry of the active site is nearly identical to that of  $\beta$ -CAs, and there is also some similarity in the fold, which has led to the

\* Correspondence: leo.syrjanen@uta.fi

<sup>1</sup>Institute of Medical Technology, University of Tampere and Tampere University Hospital, Tampere, Finland

suggestion that even  $\zeta$ -CAs might represent a distantly diverged subtype of  $\beta$ -CAs [3].

The  $\zeta$ -CAs are found only in diatoms, and the  $\delta$ -CAs in diatoms and other marine phytoplankton, whereas the three major classes ( $\alpha$ ,  $\beta$  and  $\gamma$ ) are widely distributed.  $\gamma$ -CA orthologs are present in *Archaea*, *Bacteria* and plants [6,11] but have been secondarily lost in animals and fungi [12] (M. Tolvanen, unpublished observation). The  $\alpha$  class is missing from *Archaea* (M. Tolvanen, unpublished observation) but is nearly ubiquitously present in *Bacteria* and *Eukarya*, with the exception of *Fungi*, in which only filamentous ascomycetes have  $\alpha$ -CAs [12]. In plants and animals,  $\alpha$ -CAs exist as numerous isoforms. For example, 16 and 15  $\alpha$ -CAs have been described in non-primate mammals and primates including humans, respectively [1,13].  $\beta$ -CAs appear to be the class with the widest distribution as they have been described in multiple lineages throughout the tree of life, including most species belonging to the *Archaea* and *Bacteria* domains and probably all species of plants and fungi among *Eukarya* ([14] and M. Tolvanen, unpublished data).  $\beta$ -CAs have also been characterized in a number of human pathogens such as fungi/yeasts (e.g., *Candida albicans*, *Candida glabrata*, *Cryptococcus neoformans* and *Saccharomyces cerevisiae*) [15-20] and bacteria (*Helicobacter pylori*, *Mycobacterium tuberculosis*, *Haemophilus influenzae*, *Brucella suis* etc.) [21-25], and their inhibition profiles by various agents such as sulfonamides, anions, carboxylates and boronic acids have been explored [15-24,26-30]. Given that  $\beta$ -CA genes appear to be missing in invertebrates, novel antimicrobial compounds based on the inhibition of  $\beta$ -CAs from pathogenic organisms may soon become available.

Although  $\beta$ -CAs catalyze the same reaction as  $\alpha$ -CAs and other CA forms, important structural differences between these classes exist. First, instead of functioning as obligate monomers like most  $\alpha$ -forms or trimers like  $\gamma$ -forms,  $\beta$ -CAs are found in many oligomerization states. Crystal structures of dimeric, tetrameric and octameric  $\beta$ -CAs have been reported [9,10,31]. In the active site of  $\beta$ -CAs, the zinc atom is coordinated by one histidine and two cysteine residues instead of three histidine residues present in  $\alpha$ -,  $\gamma$ - and  $\delta$ -CAs [32]. Despite this difference,  $\beta$ -class CAs basically share the same molecular mechanism for reversible hydration of carbon dioxide as  $\alpha$ -class CAs [9].  $\beta$ -CAs possess a highly conserved dyad comprising an aspartate and an arginine residue that seem to be crucial for the catalytic mechanism since mutation of these residues severely reduces the catalytic activity of the enzyme [33]. The aspartate makes a hydrogen bond with the Zn(II) coordinated water molecule, activating it for nucleophilic attack of the CO<sub>2</sub> molecule [29].

$\beta$ -CAs have been reported in many photosynthetic organisms, including plants and algae [6]. In plants,  $\beta$ -class CAs are strongly expressed in both roots and green tissues and are located in chloroplasts, cytoplasm and mitochondria with isozyme-specific patterns [13,34]. The cytoplasmic and chloroplastic CAs are suggested to be crucial in CO<sub>2</sub> accumulation and carbon fixation. The green algae *Chlamydomonas reinhardtii* also contains  $\beta$ -class CAs that are localized to mitochondria [35]. In fungi like *Cryptococcus neoformans* and *Candida albicans*,  $\beta$ -CAs have an important role in CO<sub>2</sub> sensing and, consequently, in the pathogenesis of these species [36]. In addition, many fungal  $\beta$ -CAs have been shown to be mitochondrial [12]. These and many other similar findings confirm that  $\beta$ -CAs are physiologically important enzymes with variable localization and function like  $\alpha$ -CAs in vertebrates.

The presence of  $\beta$ -CAs in the animal kingdom has been controversial or ignored due to the paucity and poor quality of the available sequences [6,14]. Here we show, however, that  $\beta$ -CAs are widespread among invertebrates. The aim of this study was to express, purify and characterize a  $\beta$ -CA enzyme from fruit fly (*D. melanogaster*), a commonly used model organism in biological sciences. The recombinant enzyme was produced in Sf9 insect cells using the baculovirus/insect cell expression system. Along with the characterization of the first arthropod  $\beta$ -CA, this study also describes its inhibition profile with inorganic anions. These results might open new strategies for developing novel anti-parasitic drugs against common diseases like schistosomiasis and malaria.

## Results

### Sequence analysis

We found  $\beta$ -CAs in all complete non-chordate animal genomes and in almost all invertebrates with at least 20,000 EST sequences in the NCBI database, plus in some with fewer sequences. We confirmed the existence of  $\beta$ -CA sequences in *Placozoa*, *Cnidaria*, *Platyhelminthes*, *Nematoda*, *Arthropoda*, and *Annelida*, and even in many classes of *Deuterostomia*, namely in *Hemichordata*, *Echinodermata*, and *Xenoturbellida*. The only major taxon of *Protostomia* with poor evidence for  $\beta$ -CA is *Mollusca*, in which we found only one EST to match 55 residues in other  $\beta$ -CAs. Of special interest were numerous pathogenic helminth species with complete  $\beta$ -CA sequences or substantial fragments, namely, filaria-causing *Brugia malayi*, mouse whipworm *Trichuris muris*, dog hookworm *Ancylostoma caninum*, and the flukes *Schistosoma mansoni* and *Schistosoma sinensis*.

In the case of chordates, the existence of a functional  $\beta$ -CA is currently unclear. The genomes and NCBI EST

sequence collections (as of 1<sup>st</sup> Oct, 2009) of *Ciona intestinalis* and *Ciona savignyi* (tunicates) lack  $\beta$ -CA, whereas we found two recognizable but incomplete  $\beta$ -CA sequences in the genome of the cephalochordate *Branchiostoma floridae* (of one locus from both haplotypes in the genome). The encoded proteins seem to lack more than 60 residues at the N-terminus, including the active site. In addition, there are two even less complete partial EST transcripts, GenBank BW824885 and BW803919. The latter contains an unrelated sequence in place of the active-site-containing exon. We can only conclude that it remains an open question whether  $\beta$ -CA in *B. floridae* is a pseudogene or an incompletely sequenced active gene.

Our survey discovered and assembled 38 seemingly complete and correct  $\beta$ -CA sequences from the genome and sequence databases of 33 metazoan species, including improved gene models for sequences already in sequence databases. Fragmentary  $\beta$ -CA sequences were found in additional 22 species. Multiple sequence alignment of all animal  $\beta$ -CAs shows perfect conservation of the known active site motifs CxDxR and HxxC and several other key residues. Figure 1 shows an alignment of the first 120 residues of selected  $\beta$ -CA sequences, including the N-terminal mitochondrial targeting peptide and active site regions (See Table 1 for identification of species). Of the active-site residues indicated below the alignment, two cysteines and one histidine are zinc-binding residues. Additional files 1 and 2 show the full alignment of the same sequences and of all of the identified sequences, respectively. The phylogenetic tree of the selected animal  $\beta$ -CA sequences is shown in Figure 2 (See Table 1 for identification of species). The tree indicates that the duplication of  $\beta$ -CA genes in nematodes is specific for the nematode lineage. Out of the two copies, the one labeled BCA2 is more strongly

conserved, as shown by shorter branches in the tree of Figure 2 and in trees we made with all available sequences. The placement of the  $\beta$ -CA from acorn worm (*S. kowalevskii*, a hemichordate) seems contrary to conventional invertebrate taxonomy, but since some of the bootstrap values are under 50%, the tree is not perfectly resolved outside the insect and nematode blocks.

No  $\beta$ -CA sequences exist in vertebrate genomes. A false positive is found in the *X. tropicalis* genome, scaffold\_1719, which we attribute to bacterial contamination, since the  $\beta$ -CA sequence and every other gene in this scaffold are highly similar (>80% identical) with known genes of *Pseudomonas*-related bacteria.

