CrossMark

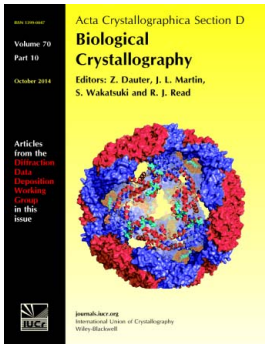
electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The structure of a tetrameric α-carbonic anhydrase from *Thermovibrio ammonificans* reveals a core formed around intermolecular disulfides that contribute to its thermostability

Paul James, Michail N. Isupov, Christopher Sayer, Vahid Saneei, Svein Berg, Maria Lioliou, Hans Kristian Kotlar and Jennifer A. Littlechild

Acta Cryst. (2014). D**70**, 2607–2618



Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The structure of a tetrameric α-carbonic anhydrase from *Thermovibrio ammonificans* reveals a core formed around intermolecular disulfides that contribute to its thermostability

Paul James,[a] Michail N. Isupov,[a] Christopher Sayer,[a] Vahid Saneei,[a] Svein Berg,[b] Maria Lioliou,[b] Hans Kristian Kotlar[b] and Jennifer A. Littlechild[a]*

[a]Henry Wellcome Building for Biocatalysis, Biosciences, College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, England, and [b]Statoil Research Centre, Statoil, Arkitekt Ebbells veg 10, Rotvoll, 7005 Trondheim, Norway

Correspondence e-mail: j.a.littlechild@exeter.ac.uk

Received 17 April 2014
Accepted 16 July 2014

PDB references: α-carbonic anhydrase, 4c3t; SAN complex, 4coq; AZM complex, 4uov

Carbonic anhydrase enzymes catalyse the reversible hydration of carbon dioxide to bicarbonate. A thermophilic *Thermovibrio ammonificans* α-carbonic anhydrase (TaCA) has been expressed in *Escherichia coli* and structurally and biochemically characterized. The crystal structure of TaCA has been determined in its native form and in two complexes with bound inhibitors. The tetrameric enzyme is stabilized by a unique core in the centre of the molecule formed by two intersubunit disulfides and a single lysine residue from each monomer that is involved in intersubunit ionic interactions. The structure of this core protects the intersubunit disulfides from reduction, whereas the conserved intrasubunit disulfides are not formed in the reducing environment of the *E. coli* host cytosol. When oxidized to mimic the environment of the periplasmic space, TaCA has increased thermostability, retaining 90% activity after incubation at 70°C for 1 h, making it a good candidate for industrial carbon-dioxide capture. The reduction of all TaCA cysteines resulted in dissociation of the tetrameric molecule into monomers with lower activity and reduced thermostability. Unlike other characterized α-carbonic anhydrases, TaCA does not display esterase activity towards *p*-nitrophenyl acetate, which appears to result from the increased rigidity of its protein scaffold.

1. Introduction

Carbonic anhydrases (CAs; carbonate dehydratases; EC 4.2.1.1) are a family of enzymes that catalyse the reversible conversion of carbon dioxide and water to bicarbonate and a proton (Tripp *et al.*, 2001). Among several others, they have potential use for CO_2 capture in industrial processes. The need for CO_2 capture and storage has generated a large amount of interest in recent years owing to concerns over global warming. Industrial-scale pilot and demonstration plants for CO_2 capture are currently being operated in fossil-fuel power plants. Carbon dioxide is trapped on an absorber column *via* an absorption process when flue gas is passed through. Relatively pure CO_2 can be then stripped off from the solvent in a heated column, compressed and stored (Savile & Lalonde, 2011). Amines are the most commonly used family of solvents for this purpose. Several solvent-based technologies have been evaluated with regard to their cost, performance, environmental impact and energy efficiency. CA enzymes have been shown to assist in such a chemical process by reducing the heat of desorption, thereby decreasing the amount of energy and potentially the recirculation of solvent needed to fulfill the process. There is an increased interest in the use of CA enzymes for these applications, since they are a cost-effective and safe alternative to the current solvent-based methods for CO_2 capture (Savile & Lalonde, 2011).

There are five distinct families of CAs (α, β, γ, δ and ε; Smith *et al.*, 1999), and while the amino-acid sequences of these proteins are conserved within each family, there is no sequence or structural similarity between the different families. The CA activity requires the presence of a catalytic zinc ion coordinated to three zinc-binding residues, which are either histidine or cysteine depending on the class of the enzyme (Silverman & Lindskog, 1988). Animals have multiple isoforms of α-CAs, which are essential for respiration (Supuran *et al.*, 2003).

Research studies on enzyme-based CO_2 capture have been concentrated on the use of α-CAs as they have a relatively short amino-acid sequence, are known to be very efficient catalysts of CO_2 hydration (De Luca *et al.*, 2013; Gai *et al.*, 2014) and are easily identifiable within thermophilic bacterial genomes. The reaction that is catalysed by α-CAs proceeds through a two-stage ping-pong mechanism. The first stage of the reaction is nucleophilic attack by the zinc-bound hydroxide ion on CO_2 (1). This results in the formation of a bicarbonate ion, which is subsequently removed and replaced by a water molecule at the active site. The zinc-bound hydroxide is then regenerated *via* the transfer of a proton between the zinc-bound water molecule and the buffer (2) (Lindskog, 1997; Domsic *et al.*, 2008). The proton transfer occurs *via* a histidine shuttle residue (His64 in human carbonic anhydrase II; HCAII), which is observed in both 'in' and 'out' conformations in many of the CA structures that have been reported (Tu *et al.*, 1989; Maupin & Voth, 2007).

$$CO_2 + CAZnOH^- \rightleftharpoons CAZnHCO_3^- \overset{H_2O}{\rightleftharpoons} CAZnH_2O + HCO_3, \quad (1)$$

$$CAZnH_2O + B \rightleftharpoons CAZnOH^- + BH^+. \quad (2)$$

The monomer of α-CAs is made up from a single domain which is built around a pleated β-sheet (Liljas *et al.*, 1972). The bound zinc is tetrahedrally coordinated by three histidine residues and a hydroxyl ligand. The active site of α-CAs has a hydrophobic pocket on one side of the catalytic zinc that is proposed to bind the CO_2 molecule (Domsic & McKenna, 2010). On the opposite side of the active site there is a hydrophilic area that favours the shuttling of protons between the catalytic site and the external buffer.

The structures of mammalian and mesophilic bacterial α-CAs and their complexes with different inhibitors have been extensively studied. Some of these inhibitors have potential medicinal applications for the treatment of glaucoma, obesity and cancer (Supuran *et al.*, 2003). Also, the human CA enzymes have been shown to have an application in cancer drug-delivery systems (Lloyd *et al.*, 2005).

