The carbonic anhydrase of Clostridium autoethanogenum represents a new subclass of β-carbonic anhydrases

Bart Pander 1 · Gemma Harris 2 · David J. Scott 2,3,4 · Klaus Winzer 1 · Michael Köpke 5 · Sean D. Simpson 5 · Nigel P. Minton 1 · Anne M. Henstra 1

Received: 14 May 2019 / Revised: 5 July 2019 / Accepted: 8 July 2019 / Published online: 25 July 2019 © The Author(s) 2019

Abstract
Carbonic anhydrase catalyses the interconversion of carbon dioxide and water to bicarbonate and protons. It was unknown if the industrial-relevant acetogen Clostridium autoethanogenum possesses these enzymes. We identified two putative carbonic anhydrase genes in its genome, one of the β class and one of the γ class. Carbonic anhydrase activity was found for the purified β class enzyme, but not the γ class candidate. Functional complementation of an Escherichia coli carbonic anhydrase knock-out mutant showed that the β class carbonic anhydrase could complement this activity, but not the γ class candidate gene. Phylogenetic analysis showed that the β class carbonic anhydrase of Clostridium autoethanogenum represents a novel subclass of β class carbonic anhydrases that form the F-clade. The members of this clade have the shortest primary structure of any known carbonic anhydrase.

Keywords Carbonic anhydrase · Clostridium autoethanogenum · Gas fermentation · Carbon dioxide · Carbon monoxide · Enzyme characterisation

Introduction
Clostridium autoethanogenum fixes carbon dioxide through the Wood–Ljungdahl pathway (WLP) and produces acetate, ethanol and 2,3-butanediol natively (Abrini et al. 1994; Köpke et al. 2011). As such, use of C. autoethanogenum provides an attractive way to mitigate the effects of global CO₂ release. Besides capture in the WLP, CO₂ is also fixed at other metabolic steps, and it was shown, for instance, that elevated CO₂ partial pressures benefit the production of 2,3-butanediol (Simpson et al. 2014). In fact, many reactions in microbial metabolism exist where CO₂ or bicarbonate are substrates or products (Smith and Ferry 2000). It was proposed that without a mechanism for the rapid interconversion of carbon dioxide and bicarbonate, the turnover rates of common carboxylation reactions that consume bicarbonate would not be feasible in Escherichia coli (Merlin et al. 2003). The interconversion of CO₂ and bicarbonate (Eq. (1)) is catalysed by carbonic anhydrase (CA), an enzyme that is essential for most forms of life. To optimise product formation and carbon fixation, knowledge about CA activity is important (Hawkins et al. 2013; Lian et al. 2016).

H₂O + CO₂↔HCO₃⁻ + H⁺ (1)

There are three major classes of carbonic anhydrase, which are as follows: α-CA, β-CA and γ-CA. The α-class protein is active in monomeric state, while the β-CAs form dimers, which in turn can stack into tetramers or octamers. The γ-CAs form homotrimeric structures (Lindskog 1997; Ferry...
Carbonic anhydrases (EC 4.2.1.1) are metalloenzymes. All known CAs function with a Zn\(^{2+}\) metal ion cofactor in their active site, but some are functional with other metal ions as well. In γ-CAs of anaerobic microorganisms, Fe\(^{2+}\) or Co\(^{2+}\) can functionally replace Zn\(^{2+}\), while the ζ-class CA is also functional with Cd\(^{2+}\) (Kumar and Ferry 2014; Supuran and Ferry 2016). The metal ions are coordinated in the active site by three histidines (for α-, γ-, δ- and η-class CAs) or one histidine and two cysteines (for β- and ζ-class CAs). These amino acids are arranged in a distinct motif for each CA class (Hewett-Emmett and Tashian 1996; Ferry 2010; Rowlett 2010; Capasso and Supuran 2015). The diverse class of microbial β-CA is divided in five clades, A–D (Smith and Ferry 2000) plus the former ε-CA class that we will recognise here as the E clade. All β-CA sport the common metal-coordinating active-site motif CxDxR-G-HxxC (Lindskog 1997; Ferry 2010; Rowlett 2010).

The physiological role of CA enzymes is diverse, but not always clear. In animals, they are essential for rapid gas exchange and other functions, such as pH homeostasis. Since specific CAs act in specific tissues with specific inhibition patterns, CAs are common drug targets (Supuran and Scozzafava 2007). In plants and algae, CAs have a function in carbon dioxide diffusion facilitation and carbon concentration mechanisms (Moroney et al. 2001; Moroney et al. 2011). In prokaryotes, CAs are almost ubiquitous. Proposed functions of CAs in prokaryotes are to act in carbon concentration (Cannon et al. 2010), carbon dioxide transport (Gai et al. 2014), facilitation of carbon dioxide or bicarbonate-consuming or -producing reactions (Smith and Ferry 2000; Merlin et al. 2003; Supuran and Ferry 2013), pH homeostasis (Sachs et al. 2005) and acetate transport facilitation (Braus-Stromeyer et al. 1997). Most bacteria can grow under low carbon dioxide partial pressures. For this, a CA gene seems to be essential since species that need high carbon dioxide partial pressures (capnophiles) often have no detectable CA activity and some have lost CA genes (Ueda et al. 2008; Ueda et al. 2012). For the capnophile Campylobacter jejuni, it was shown that it contains a CA that is only active at high pH, but not under normal physiological pH (Al-Haideri et al. 2016). Also, CA deletion mutants often can only grow under high carbon dioxide partial pressures (Kusian et al. 2002; Merlin et al. 2003; Kumar et al. 2013), making them functional capnophiles. In E. coli, a specific CA, CynT, is part of cyanate metabolism operon that is tightly controlled and induced by cyanate or azide (Guilloton et al. 1993). When the constitutively expressed CA can gene of E. coli was disrupted, the mutant strain (E. coli EDCM636) was unable to grow under atmospheric carbon dioxide pressure. Addition of azide restored its normal growth (Merlin et al. 2003). This strain is useful for studies to test if putative CA genes can complement CA activity.