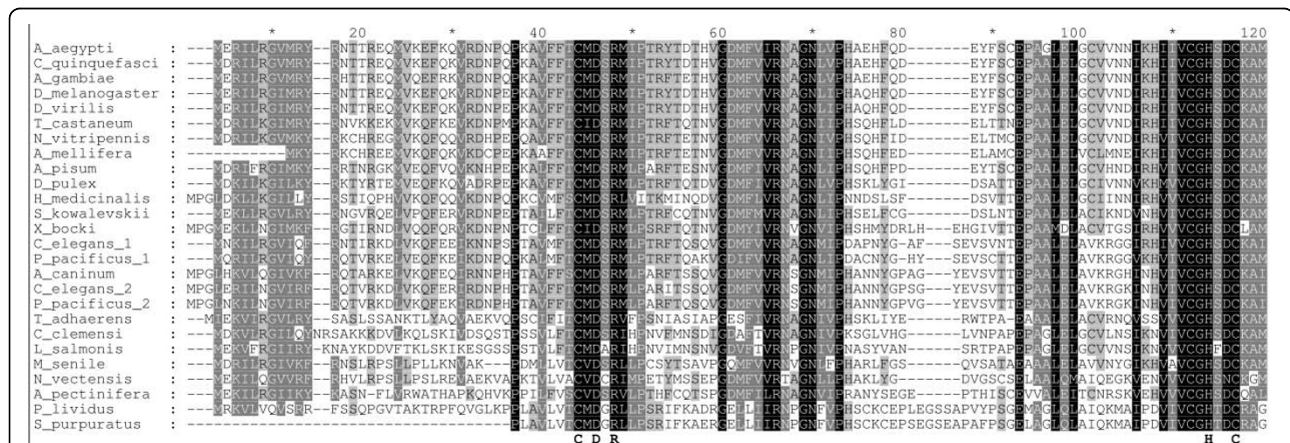
All of the complete sequences are classified as  $\beta$ -CA domains of type "BCA\_CladeB" in the NCBI Conserved Domain Database. All plant  $\beta$ -CAs are also classified as BCA\_CladeB domains, so animal  $\beta$ -CAs can well be said to be "plant-type  $\beta$ -CAs". There are also hundreds of bacterial  $\beta$ -CA sequences of type "BCA\_CladeB", and some of them cluster closest to animal sequences, and some closest to plant sequences in phylogenetic trees (data not shown).

#### Expression of $\beta$ -CA in Sf9 insect cells

Sf9 insect cells were transfected with the  $\beta$ -CA gene (*DmBCA*) obtained from *D. melanogaster* cDNA. The amount of protein obtained from 500 ml of culture supernatant was approximately 1 mg. According to SDS-PAGE, the relative molecular mass of DmBCA was approximately 28 and 27 kDa before and after thrombin treatment, respectively (Figure 3).

#### Subcellular localization of DmBCA

When the collection of 38 complete animal  $\beta$ -CA sequences was analyzed for subcellular targeting with



**Figure 1 Alignment of 26  $\beta$ -CA sequences from invertebrate species.** Only the first 120 residues are shown. Active-site residues C, D, R, H and C are highlighted below the alignment. These residues are found in all  $\beta$ -CAs. See Table 1 for identification of species.

**Table 1 Identification of sequences and species in the sequence alignment and phylogenetic tree**

Abbreviation	Full name	Common name and classification	gi or accession numbers
A_aegypti	<i>Aedes aegypti</i>	Yellow fever mosquito, insects	157110803 + 77891004
C_quinquefasciatus	<i>Culex quinquefasciatus</i>	Southern house mosquito, insects	170043321
A_gambiae	<i>Anopheles gambiae</i>	Malaria mosquito, insects	57968460
D_melanogaster	<i>Drosophila melanogaster</i>	Fruit fly, insects	24645213
D_virilis	<i>Drosophila virilis</i>	Fruit fly, insects	194152748
T_castaneum	<i>Tribolium castaneum</i>	Red flour beetle, insects	91084165
N_vitripennis	<i>Nasonia vitripennis</i>	Wasp, insects	156547528
A_mellifera	<i>Apis mellifera</i>	Honeybee, insects	110764310
A_pisum	<i>Acyrtosiphon pisum</i>	Pea aphid, insects	193713675
D_pulex	<i>Daphnia pulex</i>	Water flea, crustaceans	FE417346 + FE409868
H_medicinalis	<i>Hirudo medicinalis</i>	Medical leech, Annelid worms	EY481200 + EY505051 + EY490477
S_kowalevskii	<i>Saccoglossus kowalevskii</i>	Acorn worm, hemichordate	187043763
X_bocki	<i>Xenoturbella bocki</i>	Xenoturbella, Xenoturbellidae	117195962
C_elegans_1	<i>Caenorhabditis elegans</i> BCA1	Nematode	NP_741809.1
P_pacificus_1	<i>Pristionchus pacificus</i> BCA1	Nematode	FG098717 + GeneWise
A_caninum	<i>Ancylostoma caninum</i>	Dog hookworm, nematodes	FC551456 + FC550353
C_elegans_2	<i>Caenorhabditis elegans</i> BCA2	Nematode	NP_001041015
P_pacificus_2	<i>Pristionchus pacificus</i> BCA2	Nematode	FG106379 + GeneWise
T_adhaerens	<i>Trichoplax adhaerens</i>	Trichoplax, Placozoa	190581916
C_clemensi	<i>Caligus clemensi</i>	Sea louse, crustaceans	225719368
L_salmonis	<i>Lepeophtheirus salmonis</i>	Sea louse, crustaceans	225713548
M_senile	<i>Metridium senile</i>	Sea anemone, Cnidaria	FC835283
N_vectensis	<i>Nematostella vectensis</i>	Sea anemone, Cnidaria	XP_001632619
A_pectinifera	<i>Asterina pectinifera</i>	Starfish, Echinodermata	DB424979 + DB440523
P_lividus	<i>Paracentrotus lividus</i>	Sea urchin, Echinodermata	139313180 + 139245724
S_purpuratus	<i>Strongylocentrotus purpuratus</i>	Purple sea urchin, Echinodermata	XP_001189115

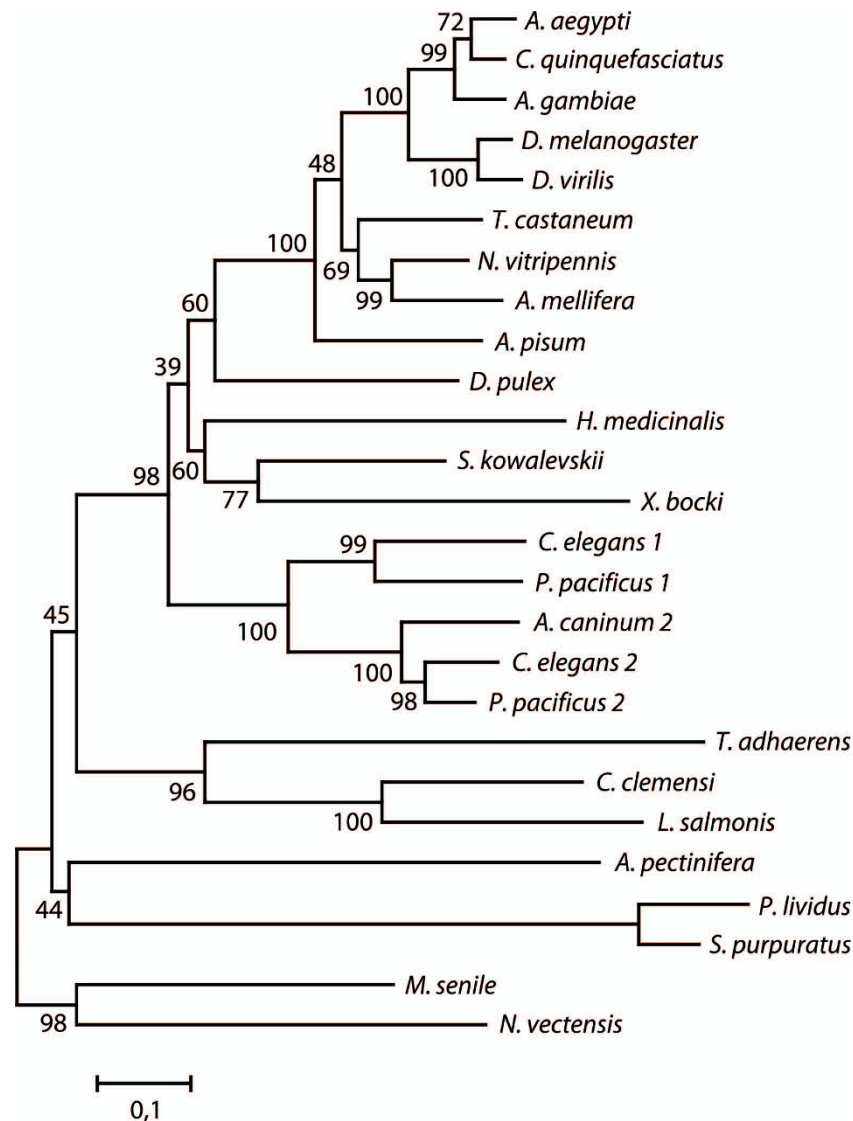
TargetP, 22 sequences were predicted to be mitochondrial, with low reliability, and most of the remaining sequences were predicted to be cytoplasmic, again with low reliability. The results from Wolf PSORT, Secretome 2.0 and MitoProt II v. 1.101 also supported the hypothesis that it is a mitochondrial enzyme, with a possibility for non-classical secretion (without a signal peptide) suggested by Secretome 2.0. A mitochondrial targeting signal sequence is also predicted in the N-terminus (prediction lengths varying from 14 in MitoProt to 49 in TargetP). Based on these findings, a DmBCA-GFP construct was designed to study the subcellular localization of the protein in Sf9 cells.