Attempts have been made to increase the thermal stability of the human α-CA enzyme for potential industrial applications whilst attempting to preserve its fast rate of reaction. However, this has achieved only limited success (Fisher *et al.*, 2012). To overcome the requirement for a more robust CA which is necessary for industrial-scale carbon-dioxide capture, bioinformatics studies have been carried out using the available thermophilic genomes in an attempt to identify more stable CA enzymes. Both β-CA and γ-CA enzymes have been identified in thermophilic archaea (Smith & Ferry, 1999), and more recently α-CA enzymes have been reported in the thermophilic bacteria *Sulfurihydrogenibium yellowstonense* (SspCA) and *S. azorense* (SazCA) (Akdemir *et al.*, 2013; Vullo *et al.*, 2013). Recently, the SspCA enzyme structure has been determined and its observed increased thermal stability has been attributed to structural compactness, an overall increase of charged residues on the protein surface and a number of ionic networks (Di Fiore *et al.*, 2013).

This paper describes the biochemical and structural characterization of a novel thermostable α-CA enzyme identified from the anaerobic, chemolithoautotrophic bacterium *Thermovibrio ammonificans*. This organism has an optimal growth temperature of 75°C and was isolated from the walls of an active deep-sea hydrothermal vent chimney on the East Pacific Rise (Vetriani *et al.*, 2004). It uses carbon dioxide as a carbon source and requires CA enzyme activity for growth. This carbonic anhydrase (TaCA) has been cloned and overexpressed in *Escherichia coli*. The X-ray crystal structures of the native TaCA and its complexes with the inhibitors acetazolamide (AZM) and sulfanilamide (SAN) have been determined to resolutions of 1.7, 1.9 and 1.5 Å, respectively.

2. Experimental

2.1. Materials

Restriction endonucleases and DNA-modifying enzymes were obtained from Thermo Scientific (Hemel Hempstead, England). All oligonucleotides were obtained from Eurofins MWG Biotech (Ebsberg, Germany). All other reagents were obtained from Sigma–Aldrich (Buchs, Switzerland) unless otherwise stated.

2.2. Cloning, overexpression and purification

The protein sequence of bovine α-CA I (BCA1; accession No. NP_848667.1) was used as a query in a *BLASTP* search (Altschul *et al.*, 1990) against the nonredundant protein sequence database to identify the TaCA sequence (accession No. YP_004152175.1). The gene was then synthesized using the GeneArt service with the codon usage optimized for *E. coli* using the *GeneOptimizer* software (Life Technologies; Fath *et al.*, 2011). Primers were then designed to subclone the optimized gene sequence into the pLATE51 vector (Thermo Scientific, Hemel Hempstead, England). The TaCA sequence plus the compatible vector overhangs was amplified by PCR using a 5′ primer containing the 5′ vector overhang (in italics) *GGTGATGATGATGACAAG*AAACGTGTTCTGGTTAC-CC and a 3′ primer containing the 3′ vector overhang (in italics) *GGAGATGGGAAGTCA*TTATTTCATAACTTTA-CGTGCATTCAG. The resulting 744 bp amplified fragment was treated with T4 DNA polymerase to create the compatible overhangs before being annealed to the vector. Restriction-enzyme digestion and DNA sequencing was used to confirm

the presence of the TaCA gene sequence in the recombinant plasmid designated pTaCA51.

Plasmid pTaCA51 was transformed into *E. coli* BL21 (DE3) Star cells (Invitrogen, Carlsbad, California, USA) and cultured overnight in LB medium. Starter cultures (10 ml) were then used to inoculate 500 ml ZYM-5052 medium (Studier, 2005) and this was left to incubate at 20°C for 2 days with shaking at 250 rev min^{-1}. The cells were then harvested at 5000*g* at 10°C for 20 min. The pellet was then resuspended in 25 ml 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl. The cells were disrupted by sonication at 10 µm (Soniprep 150; MSE, London, England) on ice for 4 min and the cell debris was removed by centrifugation at 20 000*g* at 4°C for 30 min. The clarified cell lysate was then heat-treated at 65°C for 30 min before being centrifuged at 20 000*g* at 4°C for 30 min to remove any denatured proteins.

The protein was purified using a 1 ml HisTrap FF crude column (GE Healthcare, Little Chalfont, England) using an elution gradient from 20 to 500 m*M* imidazole in 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl. The enzyme was then applied onto a calibrated Superdex 200 HiLoad 16/60 gel-filtration (GF) column (GE Healthcare, Little Chalfont, England) and was eluted with one column volume of 10 m*M* HEPES, 0.1 *M* NaCl pH 8.0 at 1.0 ml min^{-1}. To prepare the reduced protein (rTaCA), a sample of the protein was also treated with the reducing agent dithiothreitol (DTT; 10 µ*M*) and then applied onto the calibrated GF column. To prepare an oxidized sample of the protein (oTaCA) the medium was supplemented with 250 µ*M* diamide during protein expression, which has been shown to promote the formation of disulfide bonds (Andersen *et al.*, 1997; James, 2010). The oTaCA was purified as described above with 10 µ*M* diamide added to all of the buffer solutions. The proportion of protein containing the intersubunit disulfide was monitored using SDS–PAGE with nonreducing sample buffer.

2.3. Assay for activity using CO_2 as the substrate

The activity of this enzyme was assayed using a modification of the method described by Sundaram *et al.* (1986). A 250 ml volume of H_2O was hydrated with CO_2 by bubbling the gas through the solution for 1 h at room temperature (34 m*M* CO_2 solution). The assay buffer contained phenol red to act as an indicator for the reaction (10 m*M* Tris–HCl pH 8.3, 2 m*M* phenol red, 20 m*M* Na_2SO_4). 1 ml of freshly prepared CO_2 substrate was injected into 2 ml of the assay buffer with 10 µl sample (or 10 µl buffer for the buffer control). The reaction was monitored by the change in absorbance at 557 nm. One unit of modified Wilbur–Anderson activity (MWA) is defined as (time of reaction without enzyme in seconds − time of reaction with enzyme in seconds)/time of reaction with enzyme in seconds required for the pH to decrease from the value of 8.3 for the Tris–HCl buffer to the transition point of the dye (Sundaram *et al.*, 1986).

2.4. Thermal stability of TaCA

The thermal stabilities of the enzymes (TaCA, rTaCA and oTaCA) were tested by incubating the enzymes at a concentration of 1 mg ml^{-1} in 10 m*M* HEPES, 0.1 *M* NaCl pH 8.0 at 30, 40, 50, 60, 70, 80, 90 and 100°C for 1 h. Enzyme aliquots (100 µl) were withdrawn at appropriate times and cooled on ice before the residual activity was measured using the method described above. All reactions were performed in triplicate.