Production of 3-hydroxypropionate by the expression of bicarbonate-dependent reactions in Pyrococcus furiosus that lacks a functional CA benefited greatly from the expression of functional CA genes (Lian et al. 2016) underpinning the importance of CAs in a biotechnological context.

Previously, a diverse set of acetogens, i.e., bacteria using the WLP, screened for CA activity showed a range of CA activities (Braus-Stromeyer et al. 1997). One model of acetogen, Acetobacterium woodii, had high CA activity, but the other model of acetogen, Moorella thermoacetica, showed little to no activity. No close relatives of C. autoethanogenum were included in this CA screening.

Here, we identified two putative CA genes in the genome of C. autoethanogenum. One of these encoded a dimeric enzyme that was indeed active as CA. The gene formed a new F clade in the β-class of CA. This new clade represents CAs with the shortest primary structure known.

**Materials and methods**

**Bioinformatics**

To search the online protein databases, we used NCBI blastp, PSI-Blast, Delta-Blast, PHI-Blast and tblastn algorithms (Sayers et al. 2012). For alignments, we used the MUSCLE algorithm (Edgar 2004) in both Ugene (Okonechnikov et al. 2012) and Mega7.0 (Kumar et al. 2016). Consensus logos were made using JalView (Waterhouse et al. 2009). Phylogenetic analysis was performed in MrBayes (Ronquist 2010) and Mega7.0 (Kumar et al. 2016). Consensus logos were made using MrBayes (Ronquist et al. 2012) and Mega7.0, and figures were made in Figtree (Rambaut 2014). To find the optimal evolutionary model, we used ProTest (Darriba et al. 2011) and Mega7.0. Structural protein modelling was done by Phyre2 (Kelley et al. 2015). Protparam of the ExPASy server (Gasteiger et al. 2005) was used to calculate predicted parameters of proteins.

**Strains and growth conditions**

*C. autoethanogenum* JA1-1 (DSM 10061) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Germany). *E. coli* EDCM636 was from the Coli Genetic Stock Centre (CGSC, New Haven, CT, USA). *C. autoethanogenum* was grown in YTF medium (yeast extract 10 g l\(^{-1}\), tryptone 16 g l\(^{-1}\), fructose 10 g l\(^{-1}\),
Enzymes used for subcloning and PCR were purchased from New England Biolabs (NEB, MA, USA). All other chemicals were purchased from Sigma-Aldrich except for Bugbuster (Merck Millipore, Germany) and Strep-Tactin sepharose (IBA, Germany).

Plasmids for heterologous overexpression of CA genes were constructed as follows: Genes were amplified by polymerase chain reaction (PCR) using genomic DNA of C. autoethanogenum as template. The genomic DNA was isolated using the Genelute (Sigma) genomic DNA isolation kit. PCR was performed with primers (Table 1) FbCA and RbCA to obtain an untagged version of the gCA gene. These PCR fragments were formed with these plasmids. Plasmids were isolated using the Genelute (Sigma) genomic DNA isolation kit. PCR was performed with primers FbCA and RbCAstrep to obtain a C-terminal strep-tagged version of the gCA gene. These PCR fragments were subcloned into pMTL82252 (Heap et al. 2009) using Ndel, EcoRI and T4-ligase. To construct His-tagged versions of Caut-bCA and Caut-gCA, we performed PCR with primers BCAECPF and BCAECP1 or FPCA1 and RPCA1, respectively. The resulting fragments were cloned into a pET16b (Novagen) plasmid using Ndel and BamHI and T4-ligase. E. coli DH5x chemically competent cells (NEB) were transformed with these plasmids. Plasmids were isolated using Monarch plasmid isolation kit (NEB) and Sanger sequenced by Eurofins using the FpMTL8xx5x primer for the pMTL82252 plasmids and the pET16b_F primer for the pET16b plasmids.

E. coli BL21(DE3) pLysS was transformed with pET16b-bCA and pET16g-gCA. E. coli EDCM636 was transformed with pMTL82252-bCA, pMTL82252-bCAstrepC and pMTL82252-gCA. The transformed cells were plated on LB agar with Em and on LB agar with Em and Az. Colonies were restreaked on LB agar, LB agar with Em, LB agar with Em and Az and LB agar with Az, to study the ability of the caut-bCA and caut-gCA genes to complement the ΔCA mutation of E. coli EDCM636.