According to our experiments on the DmBCA-GFP fusion protein, DmBCA is indeed a mitochondrial protein, supporting the predictions made by bioinformatic tools. Figure 4A shows the DmBCA-GFP recombinant protein in Sf9 insect cells in which the positive signal was located in intracellular granular structures. Figure 4B shows the same cells labeled with a mitochondrial marker, MitoTracker Red CMXros™. Figure 4C presents an overlay of the previous panels, demonstrating the colocalization of DmBCA-GFP and MitoTracker Red CMXros™.

#### Catalytic activity and inhibition of DmBCA

DmBCA purified from Sf9 cells was kinetically analyzed in the presence or absence of acetazolamide or inorganic anions. The kinetic parameters of DmBCA ( $k_{cat}$  and  $k_{cat}/K_m$ ) were then compared with those of the thoroughly investigated CAs such as the cytosolic and ubiquitous human isozymes  $\alpha$ -CA I (HCA I) and II (HCA II), as well as the recently described fungal  $\beta$ -CAs, *C. neoformans* Can2, *C. albicans* Nce103, *S. cerevisiae* CA (ScCA) and *C. glabrata* CA (CgCA). All of these fungal enzymes are orthologous to *Nce103*, a protein found in all fungi/yeasts studied to date (Table 2).

One can appreciate from the data shown in Table 2 that DmBCA, similar to the other recently investigated  $\alpha$ - and  $\beta$ -CAs, possesses considerable CO<sub>2</sub> hydratase activity. A  $k_{cat}$  of  $9.5 \times 10^5 \text{ s}^{-1}$  and a  $k_{cat}/K_m$  of  $1.1 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$  make DmBCA one of the most efficient  $\beta$ -CA described to date. Data from Table 2 also show that DmBCA was appreciably inhibited by the clinically-used sulfonamide acetazolamide (5-acetamido-1,3,4-thiadiazole-2-sulfonamide), with an inhibition constant of 49 nM. Table 3 shows the DmBCA inhibition screening data with anionic physiological species (such as chloride, bicarbonate, sulfate, etc.) as well as other non-



**Figure 2** Minimum-evolution tree of 26 invertebrate  $\beta$ -CA sequences. Bootstrap consensus tree of 1,000 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. See Table 1 for identification of species.

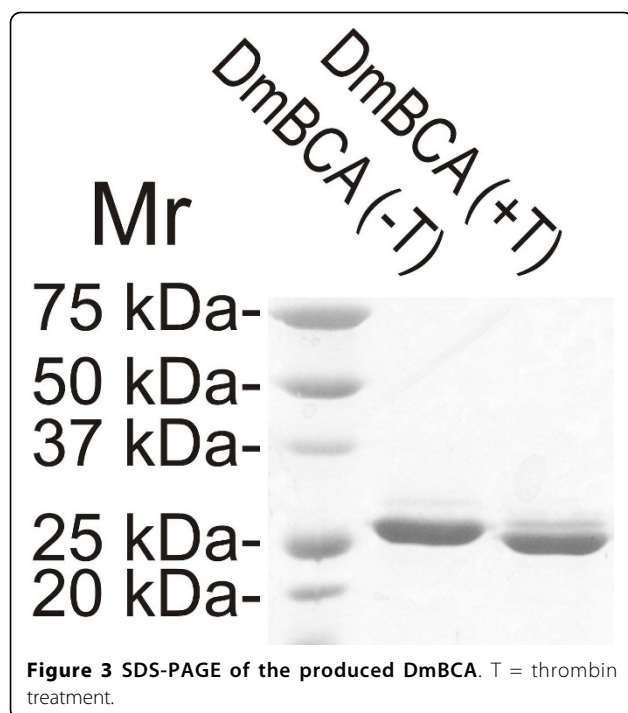
physiological anions. Note that similar to other investigated  $\alpha$ - and  $\beta$ -CAs, DmBCA has an inhibition profile with anions characteristic only for this enzyme. Several species such as bicarbonate, nitrate, nitrite, perchlorate, phenylboronic acid and phenylarsonic acid behaved as weak inhibitors ( $K_i$ s in the range of 22.4 - 200 mM), whereas other anions inhibited the enzyme more potently (e.g., cyanide, cyanate, and sulfamide, with inhibition constants in the range of 150 - 730  $\mu$ M).

#### Dynamic light scattering and gel filtration analysis

The hydrodynamic diameters of the proteins were measured by dynamic light scattering (DLS). At a temperature range from 4°C to 37°C the hydrodynamic diameter

of DmBCA was  $6.3 \pm 0.8$  nm and the diameter was found to slightly increase during the heating. According to the hydrodynamic diameter the average molecular weight of DmBCA was estimated to be 48.9 kDa, the lowest estimate being 35.4 kDa, and the highest 64.9 kDa ( $n = 30$ ). When temperature was raised above 43°C the hydrodynamic diameter of DmBCA started to clearly increase and DmBCA had a transition state at temperature range 46-49°C, where large protein aggregates appeared. This might be associated with protein denaturation.

At the temperature range from 4°C to 37°C the hydrodynamic diameter of HCA II was  $3.9 \pm 0.6$  nm leading to estimated average molecular weight of 16.1 kDa, the



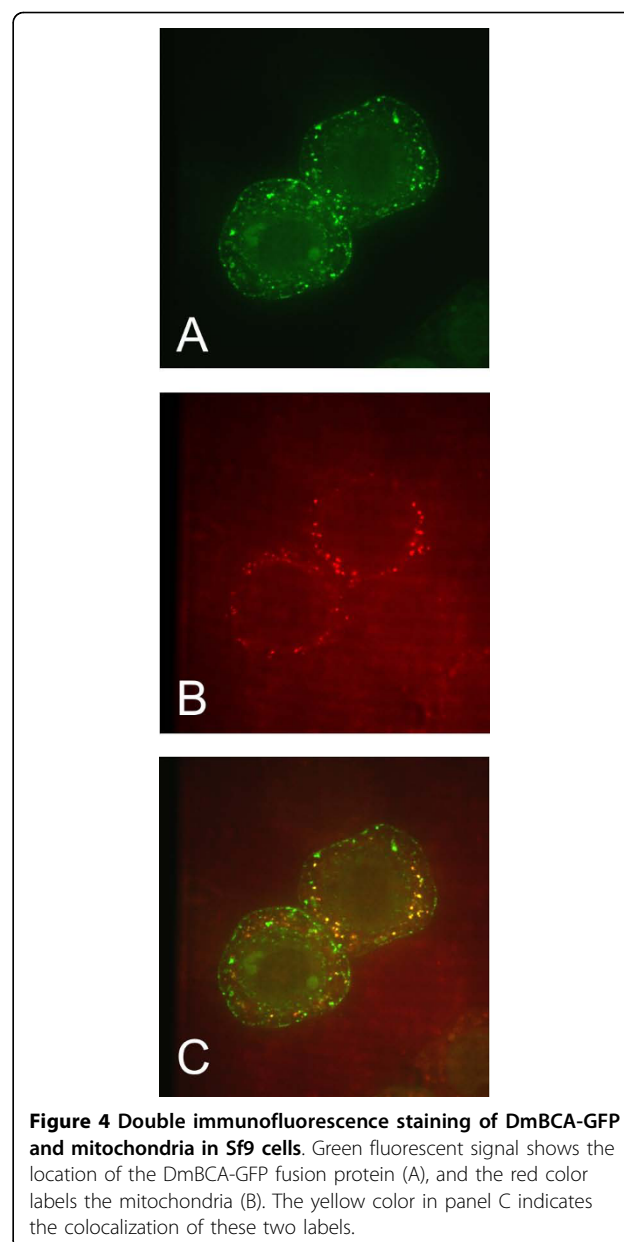
lowest estimate being 10.8 kDa, and the highest 22.8 kDa ( $n = 12$ ). When HCA II was heated from 4°C to 43°C there was a slight increase in hydrodynamic diameter. However, no transition states were observed.