2.5. Esterase assay

Activity for the hydrolysis of an ester was determined at 25°C using *p*-nitrophenyl acetate (p-NpA) in a modification of the method described by Armstrong *et al.* (1966). A 0.3 ml volume of freshly prepared 3 m*M* p-NpA in acetone was added to 0.7 ml 100 m*M* potassium phosphate buffer pH 7.5. A 10 µl aliquot of the enzyme solution (1 mg ml^{-1}) was added to the mixture and the change in A_{348} per minute was recorded for 5 min (Armstrong *et al.*, 1966). A control reaction was carried out using bovine carbonic anhydrase I (BCA1; Sigma–Aldrich, Buchs, Switzerland) at 1 mg ml^{-1} dissolved in 10 m*M* HEPES, 0.1 *M* NaCl pH 8.0.

2.6. Kinetic characterization of TaCA

The CO_2 hydration activity of TaCA was measured using an Applied Photophysics SX18MV (Leatherhead, England) stopped-flow instrument. The initial rates of the CO_2 hydration reaction were measured at 25°C using 0.2 m*M* phenol red as the indicator at an absorbance of 557 nm. The buffer was 20 m*M* HEPES pH 7.5 containing 20 m*M* Na_2SO_4 to maintain the ionic strength. The CO_2 calibration was calculated between 1.7 and 17 m*M* to determine the kinetic parameters. The non-enzymatic CO_2 hydration rates were determined and subtracted from the initial observed rates and the kinetic parameters were calculated using *PRISM* 5.0 (San Diego, California, USA) and represent mean values from five different observations.

2.7. Crystallization, data collection and structural determination

The TaCA was concentrated to ∼10 mg ml^{-1} using a 10 kDa Vivaspin membrane (Vivaproducts, Littleton, Massachusetts, USA) and microbatch crystallization trials were set up using an Oryx 6 crystallization robot (Douglas Instruments, Hungerford, England) using the JCSG+ and Morpheus screens (Molecular Dimensions, Newmarket, England; Newman *et al.*, 2005). The droplet consisted of a 50:50 ratio of protein solution to screen solution and was covered with Al's oil (a 50:50 mixture of silicone oil and paraffin) before being stored at 18°C and was regularly checked for growth of crystals using a light microscope.

Crystals appeared within two weeks in several wells, and crystals from three conditions were used for data collection: (i) 0.5 *M* LiCl, 50 m*M* sodium citrate, 10% PEG 6000 pH 4.0, (ii) 50 m*M* bis-tris propane, 12.5% PEG 3350 pH 5.5 and (iii) 9% ethylene glycol, 9% PEG 8000, 10 m*M* carboxylic acids additive mixture, 100 m*M* MES pH 6.5. A crystal from condition (i) was flash-cooled in liquid N_2 from a cryoprotectant solution consisting of 0.5 *M* LiCl, 50 m*M* sodium citrate, 50 m*M* NaCl,

8% PEG 6000, 30% PEG 400 pH 4.0 and used for native data collection. Crystals from conditions (ii) and (iii) were flash-cooled after a 30 s soak in cryoprotectant solutions consisting of 50 m*M* bis-tris propane buffer, 100 m*M* NaCl, 8% PEG 3350 pH 5.5, 30% PEG 400 plus the addition of either 2 µ*M* SAN or 2 µ*M* AZM, respectively. All three crystals used in data collection crystallized in different space groups, which can be attributed to the presence of different additives, different pH values of the medium and different PEGs.

Data were collected on beamlines I03, I02 and I04-1 at the Diamond Synchrotron light source (Didcot, England) at 100 K in a stream of gaseous nitrogen using a PILATUS detector. Data were processed and scaled using *XDS* (Kabsch, 2010) and *AIMLESS* (Evans & Murshudov, 2013) in the *xia*2 (Winter *et al.*, 2013) pipeline. All further data and model manipulation was carried out using the *CCP*4 suite of programs (Winn *et al.*, 2011).

Phases for the native structure were determined using the molecular-replacement method as implemented in *MOLREP* (Vagin & Teplyakov, 2010) using *Neisseria gonorrhoeae* α-CA (NgCA) as a model (PDB entry 1kop; Huang *et al.*, 1998). The rotation function was calculated with an integration radius of 30 Å at a resolution of 3 Å using the monomeric 1kop model, which gave a strong peak of 7.95σ with background peaks being no greater than 4.12σ. The translation function was calculated at a resolution of 3 Å with the first monomer being positioned with a correlation coefficient of 0.34 and the second monomer with a value of 0.45. The inhibitor-bound complexes, which crystallized in different space groups, were solved using molecular replacement with the coordinates of the refined native structure as the model.

Electron-density maps were calculated and the structure was positioned to give the best fit to both the $2F_o - F_c$ and $F_o - F_c$ maps. Maximum-likelihood refinement was performed using *REFMAC*5 (Murshudov *et al.*, 2011) after each session of model building performed in *Coot* (Emsley *et al.*, 2010). The dictionary definitions for the ligands were created using *JLigand* (Lebedev *et al.*, 2012). The AZM dictionary definition was based on Cambridge Structural Database entry ATDZSA (Mathew & Palenik, 1974). Statistics of the data processing and the parameters of the final refined models are given in Table 1. After refinement, the quality of the model was checked using *PROCHECK* (Laskowski *et al.*, 1993). Images were created using the *PyMOL* molecular-graphics program (DeLano, 2002).

Table 1
Summary of the data-processing and refinement statistics.

Values in parentheses are for the outer resolution shell.

Crystal	Native TaCA	TaCA–AZM complex	TaCA–SAN complex
Beamline	I03	I04-1	I02
Wavelength (Å)	0.9763	0.9200	0.9795
Space group	$P4_32_12$	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	$a = b = 80.9$, $c = 154.6$	$a = 163.8$, $b = 283.0$, $c = 52.2$	$a = 83.1$, $b = 115.4$, $c = 65.5$
No. of chains per asymmetric unit	2	6	2
V_M (Å^3 Da^{-1})	2.3	3.66	2.85
Resolution range (Å)	80.9–1.69 (1.74–1.69)	42.0–1.85 (1.90–1.85)	57.7–1.55 (1.59–1.55)
Multiplicity	12.7 (12.3)	7.5 (7.4)	5.6 (2.7)
Unique reflections	57982	206954	86560
Completeness (%)	100 (100)	99.6 (99.6)	94.2 (67.4)
R_{merge}† (%)	7.0 (123.1)	10.3 (95.0)	8.5 (49.0)
$\langle I/\sigma(I)\rangle$	18.6 (2.3)	13.7 (2.4)	11.9 (2.1)
Wilson *B* factor‡ (Å^2)	35.8	33.2	24.7
Refined residues	450	1350	450
Refined water molecules	291	1917	694
Refined Zn atoms	2	6	2
Refined inhibitor molecules	N/A	6	2
R_{cryst}§ (%)	22.1	17.2	17.8
R_{free} (5% of total data) (%)	25.2	20.4	20.6
R.m.s.d., bond lengths¶ (Å)	0.007 [0.019]	0.008 [0.019]	0.009 [0.019]
R.m.s.d., bond angles¶ (°)	1.234 [1.953]	1.266 [1.987]	1.357 [1.985]
Occupancy of inhibitor	N/A	1.0	0.35
Average *B* factor (Å^2)			
Protein	36.2	26.2	18.7
Solvent	37.7	42.7	37.2
Zinc ions and inhibitors	26.4	29.6	14.5
Ramachandran plot analysis††, residues in (%)			
Most favoured regions	90.5	92.9	92.3
Additional allowed regions	8.7	6.6	7.1
Generously allowed regions	0.8	0.5	0.3
Disallowed regions	0	0	0.3