**Protein purification**

To produce and purify the His-tagged putative CA enzymes from the BL21(DE3) cells, 5 ml LB was inoculated from a – 80 °C stock and grown overnight at 225 rpm, 37 °C. This O/N culture was used to inoculate 5 × 100 ml LB in 500 ml Erlemeyer flasks to a start OD600 of approximately 0.05 and incubated in a shaking incubator at 225 rpm, 37 °C. At OD600, 0.3–0.7 cells were induced with IPTG (Isopropyl β-D-1-thiogalactopyranol), a 0.5 mM final concentration, and incubated at 30 °C for 3–5 h. Cells were harvested by centrifugation, and cells were lysed with a QS1 probesonicator (Nanolabs, MA, USA) or with BugBuster® Plus Lysonase™. An additional centrifuge step was used to create cell-free extract. The protein was purified using a 5 ml HisTrap® HP column (GE Healthcare Life Sciences, Buckinghamshire, UK) and 300 mM imidazole for elution buffer. For the production of the STREP-tagged enzyme, 300 ml of overnight culture (LB, shaking, 37 °C) E. coli EDCM636 pMTL82252-bCAstrepC or EDCM636 pMTL82252-gCAstrepC was harvested by centrifugation and lysed using BugBuster Plus Lysonase with Avidin added at one small crystal per 30 ml. The Caut-bCA protein was purified using Strep-Tactin Sepharose using 100 mM Tris pH 8 with 150 mM NaCl as wash buffer. For elution, 0.5 mg ml⁻¹ desethylbiotin was added to the wash buffer. Samples of all purification fractions were resolved on a NuPAGE Novex 4–12% Bis-Tris Protein Gels (ThermoFisher, Waltham, MA, USA) to assess purity and yield. Protein concentration was further analysed using a NanoDrop Lite photospectrometer (Thermo-Scientific, Wilmington, DE, USA).

**Analytical ultracentrifugation**

For characterisation of the purified Caut-bCA by analytical ultracentrifugation, sedimentation velocity scans were recorded for a 2-fold protein dilution series, starting at 0.7 mg ml⁻¹. The analysis was performed at 50,000 rpm, using a Beckman XL-I analytical ultracentrifuge with an An-50Ti rotor, and results were obtained by absorbance measured at 280 nm and interference optical detection systems. The density and viscosity of the buffer were measured using a DMA 5000 M densitometer equipped with a Lovis 200ME viscometer module. The partial specific volume for the protein was calculated using Sednterp from the amino acid sequence. Data were

---

NaCl 0.2 g l⁻¹ with trace elements (H₂BO₃ 100 μg l⁻¹, MnCl₂·4H₂O 230 μg l⁻¹, FeCl₂·4H₂O 780 μg l⁻¹, CoCl₂·6H₂O 103 μg l⁻¹, NiCl₂·6H₂O 602 μg l⁻¹, ZnCl₂ 78 μg l⁻¹, CuSO₄·5H₂O 50 μg l⁻¹, Al₂(OH)₃·12H₂O 50 μg l⁻¹, Na₂SeO₃ 58 μg l⁻¹, Na₂WO₄ 53 μg l⁻¹, Na₂MoO₄·2H₂O 52 μg l⁻¹) and vitamins (p-aminobenzoate 57 μg l⁻¹, riboflavin 52 μg l⁻¹, thiamine 100 μg l⁻¹, nicotinate 103 μg l⁻¹, pyridoxine 255 μg l⁻¹, calcium pantothenate 52 μg l⁻¹, cyanoacobalamin 39 μg l⁻¹, biotin 11 μg l⁻¹, folate 24 μg l⁻¹, thionic acid 25 μg l⁻¹, at pH 5.8. All cultures of C. autoethanogenum were grown at 37 °C in a MG1000 anaerobic workstation with TG airlock (Don Whitley, UK).

All E. coli strains were grown on LB medium (tryptone 10 g l⁻¹, yeast extract 5 g l⁻¹, NaCl 10 g l⁻¹). To induce the Cyn operon and thus enable growth of C. autoethanogenum were grown at 37 °C in a MG1000 anaerobic workstation with TG airlock (Don Whitley, UK).

Protein concentration was further analysed using a NanoDrop Lite photospectrometer (Thermo-Scientific, Wilmington, DE, USA).
processed using SEDFIT, fitting to the c(s) or non-interacting discrete species (NIDS) model (Schuck 2000).

CA activity assay

To determine CA activity, we developed an assay using a Tecan M1000-Pro (Männedorf, Switzerland) plate reader with auto injector, based on previous methods (Wilbur and Anderson 1948; Sundaram et al. 1986; Fasseas et al. 2011; Gai et al. 2014). We have validated the assay using bovine CA (Sigma) and C. jejuni CanB (Al-Haideri et al. 2016) (kindly supplied by D.J. Kelly of the University of Sheffield). The assay buffer was 50 mM HEPES, 50 mM Na₂SO₄, 0.04% (w/v) phenol red at several pH values. The substrate for the hydration reaction was carbon dioxide-saturated water, produced by bubbling carbon dioxide through demineralised water at 20 °C for 30 min. This should result in 34 mM CO₂ (Diamond and Akinfiev 2003); for lower concentrations, the CO₂-saturated water was diluted with demineralised water that was sparged with N₂ for 30 min. The substrate of the dehydration reaction was KHCO₃ at 100 mM. For the hydration reaction, 120 μl of assay buffer was mixed with 10 μl enzyme sample or 10 μl sample buffer in a 96-well plate. After measuring the baseline for 4 s, 120 μl of substrate was injected. For the dehydration reaction, 140 μl and 10 μl enzyme samples or 10 μl sample buffer was used, after 4 s baseline measurement 50 μl substrate was added. Change of absorption was measured at 557 nm for 40 s at a temporal resolution of 200 ms. The average change in absorption s⁻¹ of the first ten readings was taken as the initial speed of the reaction. The Km and Vmax were calculated using the Michaelis and Menten (1913) curve fitting tool of GraphPad Prism 7.00 (La Jolla, CA, USA).