The estimated molecular weights determined by DLS are in agreement with the values obtained by analytical gel filtration which gave a molecular weight estimate for DmBCA of  $38.1 \pm 3.0$  kDa and for HCA II of  $23.4 \pm 4.0$  kDa. The molecular weight of HCA II monomer is 29 kDa. Therefore, gel filtration analysis appears to underestimate the molecular size of CAs. Overall, DLS and gel filtration analyses suggest dimeric state for DmBCA whereas HCA II appears predominantly monomeric in solution.

### Discussion

In the present study, we identified and characterized a novel  $\beta$ -CA enzyme (DmBCA) from an arthropod, *D. melanogaster*. Although  $\beta$ -CAs have been previously reported in *Archaea* and *Bacteria* domains, in addition to plants and fungi in *Eukarya*, our results suggest the widespread occurrence of at least a single-copy  $\beta$ -CA gene among animal species distinct from chordates. The loss of  $\beta$ -CA gene in the chordate lineage may have occurred either in the last common ancestor of all chordates or in the last common ancestor of tunicates and craniates. Whether cephalochordates have a functional  $\beta$ -CA gene remains an unresolved question.

Using bioinformatic tools, we discovered a single  $\beta$ -CA gene in most invertebrates with reasonable sequence



coverage. The exception is nematodes, which seem to have two  $\beta$ -CA genes in their genomes. A very recent article [37] reported the cloning and characterization of *C. elegans*  $\beta$ -CAs and the authors found that one of the two isoforms, BCA-1, has no activity and does not work in complementation assay. This might, however, be due to incorrect sequence: the sequence they show for *C. elegans* BCA-1 contains the coding sequence of the preceding *MTP18* gene fused to the  $\beta$ -CA reading frame. This sequence is a result of an incorrect gene prediction which has no support at the mRNA level and which remained in WormBase and UniProt until May 2008, subsequently corrected from our initiative. The previously fused

**Table 2 Kinetic parameters for the CO<sub>2</sub> hydration reaction catalyzed by various CAs**

Isozyme	Activity level	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_m$ (M <sup>-1</sup> s <sup>-1</sup> )	$K_i$ (nM) <sup>c</sup>
HCA I <sup>a</sup>	moderate	$2.0 \times 10^5$	$5.0 \times 10^7$	250
HCA II <sup>a</sup>	very high	$1.4 \times 10^6$	$1.5 \times 10^8$	12
Can2 <sup>a</sup>	moderate	$3.9 \times 10^5$	$4.3 \times 10^7$	10.5
Nce103 <sup>a</sup>	high	$8.0 \times 10^5$	$9.7 \times 10^7$	132
ScCA <sup>a</sup>	high	$9.4 \times 10^5$	$9.8 \times 10^7$	82
CgCA <sup>a</sup>	moderate	$3.8 \times 10^5$	$4.8 \times 10^7$	11
DmBCA <sup>b</sup>	high	$9.5 \times 10^5$	$1.1 \times 10^8$	49

Kinetic parameters for the human cytosolic isozymes human HCA I and II ( $\alpha$ -class CAs) at 20°C and pH 7.5 in 10 mM HEPES buffer and 20 mM Na<sub>2</sub>SO<sub>4</sub>, in addition to the  $\beta$ -CAs Can2, Nce103 (from *C. neoformans* and *C. albicans*, respectively), ScCA (from *S. cerevisiae*), CgCA (from *C. glabrata*) and DmBCA (from *Drosophila melanogaster*) measured at 20°C, pH 8.3 in 20 mM Tris buffer and 20 mM NaClO<sub>4</sub> are shown. Inhibition data with the clinically used sulfonamide acetazolamide (5-acetamido-1,3,4-thiadiazole-2-sulfonamide) are also provided.

<sup>a</sup>Reference [30]

<sup>b</sup>This work

<sup>c</sup>acetazolamide

UniProt entry Q8T3C8 now represents only MTP18, and a new entry BCA1\_CAEEL contains the more plausible BCA-1 sequence. Since *C. elegans* BCA-1 has all of the active-site residues and is also well conserved in other nematodes, we think it is probable that both isozymes in nematodes would be functional  $\beta$ -CAs.

Our larger phylogenetic analysis (M. Tolvanen, unpublished data) and NCBI domain classification of animal  $\beta$ -CAs clearly show them to be “plant-type”  $\beta$ -CAs. The “plant-type” group also includes hundreds of bacterial  $\beta$ -CAs, and this group contains plant/bacterial and animal/bacterial subgroups. The presence of such polyphyletic subgroups may indicate horizontal gene transfer, and further investigations are underway to study this phenomenon.

Fasseas *et al.* [37] state that the 3D structure of their *C. elegans*  $\beta$ -CA models would resemble archaeal “cab-type”  $\beta$ -CAs, but we cannot agree with their conclusion. In our eyes, the models are nearly identical to the pea BCA (panel a in their Figure 2), and in our hands Swiss-Model <http://swissmodel.expasy.org/> in fact chooses the pea  $\beta$ -CA structure as a template for both insect and nematode  $\beta$ -CAs.

In the recent paper by Fasseas *et al.* [37], the enzyme activity of *C. elegans* Y116A8C.28 was reported to be typical for  $\beta$ -CAs, with a  $k_{cat}$  and  $k_{cat}/K_m$  of  $2.77 \times 10^4$  s<sup>-1</sup> and  $6,383 \times 10^5$  M<sup>-1</sup>s<sup>-1</sup>, respectively. These values are significantly lower than the  $k_{cat}$  of  $9.5 \times 10^5$  s<sup>-1</sup>, and  $k_{cat}/K_m$  of  $1.1 \times 10^8$  M<sup>-1</sup>s<sup>-1</sup> that we report here for DmBCA. The enzymatic activity of DmBCA is one of the highest reported so far for a  $\beta$ -CA, suggesting an important physiological role for this enzyme.

Our studies with different anionic CA inhibitors did not reveal specific inhibitors of DmBCA, which is in fact

**Table 3 Inhibition constants of anionic inhibitors against various CAs**

Inhibitor	$K_i$ [mM] <sup>#</sup>				
	HCA II	Nce103	ScCA	CgCA	DmBCA
F <sup>-</sup>	>300	0.69	2.85	0.36	0.80
Cl <sup>-</sup>	200	0.85	0.85	0.58	0.97
Br <sup>-</sup>	63	0.94	0.0108	27	1.04
I <sup>-</sup>	26	1.40	0.0103	42.4	1.18
CNO <sup>-</sup>	0.03	1.18	31.7	0.60	0.73
SCN <sup>-</sup>	1.6	0.65	55.6	0.73	1.28
CN <sup>-</sup>	0.02	0.011	16.8	1.12	0.67
N <sub>3</sub> <sup>-</sup>	1.5	0.52	27.9	1.03	1.12
HCO <sub>3</sub> <sup>-</sup>	85	0.62	0.78	0.086	26.9
NO <sub>3</sub> <sup>-</sup>	35	0.69	13.9	0.097	43.7
NO <sub>2</sub> <sup>-</sup>	63	0.53	0.46	0.088	28.6
HS <sup>-</sup>	0.04	0.37	0.33	0.10	1.01
HSO <sub>3</sub> <sup>-</sup>	89	0.54	0.33	0.10	1.29
SO <sub>4</sub> <sup>2-</sup>	>200	14.15	0.58	0.58	1.36
ClO <sub>4</sub> <sup>-</sup>	>200	>200	>200	>200	>200
H <sub>2</sub> NSO <sub>2</sub> NH <sub>2</sub>	1.13	0.30	0.0087	0.42	0.15
H <sub>2</sub> NSO <sub>3</sub> H <sup>§</sup>	0.39	0.70	0.84	0.11	2.45
Ph-B(OH) <sub>2</sub>	23.1	30.8	38.2	0.10	22.4
Ph-AsO <sub>3</sub> H <sub>2</sub> <sup>§</sup>	49.2	30.8	0.40	0.098	32.6

Inhibition constants against isozymes HCA II ( $\alpha$ -CA class), and the  $\beta$ -CAs Nce103 (from *C. albicans*), ScCA (from *S. cerevisiae*), CgCA (from *C. glabrata*) and DmBCA (*D. melanogaster*) for the CO<sub>2</sub> hydration reaction at 20°C are shown.