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum_{hkl}$ is the sum over all reflections and $\sum_{hkl}$ is the sum over i measurements of the reflection. § $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. ‡ The Wilson *B* factor was estimated by *SFCHECK* (Vaguine *et al.*, 1999). ¶ Target values are given in square brackets. †† Ramachandran plot analysis was performed by *PROCHECK* (Laskowski *et al.*, 1993).

2.8. Coordinates

The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entries 4c3t, 4uov and 4coq for the native, AZM-bound and SAN-bound structures, respectively.

3. Results and discussion

3.1. Cloning, overexpression and purification of TaCA

The search for a carbonic anhydrase gene within the *T. ammonificans* genome using BCA1 as the query sequence revealed a 247-amino-acid protein that had high sequence identity (52%) to NgCA. This sequence also shows similarity to the HCAII enzyme (38% identity) as well as the recently discovered thermophilic α-CAs SazCA (46% identity) and SspCA (45% identity).

The gene was subcloned into the pLATE51 vector containing a 6×His tag at the 5′-end and the protein was expressed using the ZYM-5052 medium method of protein auto-induction (Studier, 2005). The protein was purified using nickel and gel-filtration chromatography, yielding a homogenous protein sample. The structural studies revealed that the original TaCA sample has intersubunit disulfides, which

are required for tetramer formation, and reduced cysteines, which are able to form intrasubunit disulfides. Therefore, two additional samples of protein were prepared: rTaCA, which has all cysteines reduced by the addition of DTT, and oTaCA, with disulfides enforced by the presence of diamide during protein expression and purification.

The oligomeric state of the three TaCA protein samples was analysed by gel-filtration chromatography on a calibrated Superdex 200 column. The TaCA and oTaCA samples had the same elution profile on the size-exclusion column, which corresponded to a tetrameric species of approximately 110 kDa (Fig. 1*a*). The rTaCA sample had an elution profile corresponding to a monomeric molecular weight of approximately 30 kDa. On SDS–PAGE using reducing sample buffer, all of the TaCA samples ran at a monomeric molecular weight of 28 kDa as expected (Fig. 1*b*; lanes 2, 4 and 6). On SDS–PAGE using nonreducing sample buffer the rTaCA sample ran at the monomeric molecular weight and the oTaCA sample ran at the dimeric molecular weight (Fig. 1*b*; lanes 5 and 7). This shows that in the oTaCA sample the two monomers are held together by a covalent bond (the intermolecular disulfide) which is retained when the protein is unfolded under SDS buffer conditions. The SDS–PAGE run in nonreducing buffer showed that the TaCA was a mixed species of dimeric and monomeric molecular weights at a ratio of 4:1 (Fig. 1*b*; lane 3). The probability of a single disulfide reduction in this TaCA sample can be estimated to be around 20% based on the SDS–PAGE. As most of this TaCA sample migrates as a tetramer on the size-exclusion column (Fig. 1*a*), with no dimeric or monomeric molecular-weight peaks observed, we can assume that both intermolecular disulfides of the tetramer need to be reduced for its dissociation into the monomeric species. The probability of both disulfides in the tetramer being reduced is just 4%, which is consistent with the absence of a visible monomeric peak on the size-exclusion column.

The TaCA enzyme studied in this project was cloned and overexpressed as the full-length gene including a leader peptide which is proposed to be important for export into the periplasmic space. When NgCA was overexpressed in *E. coli* it was exported to the periplasmic space with the loss of the 26-amino-acid leader peptide (Chirică *et al.*, 1997). SspCA was overexpressed without the first 20 amino acids, which were thought to be a leader peptide (Di Fiore *et al.*, 2013). α-CA from *Ralstonia eutropha* (ReCA) has recently been reported to be located in the periplasm as detected by fluorescence of RFP-tagged ReCA (Gai *et al.*, 2014). The intact gene including the predicted signal sequence (*SignalP*; Petersen *et al.*, 2011) was used for overexpression of the His-tagged TaCA, which is found in the cytoplasm of *E. coli*.