Results

To establish if C. autoethanogenum harbours CA genes in its genome, we collected 41 CA protein sequences covering the α-, β-, γ- and δ-CA classes of organisms from all domains of life. We assembled consensus sequences of the α-, β- and γ-CA classes and then used amino acid sequences and the assembled consensus sequences to search the genome of C. autoethanogenum. One gene that encoded a putative γ-CA (Caut-gCA, CLAU_2699) was found with a PSI-BLAST E-value of 3e-50 and 54% sequence identity to the

### Table 1

| Primer name     | Primer sequence                                                                 |
|-----------------|---------------------------------------------------------------------------------|
| FgCAOE          | TATACATATGATAAGAGAAATTGGAACATTAC                                               |
| RgCAOE          | TATAGAATTCTAATATCCATTACAAATTTTAC                                               |
| FgCAOESrepN     | TATACATATGTTGTCACATCTCAATTTTCAAAAATTTGGAACATTAC                                 |
| RgCAOESrepC     | TATAGAATTCTAATATCCATTACAAATTTTCAAAAATTTGGAACATTAC                                 |
| RgCAOEHisC      | TATAGAATTCTAATATCCATTACAAATTTTCAAAAATTTGGAACATTAC                                 |
| FbCAOE          | TATACATATGTTGACAGTGTTTGTGACATTAC                                               |
| RbCAOE          | TATAGAATTCTAATATCCATTACAAATTTTCAAAAATTTGGAACATTAC                                 |
| FbCAOESrepN     | TATACATATGTTGTCACATCTCAATTTTCAAAAATTTGGAACATTAC                                 |
| RbCAOESrepC     | TATAGAATTCTAATATCCATTACAAATTTTCAAAAATTTGGAACATTAC                                 |
| FpMTL8xx5x      | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| FpPGA           | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| RpgCA           | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| BCAECPF         | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| BCAECPRI        | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| FPCA1           | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| RPCA1           | GATGGTCAATATGTAAGAGAAATTGGAACATTAC                                               |
| pET16b_F        | GATCCGGCGAAATATACGA                                                               |

© Springer
γ-CA consensus sequence. This gene was annotated as a hexapeptide repeat-containing protein, a feature that exists in γ-CA (Iverson et al. 2000). It had the three important metal-binding histidine residues conserved but lacks glutamine and asparagine residues that were found to be important in the canonical γ-CA (Cam) from *Methanosarcina thermophile* (Ferry 2010).

A gene coding for a putative β-CA (Caut-bCA, locus tag CLAU_3021) was found using the β-CA consensus sequence in PSI-BLAST and matched with an *E* value of 1e−80. Despite the low similarity to other β-CAs, this gene encodes a protein that contains the motifs CxDxR and HxxC, which are known to be important for the activity of β-CAs (Smith and Ferry 2000; Smith et al. 2002).

In recently published transcriptomics data, caut-bCA is consistently higher expressed than caut-gCA (148.177 FPKM vs 71.7564 FPKM) (Marcellin et al. 2016). Downstream of the caut-gCA gene, a peptidase M14 carboxypeptidase was present, which has no obvious link to CAs except that both are Zn metalloenzymes. The caut-bCA gene was part of a cluster (CLAU_3019-CLAU_3023) of five genes with similar expression pattern (unpublished RNAseq data). A *yopX* gene and a single-stranded DNA-binding protein-encoding gene that are present in this cluster indicate a phage origin. The gene immediately downstream of Caut-bCA was annotated as a deoxyuridine 5′-triphosphate nucleotidohydrolase which indicates a potential function in pyrimidine metabolism where HCO₃⁻ is an important factor.

Protparam results showed that the 124 amino acid residues of Caut-bCA formed a stable protein with a molecular weight of 14.2 kDa and pI of 5.3. The 168 amino acid residues of Caut-gCA were expected to form an unstable protein with a mass of 18.3 kDa and pI of 6.4. Phyre2 structural modelling showed Caut-bCA as a small but not untypical β-CA with an accessible active site.
Functional complementation of E. coli Can<FLK2>

The Can disruption mutant of E. coli EDCM636 (Merlin et al. 2003), which lacks a constitutively expressed CA and is therefore unable to grow under atmospheric carbon dioxide pressure, was used to test if Caut-bCA and Caut-gCA exhibited CA activity. When transformed with plasmids carrying the caut-bCA or caut-gCA gene, only E. coli EDCM636 complemented with caut-bCA was able to grow in the absence of azide (Fig. 1). E. coli EDCM636 cells transformed with pMTL82252-bCAstrepC also grew without the absence of azide, indicating the C-terminal strep tag did not impair activity of the enzyme. We did find that E. coli EDCM636 cells infrequently reverted to the native phenotype and grow in the absence of azide. Therefore, we repeated this experiment three times with fresh batches of competent cells that did not show WT phenotypes.