<sup>§</sup>As sodium salt.

<sup>#</sup> Errors in the range of 5-10% of the shown data, from three different assays, by a CO<sub>2</sub> hydration stopped-flow assay.

normal for such simple inorganic ions [38]. The inhibition profile of DmBCA was unique for this enzyme, however, and differed significantly from other  $\beta$ -CAs studied previously.

Indeed, fluoride, chloride, cyanate, cyanide and sulfamide showed submillimolar inhibitory activity against DmBCA, with inhibition constants in the range of 150  $\mu$ M - 970  $\mu$ M. Another groups of anions, including bromide, iodide, thiocyanate, azide, hydrogen sulfide, bisulfite and sulfate showed inhibition constants close to 1 mM ( $K_i$ s in the range of 1.01 - 1.36 mM), thus making them weak DmBCA inhibitors. Other anions such as bicarbonate, nitrate and nitrite, as well as the non-anionic species phenylboronic acid and phenylarsonic acid, were much less effective inhibitors, with inhibition constants of 22.4 - 43.7 mM. It is interesting to note that bicarbonate, a CA substrate, is a weak DmBCA inhibitor, but it appreciably inhibits the fungal enzymes Nce103 (*C. Albicans*), ScCA and CgCA, with  $K_i$ s in the submillimolar range (of 86  $\mu$ M - 0.78 mM). Even though the low apparent  $K_i$  value for bicarbonate could be a reflection of allosteric inactivation of the enzyme ([39,40]), the value is still in the same range with the other anion inhibitors, suggesting that their mechanism



of inhibition might be quite similar. The present inhibition data may suggest a different physiological role for DmBCA in the insect compared to the corresponding  $\beta$ -CAs in fungal or yeast species studied previously. One should also note that DmBCA has a completely different inhibition profile not only compared to other  $\beta$ -CAs but also compared to the highly investigated  $\alpha$ -CA of human origin, HCA II (Table 3).

Animal  $\beta$ -CAs constitute a separate subgroup in the  $\beta$ -CA family according to our bioinformatic analysis. Because the main target,  $\beta$ -CA, is present in many parasites and disease carrying vectors but absent in humans, this discovery clearly carries the significant potential for the design of parasite-specific  $\beta$ -CA inhibitors. Such drugs would presumably combine high specificity with a low incidence of side-effects in humans. These drugs could, for example, provide novel opportunities to restrict malaria propagation and to treat patients suffering from helminth infections.

Predictions of subcellular localization placed most, but not all, animal  $\beta$ -CAs in the mitochondria. Our hypothesis is that all of them will be mitochondrial based on three main reasons. First, we have direct experimental evidence of mitochondrial localization of DmBCA. Second, all of the TargetP predictions were either mitochondrial or cytoplasmic, both with low-reliability. Third, the neural network used in TargetP is trained with human and Arabidopsis mitochondrial proteins; thus, it is perhaps not perfectly adjusted to detect the signals in invertebrate mitochondrial transit peptides. Because of this mitochondrial localization of animal  $\beta$ -CAs, we speculate that the  $\beta$ -CA gene in chordates might have been silenced and ultimately lost after the emergence of mitochondrial  $\alpha$ -CAs (CA V) to substitute for this function. The presence of an active CA in mitochondria has been linked to maintaining fundamental metabolic functions such as gluconeogenesis, ureagenesis and lipogenesis [1,41]. Although our experiments were not focused on the role of the DmBCA enzyme, its mitochondrial localization and high enzymatic activity suggest that knockdown studies in *D. melanogaster* are warranted to further investigate the physiological function of  $\beta$ -CAs in animals.

The expression of DmBCA has been studied at mRNA level in two publicly available DNA microarray datasets. In FlyAtlas (<http://flyatlas.org/atlas.cgi?name=CG11967-RA>, [42]) expression levels in 17 adult and 8 larval *D. melanogaster* tissues are given. The highest upregulated values in adult are found in spermatheca (female), fat body, and heart. In larval tissues (third instar) downregulated or average expression levels are seen for all tissues. A time course study published in FlyBase (<http://flybase.org/reports/FBgn0037646.html>, under Microarray data, Personal communications to FlyBase, Gaurav et al.

2008), shows upregulated expression of DmBCA in early larval stages and late in metamorphosis, but downregulated levels in embryonal, late larval, early metamorphosis, and adult stages. The specific temporal and tissue patterns of expression suggest relevant functional and/or physiological roles for DmBCA.

## Conclusions

In conclusion, fruit fly (*Drosophila melanogaster*)  $\beta$ -CA (DmBCA) is an active mitochondrial enzyme for the physiological reaction catalyzed by CAs, the hydration of CO<sub>2</sub> to bicarbonate and protons. It is inhibited by various inorganic anions, boronic/arsonic acids or sulfonamides. Mammals do not possess  $\beta$ -class CAs, but these enzymes are widespread throughout the phylogenetic tree, making them exciting new targets for parasitic drug development. Indeed,  $\beta$ -CAs are found in many pathogenic organisms and pathogen vectors of the animal kingdom, including the malaria mosquito *Anopheles*, the yellow fever mosquito *Aedes*, the filariasis vector *Culex*, *Ancylostoma* hookworms, *Brugia* filarial nematodes, the blood fluke *Schistosoma mansoni* and the liver fluke *Schistosoma sinensis*. Because animal-derived  $\beta$ -CAs probably have a different predicted structure compared to the  $\beta$ -CAs investigated so far in detail from Archaea, plants, algae and fungi, as well as the  $\alpha$ -CAs, specific inhibitors against these enzymes could be designed with minimal effect on human CAs and normal bacterial flora.

## Methods

### Bioinformatic analysis

Taking advantage of the known pea  $\beta$ -CA (UniProt P17067) as an initial query, and subsequently the found invertebrate  $\beta$ -CAs (*D. melanogaster*, AAF54311; *C. elegans* CAJ43916), the animal CA sequences were retrieved from NCBI protein databases using Blast [43] <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and from complete genomes at the UCSC Genome Bioinformatics Site <http://genome.ucsc.edu> using BLAT search algorithms [44]. Hits were taken through iterated cycles of multiple sequence alignment (ClustalW; [45]), evaluation and revision. For revision, sequences with poorly matching or missing regions were subjected to gene model generation with GeneWise ([46]; <http://www.ebi.ac.uk/Tools/Wise2/>), taking the genomic sequences from the UCSC site. EST and mRNA sequence data from NCBI were used to confirm gene models, sometimes to bridge gaps or fill ends in the genomic sequences, and to discover and assemble  $\beta$ -CAs from less than genome-wide sequenced organisms.

Phylogenetic trees were constructed from the multiple sequence alignments using MEGA 4 [47]. Preliminary Neighbour-Joining trees [48] were constructed with all



sequences, and a representative set of 26 sequences was then selected for a final tree, eliminating excessive insect and nematode sequences and severely incomplete sequences. The final tree was inferred using the Minimum evolution method [49] from a multiple sequence alignment containing columns corresponding to positions 31 to 255 (of 255) of *DmBCA*. The bootstrap consensus tree inferred from 1,000 replicates [50] was taken to represent the evolutionary history of the analyzed sequences. The tree was drawn to scale, with branch lengths proportional to the evolutionary distances used to infer the phylogenetic tree and rooted using the Cnidarian sequences as outgroup. The evolutionary distances were computed using the Poisson correction method and are presented in the units of the number of amino acid substitutions per site. The minimum evolution tree was searched using the Close-Neighbor-Interchange algorithm [51] at a search level of 3. The Neighbor-Joining algorithm [48] was used to generate the initial tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option).

Predictions of subcellular localization were made using TargetP v1.1 ([52]; <http://www.cbs.dtu.dk/services/TargetP/>), Secretome 2.0 <http://www.cbs.dtu.dk/services/SecretomeP/>, Wolf PSORT, <http://wolfsort.org/> and MitoProt II v. 1.101 <http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html>. Conserved domain types were identified at the NCBI Conserved Domains Database <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>.

### Construction of recombinant baculoviruses

Two constructs were engineered to study functional aspects of *D. melanogaster*  $\beta$ -CA (*DmBCA*). The first one contained a *GFP* (Green Fluorescent Protein) sequence fused C-terminally to the *DmBCA* cDNA for detection of the recombinant protein with confocal microscopy. The second construct contained a C-terminal histidine tag for protein purification. Both constructs contained full length  $\beta$ -CA gene, including the proposed N-terminal signal sequence. *Bgl*II and *Xho*I restriction sites and a thrombin cleavage site (for tag removal) were introduced into both constructs.