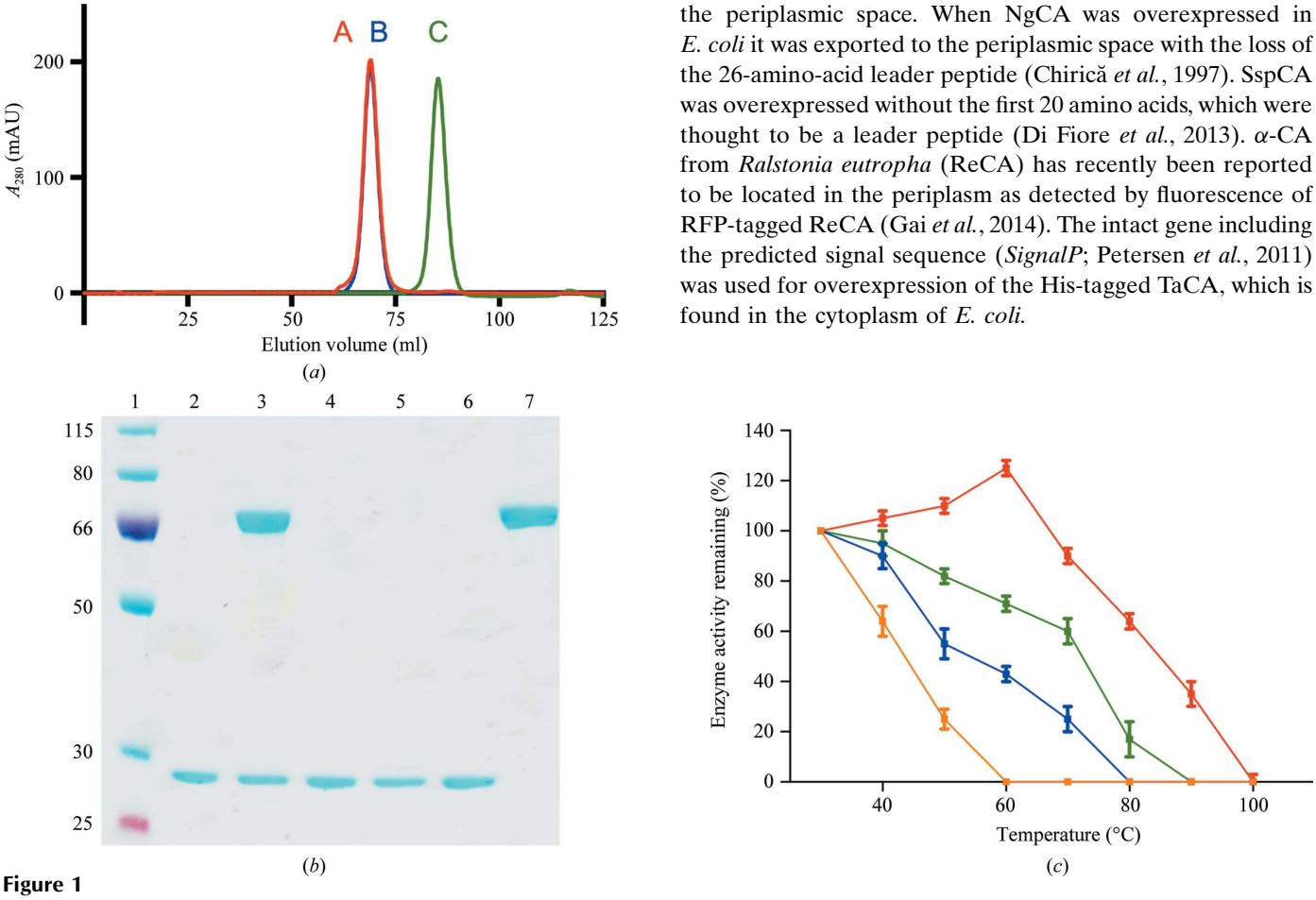


Figure 1
Biochemical characterization of the different forms of the TaCA enzyme. (*a*) The elution profiles from the calibrated Superdex 200 gel-filtration column for TaCA (red), oTaCA (blue) and rTaCA (green). Peaks A and B (∼69 ml; 110 kDa) correspond to the tetrameric form of TaCA; peak C (∼85 ml; 30 kDa) corresponds to the monomeric form of TaCA. (*b*) SDS–PAGE gel. Lane 1, molecular-weight markers labelled in kDa. Lane 2, TaCA (reduced sample buffer). Lane 3, TaCA (nonreduced sample buffer). Lane 4, rTaCA (reduced sample buffer). Lane 5, rTaCA (nonreduced sample buffer). Lane 6, oTaCA (reduced sample buffer). Lane 7, oTaCA (nonreduced sample buffer). (*c*) The percentage of enzyme activity for the different TaCA enzyme samples (TaCA, green; oTaCA, red; rTaCA, blue) remaining after incubation for 1 h at various temperatures. The increase in activity of oTaCA is thought to be owing to optimization of folding, since its expression was in a mesophilic host. The remaining enzyme activities after incubation at the same temperatures are also shown for BCA1 as a control (orange). Enzyme activity was calculated with reference to the activity of the enzyme incubated at 30°C prior to the assay.

3.2. Carbonic anhydrase activity and thermal stability of TaCA

Carbonic anhydrase activity was measured at room temperature using a variation of the method previously described by Sundaram *et al.* (1986). Using this method with CO_2 as the substrate the protein had a specific activity of 5236 MWA units per milligram of enzyme. This value is comparable to the specific activity of BCA1, which showed a specific activity of 5637 MWA units per milligram under the same assay conditions. The monomeric rTaCA had a lower specific activity of 3545 MWA units per milligram.

The TaCA enzyme showed a marked increase in stability at higher temperatures when compared with the BCA1 enzyme, retaining 60% of its activity after incubation at 70°C for 1 h (Fig. 1*c*). The rTaCA enzyme was found to be less stable, retaining 30% of its activity after incubation at 70°C for 1 h. This indicates that the formation of the disulfide core which holds together the TaCA tetramer contributes to its increased thermal stability. The oTaCA enzyme was found to be significantly more stable than the TaCA enzyme, retaining ∼90% of its activity after incubation at 70°C for 1 h. This result indicates that the formation of the conserved intramolecular disulfide bond between Cys47 and Cys202 within each monomer in oTaCA increases its thermal stability.

The kinetic parameters (Table 2) for the TaCA enzyme were determined at room temperature using the stopped-flow assay method (Khalifah & Edsall, 1972) with a resulting k_{cat} of $1.6 \times 10^6\ s^{-1}$. This value of k_{cat} is comparable to the corresponding values obtained for both HCAII ($1.40 \times 10^6\ s^{-1}$; Akdemir *et al.*, 2013) and the recently identified thermophilic enzymes SspCA ($9.35 \times 10^5\ s^{-1}$; Vullo *et al.*, 2013) and SazCA ($4.40 \times 10^6\ s^{-1}$; Akdemir *et al.*, 2013).

The TaCA enzyme was also tested for esterase activity using the substrate p-NpA since all previously characterized α-CAs have shown this activity (Vullo *et al.*, 2013). Neither TaCA nor rTaCA showed any detectable esterase activity with p-NpA, while a control reaction with BCA1 demonstrated high activity under the same assay conditions.

3.3. Quality of the models

The native TaCA enzyme crystallized in space group $P4_32_12$. The asymmetric unit contained two subunits of the biological tetramer which forms the intermolecular disulfide. This crystal form has a calculated solvent content of 46%. The observed dimer forms a biological tetramer with its crystallographic symmetry equivalent. Subunit *A* is involved in many crystal contacts and the electron density is good for residues of this subunit, with the average *B* factor for all atoms of subunit *A* being 26 Å^2. Subunit *B* is less restricted in movement since it is involved in fewer crystal contacts and is therefore less ordered, with an average *B* factor for all atoms reaching 48 Å^2. Poor electron density is observed for significant parts of subunit *B* and hindered modelling of the solvent in the native structure. The best strategy for the refinement of this structure was the imposition of local NCS restraints in *REFMAC*5 (Murshudov *et al.*, 2011).

Table 2
Comparison of the kinetic parameters of TaCA with three previously reported α-CAs.

Enzyme	k_{cat} (s^{-1})	K_m (m*M*)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)
TaCA	1.60×10^6	9.9	1.6×10^8
SazCA	4.40×10^6	12.5	3.5×10^8
SspCA	9.35×10^5	8.4	1.1×10^8
HCAII	1.40×10^6	9.3	1.5×10^8

The space group of the AZM inhibitor complex crystal was determined to be $P2_12_12$. The asymmetric unit contained six subunits forming the biological tetramer and a dimer which forms a tetramer with its symmetry equivalent. This crystal form has a calculated solvent content of 66%. The space group of the SAN inhibitor complex crystal was determined to be $P2_12_12$ and its asymmetric unit contained two subunits of the biological tetramer, with a calculated solvent content of 57%. The structures of both the AZM and SAN inhibitor complexes were refined with local NCS restraints.