Development of a high-throughput CA activity assay method

To establish whether the identified genes encode for active CA enzymes, we developed a high-throughput activity assay that was performed in 96-well format. As far as the authors know, this is the first description of performing CA assays in 96-well format. The performance of the assay was assessed with bovine α-CA with a dose–effect response in the range of 0.003–3 μg ml⁻¹ of enzyme. It was further validated by measuring the $K_M$ of bovine α-CA and the pH-specific activity of the C. jejuni CanB, a β-CA (Al-Haideri et al. 2016). The method replicated the specific pH profile of C. jejuni CanB, typical of type II β-CAs, and the $K_M$ was determined as 4.7 ± 2 mM at 20 °C. Al-Haideri et al. found a $K_M$ of 34 ± 10 mM at 4 °C; however, this temperature could not be replicated in our Tecan M1000-Pro. The $K_M$ of bovine CA was determined as 17 ± 4 mM while published values vary between 12 and 1.1 mM (Kernohan 1964; Iqbal et al. 2014). We continued to apply this assay in the characterisation of C. autoethanogenum CAs.

Enzyme characterisation

To characterise the specific activity of the putative CAs of C. autoethanogenum, they were heterologously expressed, purified and assayed for the activity. Purification of the Caut-bCA and Caut-gCA with N-terminal His-tag (Hochuli et al. 1988) on a pET16b plasmid expressed in BL21(DE3) pLysS cells did not reliably yield active enzymes (data not shown). A C-terminal STREPII-tagged (Schmidt and Skerra 2007) Caut-bCA protein, expressed from pMTL82252 in the E. coli EDCM636 strain, was reliable purified (Fig. 2) and used for further characterisation of the enzyme. We did not manage to show CA activity for Caut-gCA or Caut-gCAstrepC with any of the tested systems and did not further characterise this enzyme. The developed high-throughput CA assay method was used to measure the kinetic parameters of the hydration reaction of the Caut-bCAstrepC enzyme (Fig. 3). The $K_M$ for the hydration reaction was measured at 6.8 ± 1.6 mM, and, for the dehydration reaction, it was 10.5 ±
2.5 mM. The measured kinetic parameters of the hydration reaction were comparable with those of the other reported β-CAs (Table 2), and comparative data is mostly lacking for the dehydration reaction. Incubation of the purified enzyme at 95 °C for 10 min caused total loss of activity (data not shown).

Analytical ultracentrifugation of Caut-bCAstrepC (Table 3 and Fig. 4) showed that molecular weight of 30 kDa obtained for the main peak of the derived sedimentation coefficient distribution was consistent with that of a dimer.

### Table 2 Kinetic parameters of Caut-bCA and previously described beta-carbonic anhydrases

| Species                          | Monomer size (kDa) | Oligomeric state | Clade | Hydration | Dehydration | Reference                      |
|----------------------------------|--------------------|------------------|-------|-----------|-------------|--------------------------------|
| *Clostridium autoethanogenum*    | 14.2               | Dimer            | F     | 2.1 × 10^5 | 6.3 × 10^4 | 3.1 × 10^7 | 6.0 × 10^6 | This study |
| *Pisum sativum*                  | 24.2               | Octamer          | B     | 4 × 10^5   | 1.8 × 10^7 | NA    | NA    | Kisiel and Graf (1972); Johanson and Forsman (1993) |
| *Cryptococcus neoformans*        | 26                 | Dimer            | A     | 3.9 × 10^5 | NA    | 4.3 × 10^7 | NA    | Innocenti et al. (2008) |
| *Drosophila melanogaster*        | 30.0               | Dimer            | B     | 9.5 × 10^5 | NA    | 1.1 × 10^8 | NA    | Syrjänen et al. (2010) |
| *Clostridium perfringens* strain 13 | 21.3              | Tetramer         | D     | 1.5 × 10^6 | NA    | 4.8 × 10^6 | NA    | Kumar et al. (2013) |
| *Methanobacterium thermautotrophicus* | 18.9           | Tetramer         | D     | 1.7 × 10^4 | NA    | 5.9 × 10^6 | NA    | Smith and Ferry (1999) |
| *Halothiobacillus neapolitanus*  | 57.3               | Dimer            | E     | 8.9 × 10^5 | 4.6 × 10^4 | 2.8 × 10^7 | 4.9 × 10^6 | Sawaya et al. (2006); Heinhorst et al. (2006) |
| *Salmonella enterica* (stCAI)    | 24.8               | NA               | A     | 7.9 × 10^5 | NA    | 8.3 × 10^7 | NA    | Nishimori et al. (2011) |
| *Salmonella enterica* (stCAII)   | 26.6               | NA               | C     | 8.9 × 10^5 | NA    | 5.2 × 10^7 | NA    | Nishimori et al. (2011) |

NA, not available

2.5 mM. The measured kinetic parameters of the hydration reaction were comparable with those of the other reported β-CAs (Table 2), and comparative data is mostly lacking for the dehydration reaction. Incubation of the purified enzyme at 95 °C for 10 min caused total loss of activity (data not shown).

Analytical ultracentrifugation of Caut-bCAstrepC (Table 3 and Fig. 4) showed that molecular weight of 30 kDa obtained for the main peak of the derived sedimentation coefficient distribution was consistent with that of a dimer.