Total RNA extracted from *D. melanogaster* S2 cells (TRIzol® reagent and protocol, Invitrogen) was precipitated using sodium acetate at a final concentration of 100 mM and 100% ethanol. The solution was centrifuged at 13,000×g for 15 min at +4°C. The RNA sample was washed once with 70% ethanol and recentrifuged in the same manner. The sample was evaporated at room temperature and then suspended in sterile water. Precipitated RNA was transcribed into cDNA using a First strand cDNA Synthesis Kit #K1612 (Fermentas) according to the manufacturer's instructions.

The *DmBCA* gene was identified and amplified from cDNA by PCR using Phusion™ Hot Start High Fidelity DNA Polymerase (Finnzymes, Espoo, Finland). Sequence-specific primers were ordered from Sigma-Aldrich (St. Louis, MO). The forward primer was 5'-ATGGAGCGTATTTTGAGGGGAATC-3' (F1), and the reverse primer was 3'-CTACGAATAGAATCTTCTGACCTC-5' (R1). PCR was performed in a PTC 2000 thermal cycler (MJ Research, Waltham, MA), and the program consisted of a single 98°C denaturation step for 30 s, followed by 33 cycles of denaturation at 98°C for 10 s, annealing at 53°C for 15 s and extension at 72°C for 25 s. A final extension was performed at 72°C for 5 min. The PCR product band was separated from the gel and dissolved using Illustra™ GFX PCR DNA and GEL Band Purification Kit (GE Healthcare Life Sciences, Buckinghamshire, UK).

To engineer the *DmBCA-GFP* construct, the sequences of *DmBCA* and *GFP* were first amplified separately using sequence specific primers. The templates used for *DmBCA* and *GFP* were *DmBCA* gene product obtained from cDNA and pEGFP-N1™ Vector (Clontech), respectively. The constructs were then combined using PCR reactions whose R2 and F3 primer sequences partly overlapped (bolded in the primer sequences, see below). This allowed the ends of the PCR products to recognize each other. The forward primer used for *DmBCA* amplification was 5'-GGCCAGATC-TATGGAGCGTATTTTGAGGGGA-3' (F2), and the reverse primer was 5'-CACGGAACCACGGGGCAC-CAGCGAATAGAATCTTCTGACCTC-3' (R2). The bolded sequence was designed to recognize the thrombin site and part of the *GFP* PCR product, while the other half recognized the *DmBCA* PCR product.

The forward primer used for *GFP* amplification was 5'-TCGCTGGTGCCCCGTTGGTTCCGTGAGCAAGGCGAGGAGCTG-3' (F3), and the reverse primer was 5'-CCGCTCGAGTTACTTGTACAGCTCGTCCAT-3' (R3). The bolded sequence in the F3 primer was designed to recognize the thrombin site and part of the *DmBCA* PCR product, while the other half recognized the *GFP* PCR product. The PCR program was as follows: 98°C for 2 min; then 33 cycles of 98°C for 10 s, 55°C for 15 s, and 72°C for 30 s; and finally 72°C for 7 min.

Both PCR products were run on agarose gel, purified and used as templates in the next PCR reaction. The PCR program was as follows: 98°C for 2 min; then 33 cycles of 98°C for 10 s, 55°C for 15 s and 72°C for 40 s; and finally 72°C for 8 min. The forward primer used in this PCR was (F2), and the reverse primer was (R3). The extra sequence overlaps at the 3' end of *DmBCA* and at the 5' end of the *GFP* allowed these PCR products to anneal to each other.

The PCR product was run on an agarose gel, and the obtained band was purified. pFastBac™ 1 plasmid (Invitrogen) and the PCR product were digested at +37°C overnight with *Bam*HI and *Xho*I restriction enzymes (New England Biolabs). The digested plasmid and *DmBCA-GFP* construct were purified and then ligated overnight at +4°C using T4 DNA ligase (New England Biolabs). The ligated product was transformed into TOP10 bacteria (Invitrogen). Overnight cultures (8 ml) were made from these colonies, and plasmids were purified using a QIAprep Spin Miniprep Kit™ (Qiagen, Hilden, Germany). Sequencing was performed to verify the validity of the *DmBCA-GFP* construct. The construction of baculoviral genomes encoding the recombinant proteins has been described previously [53].

For recombinant protein production, the *DmBCA* gene construct with a C-terminal polyhistidine tag of six histidines was constructed and cloned into the pFast-Bac1™ vector. The forward primer used in the initial amplification of the *DmBCA* gene was (F2), and the reverse primer was 5'-GCCCTCGAGTTAATGGTGGTGATGGTGGTGGGAACCACGGGGCACCAGC-GAATAGAATCTTCTGACCTC -3' (R4). The latter primer contains nucleotide repeats to create the polyhistidine tag (bolded in the primer sequence). The PCR program was as follows: 98°C for 60 s; then 35 cycles of 98°C for 10 s, 66°C for 15 s, and 72°C for 60 s; and finally 72°C for 5 min. Otherwise, the construct was made essentially in the same way as the *DmBCA-GFP* construct described above.

#### Study of the subcellular localization of DmBCA

A total of 100 µl of Sf9 insect cells (2 million cells/ml) were infected with 10 µl of baculovirus stock. The cells were kept at +27°C in incubator for three days in Lab-Tek™ Chamber Slide™ System™ plates (Nunc). The medium was then removed, and the cells were incubated in 600 µl of medium containing 100 nM MitoTracker Red CMXros™ (Invitrogen) for 20 minutes at +27°C. The cells were washed three times with 600 µl of medium and kept at +27°C in an incubator for two hours. The cells were then washed with PBS, fixed with 4% paraformaldehyde for five minutes and washed again with PBS. The cells were mounted in Vecta-Shield® Mounting Medium (Vector Laboratories), covered with cover slips and analyzed using a confocal scanning laser microscope (Perkin Elmer-Cetus/Wallac UltraView LCI system™) with two different wavelengths: 488 nm for GFP detection and 579 nm for MitoTracker™. Image acquisition was performed with an Andor iXon™ DV885 EMCCD camera and the Andor iQ™ software (Andor).

#### Production and Purification of Recombinant Insect β-CA

The Sf9 insect cells were grown in HyQ SFX-Insect serum-free cell culture medium (HyClone, Logan, UT) in an orbital shaker at 27°C (125 rpm) for three days after infection. Although much of the DmBCA was associated with the cell pellet, protein purification was performed after centrifugation (5000 × g, 20°C, 8 min) from the supernatant and yielded highly pure DmBCA protein for characterization. Purification was performed using the ProBond™ Purification System (Invitrogen) under native binding conditions, with wash and elution buffers made according to the manufacturer's instructions. The purification procedure per 500 ml of insect cell medium was as follows: 1 liter of native binding buffer and 25 ml of the nickel-chelating resin were added to the medium, and the His-tagged protein was then allowed to bind to the resin on a magnetic stirrer at 25°C for 3 h. The resin was washed with 100 + 30 ml of washing buffer (Invitrogen). The protein was then eluted with elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 250 mM imidazole, pH 8.0).

The purified DmBCA recombinant protein was transferred to a buffer of 50 mM Tris-HCl, pH 7.5, using an Amicon Ultracel™ - 10 k centrifugal filter device (Millipore) according to the manufacturer's instructions. To remove the His tag, the recombinant protein was treated with 60 µl of resin-coupled thrombin (Thrombin Clean-Cleave KIT™, Sigma) per 1 mg of protein with gentle shaking at 25°C for 1 h, according to the manufacturer's instructions. Protein concentration was determined using the DC Protein Assay™ (Bio-Rad) with three different dilutions. Purified recombinant DmBCA proteins were analyzed using 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions. The gels were stained using the Colloidal Blue Staining Kit™ (Invitrogen).

#### CA activity measurements

An Applied Photophysics stopped-flow instrument was used to assay the CA-catalyzed CO<sub>2</sub> hydration activity. Phenol red (at a concentration of 0.2 mM) was used as an indicator, working at the absorbance maximum of 557 nm, with 10 - 20 mM HEPES (pH 7.5) or Tris(pH 8.3) as buffers and 20 mM Na<sub>2</sub>SO<sub>4</sub> or 20 mM NaClO<sub>4</sub> (for maintaining constant ionic strength), following the initial rates of the CA-catalyzed CO<sub>2</sub> hydration reaction for a period of 10 - 100 s. The CO<sub>2</sub> concentrations ranged from 1.7 to 17 mM for the determination of kinetic parameters and inhibition constants. For each inhibitor at least six traces of the initial 5-10% of the reaction were used to determine the initial velocity. The uncatalyzed rates were determined in the same manner and subtracted from the total observed rates. Stock solutions of inhibitor (100 mM)

were prepared in distilled-deionized water, and dilutions up to 0.01  $\mu\text{M}$  were made thereafter with distilled-deionized water. Inhibitor and enzyme solutions were pre-incubated together for 15 min at room temperature prior to the assay to allow for the formation of the E-I complex. The inhibition constants were obtained by non-linear least-squares methods using PRISM 3, whereas the kinetic parameters for the uninhibited enzymes were obtained from Lineweaver-Burk plots, each representing the mean of at least three different determinations.