The different crystal structures of the TaCA enzyme were refined to a resolution of 1.85 Å or better with a final R_{free} of 0.25 or better (Table 1). Owing to disorder, residues belonging to the leader peptide 1–22 were not built in any of the structures. The *G*-factors calculated for each model confirmed their good stereochemical properties. The electron densities for the inhibitor active-site complexes with both AZM and SAN were confirmed by calculating $F_o - F_c$ OMIT maps. The AZM inhibitor was modelled at full occupancy as judged by the *B* factors and the electron-density maps. The SAN inhibitor was modelled with a lower occupancy of 0.35 owing to disorder of the aromatic group.

The proline residues 55, 116 and 201 are in the *cis* conformation in all monomers of the different TaCA structures. The two cysteine residues, Cys47 and Cys202, have been modelled in two conformations, as an intrasubunit disulfide and as nonbonded free thiols, since the expression conditions were not sufficiently oxidative for full occupancy of the disulfide bond. These disulfides were modelled with low occupancies of 0.2–0.3 in the native and the SAN complex structures. There are no intrasubunit disulfides in the AZM complex structure. One cysteine residue of each subunit, Cys67, forms an intermolecular disulfide bond with the most remote subunit of the tetramer. This was modelled at full occupancy in the native and the AZM complex structures, while it has 80% occupancy in the SAN complex structure. The catalytic zinc has full occupancy in both complex structures and in subunit *A* of the native structure. In subunit *B* of the native structure it has an occupancy of 0.7.

The structure of the SAN complex contains a long PEG molecule with six repeats of the ethylene glycol unit. The structure of the AZM complex contains a molecule of bis-tris propane buffer bound to several monomers and multiple PEG molecules, the longest of which contained six ethylene glycol repeats. Unexplained electron density present on some of the crystal contacts close to the long PEG molecule was modelled as L-(+)-tartaric acid, which may have come from the crystallization condition.

Many residue side chains were modelled with alternative conformations and for some residues second positions of the main-chain O atom were also modelled. The Ramachandran plot for the native structure revealed that 90% of the residues are in the most favoured regions and a further 9% are in additionally allowed regions. The inhibitor-bound structures have similar statistics.

3.4. The TaCA monomer structure

The monomer of TaCA is built around a central ten-stranded mixed-type pleated β-sheet with strand direction + + − + + − + − + − and connectivity +8x, −1, −1, −1, −1, +5, −7x, +1, −2x (Richardson, 1981). A small two-stranded parallel β-sheet is formed by residues 57–58 and 120–121. Two α-helices are positioned on one side of the central β-sheet towards the N-terminus of the subunit, with an additional two larger helices on the other side of the sheet (Fig. 2).

The TaCA enzyme has two cysteine residues, Cys47 and Cys202, that are capable of forming an intramolecular disulfide bond under oxidizing conditions. This disulfide is only partially formed in the TaCA structures. These two cysteines are also observed in NgCA and SspCA and in some isoforms of mammalian and algal α-CAs (Huang *et al.*, 1998), but they are not conserved in HCAII. This intramolecular disulfide bond connects the N-terminal region of the monomer to a loop holding a number of active-site residues. It is fully oxidized in the NgCA and SspCA structures, even though the purification and crystallization buffers used for NgCA contained 2 m*M* mercaptoethanol (Huang *et al.*, 1998).

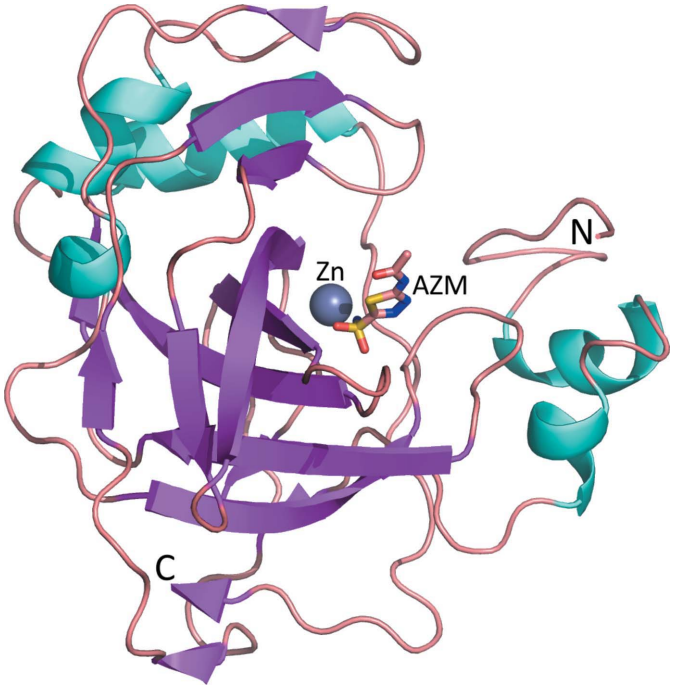


Figure 2
The folding of the TaCA monomer is shown as a cartoon model with the N- and C-termini shown. The α-helices are shown in cyan, the β-strands in magenta and the loops in purple. The catalytic zinc ion is shown as a sphere and the inhibitor AZM as a stick model. Figs. 2–5 were generated in *PyMOL* (DeLano, 2002).

3.5. The quaternary structure of TaCA

Size-exclusion chromatography suggested that TaCA and oTaCA both have a tetrameric quaternary structure, while rTaCA elutes from a gel-filtration column as a monomer. The tetrameric quaternary structure has been confirmed in the structures of TaCA obtained from three different crystal forms, making TaCA the first reported tetrameric α-CA enzyme.

In comparison, most mammalian α-CAs are monomeric proteins. SspCA forms a dimer which is quite similar to that