### Table 3 Caut-bCAstrepC estimated molecular weights from c(s) and NIDS analysis by analytical ultracentrifugation

| Monomer MW (kDa) | Detection method | Concentration (mg/mL) | Major species | Peak 1 | Peak 2 | f/foa |
|------------------|------------------|-----------------------|---------------|--------|--------|-------|
|                  |                  |                       |               | MW (kDa) | Sed. co (S) | MW (kDa) | Sed. Co (S) | |
| C(s)             |                  |                        |               | 0.70   | 9.3     | 1.03   | 34.6   | 2.48 | 1.44 |
| Asorbance        | 0.35             | 8.4                   | 1.01          | 32.8   | 2.49 | 1.37 |
|                  | 0.18             | 9.3                   | 1.02          | 35.1   | 2.48 | 1.45 |
| 15.2             |                  |                        |               | 0.70   | –       | –      | 31.3   | 2.48 | 1.29 |
| Interference     | 0.35             | 13.0                  | 1.45          | 29.3   | 2.49 | 1.23 |
|                  | 0.18             | 11.9                  | 1.32          | 30.6   | 2.48 | 1.27 |
| NIDS             |                  |                        |               | 0.70   | 10.1    | 1.04   | 29.3   | 2.49 | N/A |
| 15.2             | Asorbance        | 0.35                  | 9.7           | 1.02   | 29.4   | 2.51 | N/A |
|                  | 0.18             | 3.6                   | 0.94          | 29.8   | 2.49 | N/A |

a Best-fit frictional ratio f0

### Phylogeny

The low identity of the putative Caut-bCA with known β-CAs triggered further phylogenetic analysis. A Bayesian phylogenetic tree of 60 β-CA sequences was constructed (Fig. 5) which present the overall topology of identified clades, consistent with that found with other methods (maximum likelihood) and larger sets of β-CA sequences (data not shown). The analysis included sequences of the majority of previously described β-CAs as well as β-CAs
from the major taxonomic groups of life. We identified six major clades within the β-CAs (Fig. 5). The A, B, C and D clades are as described previously (Smith and Ferry 2000), and the E clade was formed by the former ε-CAs. The Caut-bCA-like proteins formed a distinct F clade, of not previously studied β-CAs.

An alignment of 160 β-CA protein sequences was summarised as consensus logos for all six identified subclades (Online Resource 1). Only two motifs interspaced by 48–77 amino acids were fully conserved in all selected β-CAs. These were the HxxC motifs, which bind the active site metal ion and the CxDxR motif that completes the active site (Rowlett 2010). Additional motifs were identified for subgroups of the clades. Enzymes in the A, B and C clades all contained a QxP motif of six amino acids N-terminal of the CxDxR motif, while D and E clades lack this motif. In most of the novel F clade, the QxP motif was found on the C-terminal side of the CxDxR motif. The A, B and C clades further contained a conserved G [D/E] xFxR motif in sequence that was flanked by the CxDxR and HxxC motifs. In the E clade, only GxxF was conserved, and in the D clade, G [D/E] was conserved. The distance between CxDxR and HxxC was larger than average in the E clade with approximately 75 aa, and to a lesser extent in the F clade with average distance of 64 aa, where A, B, C and D clades had an average distance of 55 aa.

Both phylogenetic and alignment analysis showed that the F clade was further split in two sub-clades, one defined by the putative β-CA of M. thermoacetica (Mtherm-bCA-like) and the other by Caut-bCA. F clade CAs were mostly of shorter sequence than the previously described CAs of the other clades. The Mtherm-bCA-like subclade harbours six proteins of considerable length (172–183 aa). The Caut-bCA-like clade consists solely of relatively short CAs with a length of between 124 and 142 residues.

### Discussion

Two putative CA genes, Caut-bCA and Caut-gCA, were identified in the genome of C. autoethanogenum as potential members of the β- and γ-classes of CA. However, a low sequence similarity of Caut-bCA with other members of the β-CA and a reported lack of CA activity for several γ-CA homologs (Martin et al. 2009; Al-Haideri et al. 2016; Kaur et al. 2010; Ferry 2010), did not allow to ascribe CA function to either of these genes based on gene sequence alone. Neither transcription profile nor genomic contexts of the identified genes revealed a specific function for these genes. To assess CA activity for the proteins encoded by these genes, these were heterologously expressed in a Can disruption mutant of E. coli (strain EDCM636). This showed that Caut-bCA could complement the lack of CA activity in this strain while Caut-gCA did not. This confirmed that Caut-bCA is a carbonic anhydrase while Caut-gCA is not a functional CA in this context. The results of activity assays on purified Caut-gCA and Caut-bCA proteins were consistent with this complementation study as no tested gCA construct yielded any activity while the pMTCaut-bCAstrepc constructs yielded active purified enzymes. We suspect that the bulky metal binding 10× His residue close to the metal ion-binding active site interfered with the activity of Caut-bCA N-terminal His-tag (Hochuli et al. 1988) on a pET16b plasmid expressed in BL21(DE3) pLysS cells and therefore did not reliably yield active enzyme. Since E. coli EDCM636 does not grow without Caut-bCA complementation, Caut-bCA must be produced as an active enzyme. We suspect that this further enhanced the reproducibility of this system.

The measured kinetic parameters were similar to other reported β-CAs. The $K_M$ of 6.8 mM was slightly higher than that for the β-CAs of Clostridium perfringens and Methanothermobacter thermautotrophicus but lower than those of Salmonella enterica or Helicobacter pylori. The particular low $K_M$ of the C. perfringens CA is interpreted as an indication for a function in retaining intracellular levels for anaplerotic CO2 fixation reactions (Kumar et al. 2013).