Kinetic measurements have been performed also with m-cresol purple (as indicator) - bicine (as buffer) (data not shown), and the results were the same ( $\pm 5$ -10% of the reported values, which is the error range of this method) both for the kinetic parameters of  $\text{CO}_2$  hydration and for the inhibition constants of anionic inhibitors investigated here. Thus, the standard method reported in this paper is reliable for the investigation of  $\beta$ -CAs (in addition to the  $\alpha$ -CAs) as reported by this group for several enzymes, such as the three  $\beta$ -class enzymes from *Mycobacterium tuberculosis*, *Helicobacter pylori* and *Brucella suis*, and the fungal class enzymes (for example [25]).

#### Dynamic light scattering analysis

The hydrodynamic diameters of the proteins were determined by dynamic light scattering (DLS) using Zetasizer ZS (Malvern Instruments Ltd., Worcestershire, United Kingdom). A 100  $\mu\text{l}$  sample of DmBCA (200  $\mu\text{g}/\text{ml}$ ) in elution buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , 500 mM NaCl, 250 mM imidazole, pH 8.0) was analyzed. Human CA II (HCA II) (180  $\mu\text{g}/\text{ml}$ ) in 0.1 M Tris, 0.4 M  $\text{NaN}_3$ , 1 mM benzamidine, 20% glycerol, pH 7.0 was analyzed to support the results of DmBCA analysis. The small molecules such as imidazole and glycerol of the elution buffers seemed to dominate in the DLS analysis, since most of the light scattering was from particles having diameter smaller than one nm. Therefore, proteins were exchanged to 50 mM  $\text{Na}_2\text{HPO}_4$  pH 7.0 containing 100 mM NaCl using protein desalting spin columns (Pierce). DLS analysis was then performed for 100  $\mu\text{l}$  sample by using temperature scanning mode where the temperature was raised from 4°C to 50°C at 3°C intervals. Sample was let to equilibrate to each measurement temperature two minutes before data acquisition. For DmBCA three parallel measurements were carried out at each temperature and for HCA II only one measurement was performed at each temperature. The molecular weight of the protein was estimated from hydrodynamic diameter using globular protein standard curve provided by the manufacturer.

#### Analytical gel filtration

The molecular size of the protein in solution was measured by size-exclusion chromatography using

Superdex200 10/300GL column (GE Healthcare) connected to ÄKTA™ purifier-100 equipped with UV-900 monitor (GE Healthcare). The analysis was done using 50 mM  $\text{Na}_2\text{HPO}_4$ , 650 mM NaCl (pH 7.0) as mobile phase. 20-30  $\mu\text{g}$  protein was injected per run. All the analyses were done with flow rate 0.3 ml/min, and absorbencies at 280 nm and 205 nm were used to locate the protein peaks in the chromatograms. Molecular weight calibration curve was prepared by analyzing gel filtration standard protein mixture containing thyroglobulin (670kDa),  $\gamma$ -globulin (158kDa), ovalbumin (44kDa), myoglobin (17kDa) and vitamin B<sub>12</sub> (1,35kDa) (Bio-Rad).

#### Additional material

**Additional file 1: The full alignment of the 26 invertebrate  $\beta$ -CA sequences (partly shown in Figure 1).**

**Additional file 2: The full sequence alignment of all of the identified invertebrate  $\beta$ -CA sequences.**

#### Acknowledgements

We thank Professor Hans Spelbrink and Dr. Peter Martinsson for help with fluorescence confocal microscopy and Aulikki Lehmus and Jukka Lehtonen for skilful technical assistance. We thank Joanna Zmurko for assistance in gel filtration analysis. This research was financed in part by grants from the 6<sup>th</sup> Framework Programme of the European Union (DeZnIT project) and the Competitive Research Funding of the Tampere University Hospital (9L071).

#### Author details

<sup>1</sup>Institute of Medical Technology, University of Tampere and Tampere University Hospital, Tampere, Finland. <sup>2</sup>School of Medicine, University of Tampere and Tampere University Hospital, Tampere, Finland. <sup>3</sup>Centre for Laboratory Medicine, Tampere University Hospital, Tampere, Finland. <sup>4</sup>Laboratorio di Chimica Bioinorganica, Università degli Studi di Firenze, Sesto Fiorentino (Firenze), Italy. <sup>5</sup>Vetsuisse Faculty and Zurich Center for Integrative Human Physiology (ZIHP), Institute of Veterinary Physiology, University of Zurich, Zurich, Switzerland.

#### Authors' contributions

LS carried out protein construct design, protein production and purification, colocalization studies and drafted the manuscript. MH contributed to protein construct design and co-localization studies. MT made all the bioinformatic analysis and helped to draft the manuscript. AO contributed in bioinformatic analyses. CTS, AI and AS carried out the kinetic measurements. JL and VPH carried out the DLS analysis. BN and VPH performed the gel filtration studies. TAG conceived the study, provided materials and helped to draft the manuscript. CTS and SP conceived the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Received: 22 February 2010 Accepted: 26 July 2010

Published: 26 July 2010

#### References

1. Sly WS, Hu PY: Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu Rev Biochem* 1995, **64**:375-401.
2. Lane TW, Saito MA, George GN, Pickering IJ, Prince RC, Morel FM: Biochemistry: a cadmium enzyme from a marine diatom. *Nature* 2005, **435**(7038):42.
3. Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FM: Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* 2008, **452**(7183):56-61.