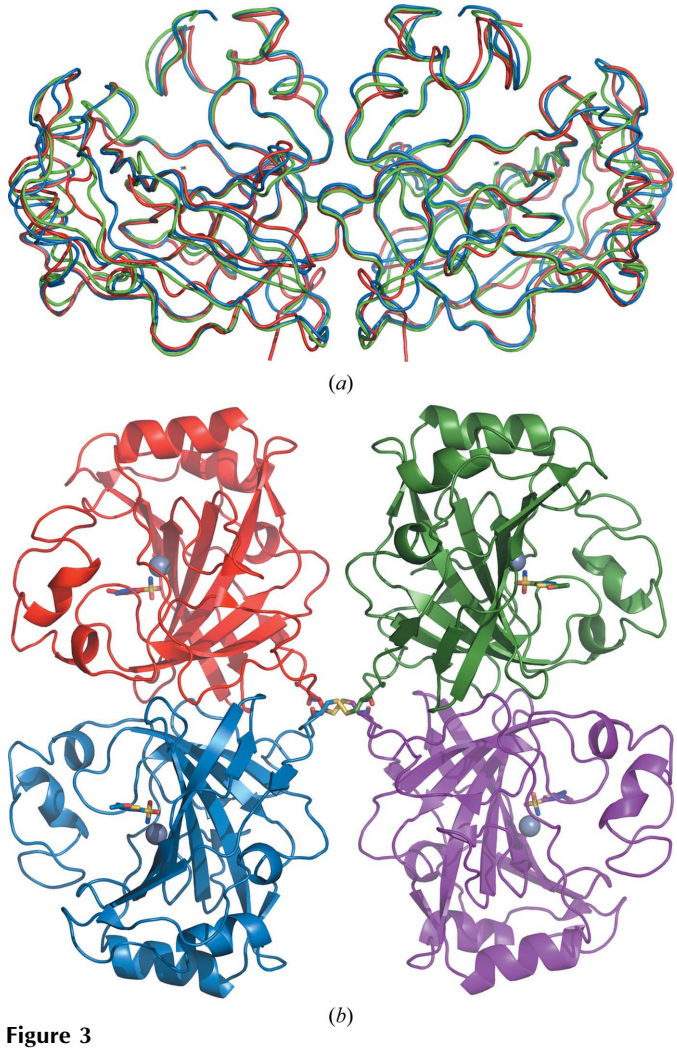


Figure 3
(*a*) Ribbon representations of superimposed dimeric molecules of SspCA (red), NgCA (green) and the 'bacterial dimer' of TaCA (green) are shown viewed normally to the molecular dyad. (*b*) A cartoon diagram showing the overall folding of the noncrystallographic TaCA enzyme tetramer from the AZM complex structure viewed approximately along the molecular dyad. The individual monomers are shown in different colours with the inhibitor AZM molecules displayed as stick models. The Cys67 residues of each monomer which form the intermolecular disulfides are displayed as stick models and the catalytic zinc ions are shown as spheres. The distance between the green and the red monomers is smaller than that between the blue and the magenta monomers. This feature underlines the breakdown of the molecular 222 point-group symmetry owing to the crystal packing. The colour of the C atoms in the stick models matches the colour of the monomer that they belong to. This demonstrates that each monomer forms an intermolecular disulfide with the most distant monomer: red with magenta and green with blue.

observed in the NgCA structure (Di Fiore *et al.*, 2013). This led to the proposal (Di Fiore *et al.*, 2013) that all bacterial α-CAs form a conserved dimer. Such a dimer has been observed in the TaCA structure and will henceforth be called a 'bacterial dimer' (Fig. 3*a*). The formation of the 'bacterial dimer' of the TaCA buries around 900 Å^2 of solvent-accessible area, which is about 9% of the total monomer surface area. Residues 65–67 from two subunits of the 'bacterial dimer' form an antiparallel two-stranded β-sheet. The TaCA tetramer can be described as a dimer of two 'bacterial dimers' (Fig. 3*b*) held together by an unusual novel core built around intersubunit disulfides in the centre of the molecule. The 'bacterial dimers' of the three different TaCA structures are quite similar.

3.6. The TaCA tetramer core

Cys67 of each monomer of the 'bacterial dimer' forms an intermolecular disulfide with the most remote monomer of the adjacent 'bacterial dimer' (Fig. 4*a*). The S$^\gamma$ atoms of the two thus formed intermolecular disulfides are within 7 Å distance and are located in the centre of the tetramer. The interaction between two different 'bacterial dimers' is complemented by an intersubunit ion pair formed between the side chain of Lys65 of each subunit and the carboxyl group of the C-terminal Lys247 belonging to a different subunit of the adjacent 'bacterial dimer'. This limits the interactions between the catalytic dimers to a small core formed by disulfides and ionic interactions in the centre of the tetramer. No other interactions exist between the 'bacterial dimers' of TaCA, and the reduction of the intermolecular disulfide results in tetramer breakdown. Around 250 Å^2 (less than 3%) of the total subunit solvent-accessible surface area is buried upon TaCA tetramer formation.

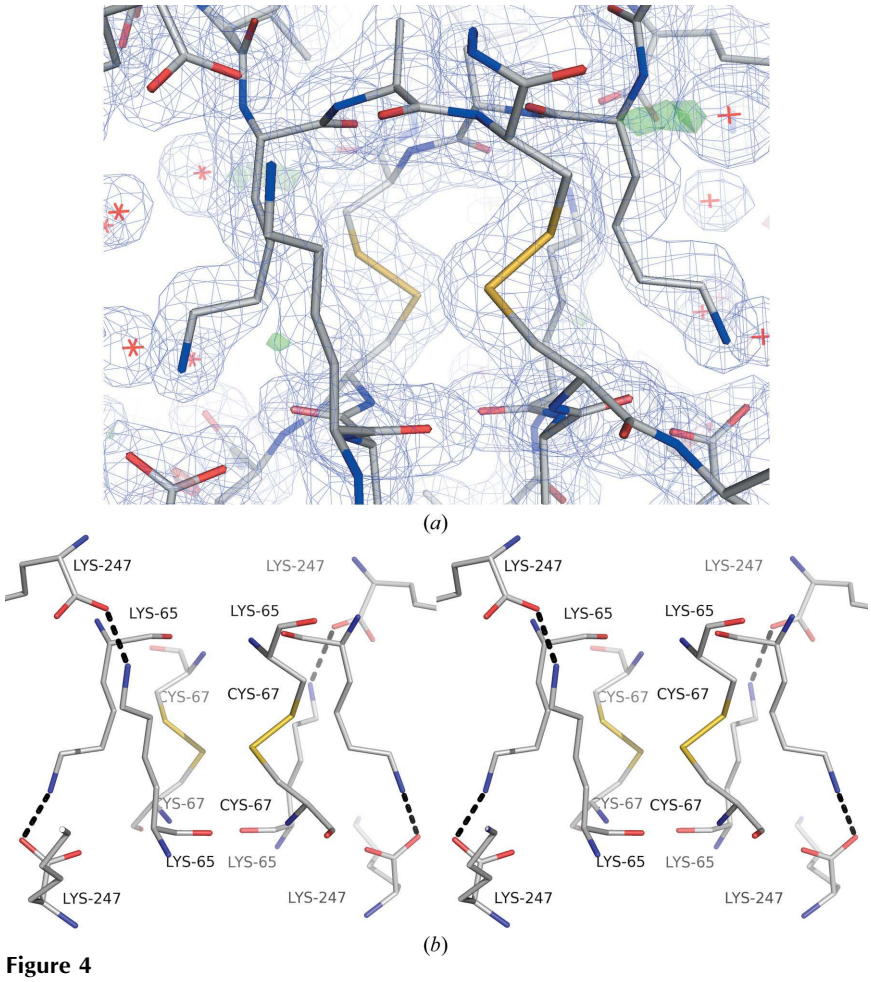


Figure 4
(*a*) Electron-density maps for the native structure of TaCA, showing the tetramer core formed by the side chains of Lys65 of each monomer and the intermolecular disulfides which link the Cys67 residues of the opposite monomers within the tetramer. The $2F_o - F_c$ map (blue) is contoured at 1.0σ and the $F_o - F_c$ map is contoured at 3.0σ (green) and -3.0σ (red). The amino-acid residues are shown as stick models. The solvent molecules which are excluded from the inside of the core are shown as red stars. (*b*) A stereo diagram showing the interactions in the core of the TaCA tetramer. The intersubunit disulfides formed by the Cys67 residues of the opposite monomers are surrounded by the Lys65 residues from each monomer to form an ion pair with the carboxyl group of the C-terminal Lys247 residues from adjacent monomers. The amino-acid residues are shown as stick models.