The molecular weight of ~30 kDa determined by analytical ultra-centrifuge shows that Caut-bCA is a small dimeric protein. A Rhodospirillum rubrum CA was purified with a similar size (28 kDa) and oligomeric state (dimer); however, no matching gene or protein sequence was identified, and later analysis of this species of β-CA showed larger sizes for its β-CA monomers (Gill et al. 1984; Smith et al. 1999). This makes Caut-bCA the smallest confirmed β-CA. The measured molecular weight of the dimer is close to some of the monomers of β-CAs of other species, illustrating the compact nature of Caut-bCA (Table 2).

Because of the low identity of the putative Caut-bCA with known β-CAs, further phylogenetic analysis was performed. We identified multiple orthologs of Caut-bCA in both bacteria.
and archaea annotated as hypothetical genes. It was striking to find these specific Caut-bCA-like CA sequences relatively often in the deposited genomes of uncultivated bacteria and archaea or recent isolates of candidate species. However, we limited our further phylogenetic analysis to previously cultured species. The phylogenetic analysis revealed that the group of Caut-bCA-like CAs forms a distinct F clade of not previously studied $\beta$-CAs that are somewhat distantly related to $\beta$-CAs from other groups. Fig. 5 $\beta$-CA phylogeny. The phylogeny of $\beta$-CAs was reconstructed using MrBayes. All posterior probabilities are above 0.85 except for branches labelled with * which have a probability between 0.55 and 0.7. A similar topology was obtained with more extended sets of sequences, both with maximum Likelihood method and Bayesian analysis (data not shown). The tree was edited for clarity using Figtree and Inkscape.
to the other β-CAs of the A, B, C and D subclasses. The further split of the F clade in two subclasses (Mtherm-bCA-like and Caut-bCA-like) might be of consequence since M. thermoacetica showed little CA activity in previous studies (Braus-Stromeyer et al. 1997). Therefore, some caution is in place to assume functionality on the Mtherm-bCA-like proteins without further study. Based on the similarity of the Caut-bCA-like CAs to Caut-bCA, we assume that these are active CAs. Knowledge on CA in C. autoethanogenum could be used to improve CO₂ utilization or to modulate product spectrum in vivo and for CO₂ capture and storage in vitro (Alvizo et al. 2014; Warden et al. 2015; Lian et al. 2016).

Acknowledgments We would like to thank D.J. Kelly from the University of Sheffield for supplying C. jejuni β-CA Can used for assay validation.

Funding This work was supported by the Biotechnology and Biological Sciences Research Council (grant numbers BB/L502030/1, BB/K00283X/1) as part of the BBSRC Longer and Larger Grant GASCHEM and the industrial partner LanzaTech Inc. GH is supported by the Medical Research Council (UK) grant to the RCaH, and DJS is supported by the Science and Technology Facilities Council (UK).

Compliance with ethical standards

Conflict of interest BP declares that she has no conflict of interest. GH declares that he has no conflict of interest. MK is employed by LanzaTech. LanzaTech has commercial interest in CO₂ utilization. SDS is employed by LanzaTech. LanzaTech has commercial interest in CO₂ utilization. NPM declares that he has no conflict of interest. AMH declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Abrini J, Naveau H, Nyns E-J (1994) Clostridium autoethanogenum, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. Arch Microbiol 161:345–351. https://doi.org/10.1007/BF00303591

Al-Haidery H, White M, Kelly DJ (2016) Major contribution of the type II beta carbonic anhydrase CanB (Cj0237) to the capnophilic growth phenotype of Campylobacter jejuni. Environ Microbiol 18:721–735. https://doi.org/10.1111/1462-2920.13092

Alvizo O, Nguyen LJ, Savile CK, Bresson JA, Lakhapatri SL, Solis EOP, Fox RJ, Broering JM, Benoit MR, Zimmerman SA, Novick LJ, Lajonde JJ (2014) Directed evolution of an ultrastable carbonic anhydrase for highly efficient carbon capture from flue gas. Proc Natl Acad Sci U S A 111:16436–16441. https://doi.org/10.1073/pnas.1411461111

Braus-Stromeyer S, Schnappauf G, Braus G, Gossner A, Drake H (1997) Carbonic anhydrase in Acetobacter woodii and other acetogenic bacteria. J Bacteriol 179:719–7200

Cannon GC, Heinrich S, Kerfeld CA (2010) Carboxysomal carbonic anhydrases: structure and role in microbial CO₂ fixation. Biochim Biophys Acta 1804:382–392. https://doi.org/10.1016/j.bbapap.2009.09.026

Capasso C, Supuran CT (2015) An overview of the alpha-, beta- and gamma-carbonic anhydrases from bacteria: can bacterial carbonic anhydrases shed new light on evolution of bacteria? J Enzyme Inhib Med Chem 30:325–332. https://doi.org/10.3109/14756366.2014.910202

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165. https://doi.org/10.1093/bioinformatics/btr088

Del Prete S, Vullo D, Fisher GM, Andrews KT, Poulsen S-A, Capasso C, Supuran CT (2014) Discovery of a new family of carbonic anhydrases in the malaria pathogen Plasmodium falciparum—the η-carbonic anhydrases. Bioorg Med Chem Lett 24:4389–4396. https://doi.org/10.1016/j.bmcl.2014.08.015

Diamond LW, Akinfiev NN (2003) Solubility of CO₂ in water from –1.5 to 100 °C and from 0.1 to 100 MPa: evaluation of literature data and thermodynamic modelling. Fluid Phase Equilib 208:265–290. https://doi.org/10.1016/S0378-3812(03)00041-4