4. Tripp BC, Bell CB, Cruz F, Krebs C, Ferry JG: **A role for iron in an ancient carbonic anhydrase.** *J Biol Chem* 2004, **279**(8):6683-6687.
5. Macaulay SR, Zimmerman SA, Apolinario EE, Evilia C, Hou YM, Ferry JG, Sowers KR: **The archetype gamma-class carbonic anhydrase (Cam) contains iron when synthesized in vivo.** *Biochemistry* 2009, **48**(5):817-819.
6. Zimmerman SA, Ferry JG: **The beta and gamma classes of carbonic anhydrase.** *Curr Pharm Des* 2008, **14**(7):716-721.
7. So AK, Espie GS, Williams EB, Shively JM, Heinhorst S, Cannon GC: **A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell.** *J Bacteriol* 2004, **186**(3):623-630.
8. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, Yeates TO, Kerfeld CA: **The structure of beta-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two.** *J Biol Chem* 2006, **281**(11):7546-7555.
9. Strop P, Smith KS, Iverson TM, Ferry JG, Rees DC: **Crystal structure of the "cab"-type beta class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*.** *J Biol Chem* 2001, **276**(13):10299-10305.
10. Kimber MS, Pai EF: **The active site architecture of *Pisum sativum* beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases.** *EMBO J* 2000, **19**(7):1407-1418.
11. Parisi G, Perales M, Fornasari MS, Colaneri A, Gonzalez-Schain N, Gomez-Casati D, Zimmermann S, Brennicke A, Araya A, Ferry JG, Echave J, Zabaleta E: **Gamma carbonic anhydrases in plant mitochondria.** *Plant Mol Biol* 2004, **55**(2):193-207.
12. Elleuche S, Poggeler S: **Evolution of carbonic anhydrases in fungi.** *Curr Genet* 2009, **55**(2):211-222.
13. Fabre N, Reiter IM, Becuwe-Linka N, Genty B, Rumeau D: **Characterization and expression analysis of genes encoding alpha and beta carbonic anhydrases in *Arabidopsis*.** *Plant Cell Environ* 2007, **30**(5):617-629.
14. Hewett-Emmett D: **Evolution and distribution of the carbonic anhydrase gene families.** *EXS* 2000, **90**(90):29-76.
15. Klengel T, Liang WJ, Chaloupka J, Ruoff C, Schroppel K, Naglik JR, Eckert SE, Mogensen EG, Haynes K, Tuite MF, Levin LR, Buck J, Muhlschlegel FA: **Fungal adenyllyl cyclase integrates CO2 sensing with cAMP signaling and virulence.** *Curr Biol* 2005, **15**(22):2021-2026.
16. Bahn YS, Cox GM, Perfect JR, Heitman J: **Carbonic anhydrase and CO2 sensing during *Cryptococcus neoformans* growth, differentiation, and virulence.** *Curr Biol* 2005, **15**(22):2013-2020.
17. Innocenti A, Muhlschlegel FA, Hall RA, Steegborn C, Scozzafava A, Supuran CT: **Carbonic anhydrase inhibitors: inhibition of the beta-class enzymes from the fungal pathogens *Candida albicans* and *Cryptococcus neoformans* with simple anions.** *Bioorg Med Chem Lett* 2008, **18**(18):5066-5070.
18. Isik S, Kockar F, Arslan O, Guler OO, Innocenti A, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the beta-class enzyme from the yeast *Saccharomyces cerevisiae* with anions.** *Bioorg Med Chem Lett* 2008, **18**(24):6327-6331.
19. Zimmerman SA, Ferry JG, Supuran CT: **Inhibition of the archaeal beta-class (Cab) and gamma-class (Cam) carbonic anhydrases.** *Curr Top Med Chem* 2007, **7**(9):901-908.
20. Isik S, Kockar F, Aydin M, Arslan O, Guler OO, Innocenti A, Scozzafava A, Supuran CT: **Carbonic anhydrase inhibitors: inhibition of the beta-class enzyme from the yeast *Saccharomyces cerevisiae* with sulfonamides and sulfamates.** *Bioorg Med Chem* 2009, **17**(3):1158-1163.
21. Nishimori I, Minakuchi T, Kohsaki T, Onishi S, Takeuchi H, Vullo D, Scozzafava A, Supuran CT: **Carbonic anhydrase inhibitors: the beta-carbonic anhydrase from *Helicobacter pylori* is a new target for sulfonamide and sulfamate inhibitors.** *Bioorg Med Chem Lett* 2007, **17**(13):3585-3594.
22. Minakuchi T, Nishimori I, Vullo D, Scozzafava A, Supuran CT: **Molecular cloning, characterization, and inhibition studies of the Rv1284 beta-carbonic anhydrase from *Mycobacterium tuberculosis* with sulfonamides and a sulfamate.** *J Med Chem* 2009, **52**(8):2226-2232.
23. Nishimori I, Minakuchi T, Vullo D, Scozzafava A, Innocenti A, Supuran CT: **Carbonic anhydrase inhibitors. Cloning, characterization, and inhibition studies of a new beta-carbonic anhydrase from *Mycobacterium tuberculosis*.** *J Med Chem* 2009, **52**(9):3116-3120.
24. Cronk JD, Rowlett RS, Zhang KY, Tu C, Endrizzi JA, Lee J, Gareiss PC, Preiss JR: **Identification of a novel noncatalytic bicarbonate binding site in eubacterial beta-carbonic anhydrase.** *Biochemistry* 2006, **45**(14):4351-4361.
25. Joseph P, Turtaut F, Ouahrani-Bettache S, Montero JL, Nishimori I, Minakuchi T, Vullo D, Scozzafava A, Kohler S, Winum JY, Supuran CT: **Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from *Brucella suis*.** *J Med Chem* 2010, **53**(5):2277-2285.
26. Innocenti A, Hall RA, Schlicker C, Muhlschlegel FA, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the beta-class enzymes from the fungal pathogens *Candida albicans* and *Cryptococcus neoformans* with aliphatic and aromatic carboxylates.** *Bioorg Med Chem* 2009, **17**(7):2654-2657.
27. Innocenti A, Hall RA, Schlicker C, Scozzafava A, Steegborn C, Muhlschlegel FA, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition and homology modeling studies of the fungal beta-carbonic anhydrase from *Candida albicans* with sulfonamides.** *Bioorg Med Chem* 2009, **17**(13):4503-4509.
28. Innocenti A, Winum JY, Hall RA, Muhlschlegel FA, Scozzafava A, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the fungal beta-carbonic anhydrases from *Candida albicans* and *Cryptococcus neoformans* with boronic acids.** *Bioorg Med Chem Lett* 2009, **19**(10):2642-2645.
29. Schlicker C, Hall RA, Vullo D, Middelhaufe S, Gertz M, Supuran CT, Muhlschlegel FA, Steegborn C: **Structure and inhibition of the CO2-sensing carbonic anhydrase Can2 from the pathogenic fungus *Cryptococcus neoformans*.** *J Mol Biol* 2009, **385**(4):1207-1220.
30. Innocenti A, Leewattanapasuk W, Muhlschlegel FA, Mastrolorenzo A, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the beta-class enzyme from the pathogenic yeast *Candida glabrata* with anions.** *Bioorg Med Chem Lett* 2009, **19**(16):4802-4805.
31. Smith KS, Cosper NJ, Stalhandske C, Scott RA, Ferry JG: **Structural and kinetic characterization of an archaeal beta-class carbonic anhydrase.** *J Bacteriol* 2000, **182**(23):6605-6613.
32. Cox EH, McLendon GL, Morel FM, Lane TW, Prince RC, Pickering IJ, George GN: **The active site structure of *Thalassiosira weissflogii* carbonic anhydrase 1.** *Biochemistry* 2000, **39**(40):12128-12130.
33. Smith KS, Ingram-Smith C, Ferry JG: **Roles of the conserved aspartate and arginine in the catalytic mechanism of an archaeal beta-class carbonic anhydrase.** *J Bacteriol* 2002, **184**(15):4240-4245.
34. Tetu SG, Tanz SK, Vella N, Burnell JN, Ludwig M: **The *Flaveria bidentis* beta-carbonic anhydrase gene family encodes cytosolic and chloroplastic isoforms demonstrating distinct organ-specific expression patterns.** *Plant Physiol* 2007, **144**(3):1316-1327.
35. Mitra M, Lato SM, Ynalvez RA, Xiao Y, Moroney JV: **Identification of a new chloroplast carbonic anhydrase in *Chlamydomonas reinhardtii*.** *Plant Physiol* 2004, **135**(1):173-182.
36. Bahn YS, Muhlschlegel FA: **CO2 sensing in fungi and beyond.** *Curr Opin Microbiol* 2006, **9**(6):572-578.
37. Fasseas MK, Tsikou D, Fliemetakis E, Katinakis P: **Molecular and biochemical analysis of the beta class carbonic anhydrases in *Caenorhabditis elegans*.** *Mol Biol Rep* 2010, **37**(6):2941-50.
38. Supuran CT: **Carbonic anhydrases: novel therapeutic applications for inhibitors and activators.** *Nat Rev Drug Discov* 2008, **7**(2):168-181.
39. Rowlett RS, Tu C, Lee J, Herman AG, Chapnick DA, Shah SH, Gareiss PC: **Allosteric site variants of *Haemophilus influenzae* beta-carbonic anhydrase.** *Biochemistry* 2009, **48**(26):6146-6156.
40. Rowlett RS, Hoffmann KM, Failing H, Mysliwiec MM, Samardzic D: **Evidence for a bicarbonate "escort" site in *Haemophilus influenzae* beta-carbonic anhydrase.** *Biochemistry* 2010, **49**(17):3640-3647.
41. Hazen SA, Waheed A, Sly WS, LaNoue KF, Lynch CJ: **Differentiation-dependent expression of CA V and the role of carbonic anhydrase isozymes in pyruvate carboxylation in adipocytes.** *FASEB J* 1996, **10**(4):481-490.
42. Chintapalli VR, Wang J, Dow JA: **Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease.** *Nat Genet* 2007, **39**(6):715-720.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
44. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**(4):656-664.
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD,

- Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947-2948.
46. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Res* 2004, **14**(5):988-995.
  47. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596-1599.
  48. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**(4):406-425.
  49. Zhakhetsky A, Nei M: **METREE: a program package for inferring and testing minimum-evolution trees.** *Comput Appl Biosci* 1994, **10**(4):409-412.
  50. Zharkikh A, Li WH: **Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique.** *Mol Phylogenet Evol* 1995, **4**(1):44-63.
  51. Nei M: **Phylogenetic analysis in molecular evolutionary genetics.** *Annu Rev Genet* 1996, **30**:371-403.
  52. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**(4):1005-1016.
  53. Hilvo M, Baranauskiene L, Salzano AM, Scaloni A, Matulis D, Innocenti A, Scozzafava A, Monti SM, Di Fiore A, De Simone G, Lindfors M, Janis J, Valjakka J, Pastorekova S, Pastorek J, Kulomaa MS, Nordlund HR, Supuran CT, Parkkila S: **Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes.** *J Biol Chem* 2008, **283**(41):27799-27809.

doi:10.1186/1471-2091-11-28

**Cite this article as:** Syrjänen et al.: Characterization of the first beta-class carbonic anhydrase from an arthropod (*Drosophila melanogaster*) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates. *BMC Biochemistry* 2010 **11**:28.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