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Fasseas MK, Tsikou D, Flemetakis E, Katinakis P (2011) Molecular and biochemical analysis of the α class carbonic anhydrases in Caenorhabditis elegans. Mol Biol Rep 38:1777–1785. https://doi.org/10.1007/s11033-010-0292-y

Ferry JG (2010) The gamma class of carbonic anhydrases: structure and role in microbial CO₂ fixation. Biochim Biophys Acta 1804:374–381. https://doi.org/10.1016/j.bbapap.2009.08.026

Gai CS, Lu J, Brigham CJ, Bernardi AC, Sinskey AJ (2014) Insights into bacterial CO₂ metabolism revealed by the characterization of four carbonic anhydrases in Raistonia eutrophila H16. AMB Express 4:2. https://doi.org/10.1186/2191-0855-4-2

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools in the ExPASy server. In: Walk JM (ed) Proteomics protoc. handbook. Humana Press, New York 2005

Gill SR, Fedorka-Cray PJ, Tweten RK, Sleeper BP (1984) Purification and properties of the carbonic anhydrase of Rhodospirillum rubrum. Arch Microbiol 138:113–118. https://doi.org/10.1007/BF00413010

Guillon MB, Lamblin AF, Kozliak EI, Gerami-Nejad M, Tu C, Silverman D, Anderson PM, Fuuchs JA (1993) A physiological role for cyanate-induced carbonic anhydrase in Escherichia coli. J Bacteriol 175:1443–1451

Hawksb AS, McTernan PM, Lian H, Kelly RM, Adams MW (2013) Biological conversion of carbon dioxide and hydrogen into liquid fuels and industrial chemicals. Curr Opin Biotechnol 24:376–384. https://doi.org/10.1016/j.copbio.2013.02.017

Heap JT, Pennington OJ, Carntan ST, Minton NP (2009) A modular system for Clostridium shuttle plasmids. J Microbiol Methods 78:79–85. https://doi.org/10.1016/j.mimet.2009.05.004

Heinhorst S, Williams EB, Cai F, Murin CD, Shively JM, Cannon GC (2006) Characterization of the carboxysomal carbonic anhydrase CsoSCA from Halothiobacillus neapolitanus. J Bacteriol 188:8087–8094. https://doi.org/10.1128/JB.00990-06
modeling. Biophys J 78:1606–1619. https://doi.org/10.1016/S0006-3495(00)76713-0

Simpson SD, Koepke M, Smart KF, Tran LP, Sechrist P (2014) System and method for controlling metabolite production in a microbial fermentation. https://patents.google.com/patent/WO2014151158A1. Accessed 19 June 2019

Smith KS, Ferry JG (1999) A plant-type (beta-class) carbonic anhydrase in the thermophilic methanorarchaeon Methanobacterium thermoautotrophicum. J Bacteriol 181:6247–6253

Smith KS, Ferry JG (2000) Prokaryotic carbonic anhydrases. FEMS Microbiol Rev 24:335–366. https://doi.org/10.1016/S0168-6445(00)00030-9

Smith KS, Jakubzick C, Whittam TS, Ferry JG (1999) Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. Proc Natl Acad Sci 96:15184–15189. https://doi.org/10.1073/pnas.96.26.15184

Smith KS, Ingram-Smith C, Ferry JG (2002) Roles of the conserved aspartate and arginine in the catalytic mechanism of an archaeal-class carbonic anhydrase. J Bacteriol 184:4240–4245. https://doi.org/10.1128/JB.184.15.4240-4245.2002

Sundaram V, Rumbolo P, Grubb J, Strisciuglio P, Sly WS (1986) Carbonic anhydrase II deficiency: diagnosis and carrier detection using differential enzyme inhibition and inactivation. Am J Hum Genet 38:125–136

Supuran CT (2016) Structure and function of carbonic anhydrases. Biochem J 473:2023–2032. https://doi.org/10.1042/BCJ20160115

Supuran CT, Ferry JG (2013) Carbonic anhydrases of anaerobic microbes. Bioorg Med Chem 21:1392–1395

Supuran CT, Scozzafava A (2007) Carbonic anhydrases as targets for medicinal chemistry. Bioorg Med Chem 15:4336–4350. https://doi.org/10.1016/j.bmc.2007.04.020

Syrjänen L, Tolvanen M, Hilvo M, Olatubosun A, Innocenti A, Scozzafava A, Leppimäki J, Niederhauser B, Hytönen VP, Gorr TA, Parkkila S, Supuran CT (2010) Characterization of the first beta-class carbonic anhydrase from an arthropod (Drosophila melanogaster) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates. BMC Biochem 11:28. https://doi.org/10.1186/1471-2091-11-28

Ueda K, Tagami Y, Kamihara Y, Shiratori H, Takano H, Beppu T (2008) Isolation of bacteria whose growth is dependent on high levels of CO2 and implications of their potential diversity. Appl Environ Microbiol 74:4535–4538. https://doi.org/10.1128/AEM.00491-08

Ueda K, Nishida H, Beppu T (2012) Dispensabilities of carbonic anhydrase in proteobacteria. Int J Evol Biol 2012:324549. https://doi.org/10.1155/2012/324549

Warden AC, Williams M, Peat TS, Seabrook SA, Newman J, Dojchinov G, Haritos VS (2015) Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. Nat Commun 6:10278. https://doi.org/10.1038/ncomms10278

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wilbur KM, Anderson NG (1948) Electrometric and colorimetric determination of carbonic anhydrase. J Biol Chem 176:147–154

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.