The four Lys65 side chains, one from each subunit, are found in an extended conformation and together with the two intersubunit disulfides form a core with a nearly regular hexagonal cross-section (Fig. 4*b*). The distances between the C$^\beta$ atoms of the Lys65 residues from subunits *B* and *D* which form the 'bacterial dimer' are 9.7 Å, which is similar to the distances between the C$^\beta$ atoms of the Cys67 residues from these subunits, which are involved in different intersubunit disulfides (9.3 Å). Solvent molecules are excluded from the inside of the core and the S$^\gamma$ atoms of the cysteine residues point inside the core. Such an arrangement of the core appears to hinder the interaction of reducing agents, such as thioredoxin or gluthathione, with the disulfides maintaining their oxidized state in the reducing environment of the bacterial cell.

As the interaction between two 'bacterial dimers' is limited to a small core in the centre of the tetramer, one would expect a significant degree of rotation between the dimers in the different crystal forms obtained. Indeed, when one 'bacterial dimer' of the native TaCA tetramer is superimposed onto the dimer of the AZM complex, the other 'bacterial dimer' is rotated around the molecular dyad by about 7° in relation to the second dimer of the complex, with displacement of some C$^\alpha$ positions by up to 5 Å. This rotation is relatively small and the tetramers of the two complexes are similar. However, relative movement of 'catalytic dimers' is not limited to rotation around the molecular dyad. A more complex twist, resulting in a breakdown of the 222 point-group symmetry of the TaCA noncrystallographic tetramer, is observed in the AZM complex (Fig. 4*b*), where the distances between pairs of

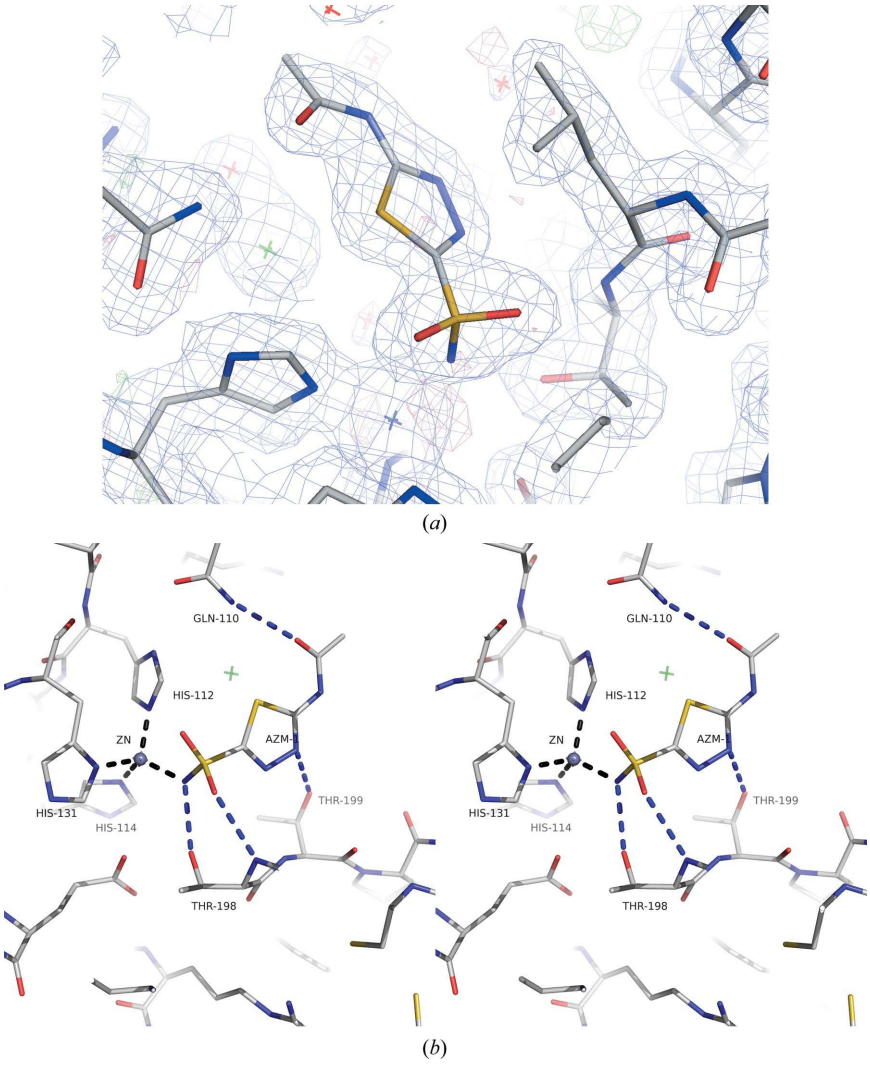


Figure 5
(*a*) Electron-density maps of the active site for the TaCA–AZM inhibitor complex. The $2F_o - F_c$ map (blue) is contoured at 1.1σ and the $F_o - F_c$ map is contoured at 3.0σ (green) and -3.0σ (red). The neighbouring residues and the AZM molecule are shown as stick models and the zinc ion (grey) and the solvent molecules (green for Cl^- ion, red for waters) are shown as stars. (*b*) Interactions in the active site of the AZM complex of TaCA. The neighbouring residues and the AZM molecule are shown as stick models and the zinc ion is shown as a grey sphere. Hydrogen bonds are shown as blue dashed lines and metal coordination as black dashed lines.

equivalent C^{α} atoms in the two 'bacterial dimers', which should be nearly equal, differ by up to 5 Å.

3.7. Structural features responsible for the thermostability of TaCA

The TaCA enzyme has relatively few ion pairs in comparison to SspCA, where ionic networks were reported to be essential for thermostability (Di Fiore *et al.*, 2013). The pair of disulfide-linked 'bacterial dimers' associated as a tetrameric complex is a novel feature of TaCA which has not been observed previously for any other α-CA. The tetramer association involves formation of four ion pairs and two disulfide bonds. This appears to be the main means of thermal adaptation of the TaCA enzyme in this thermophilic bacterium. Increased oligomerization has been observed for other thermostable enzymes; for example, methionine aminopeptidase, which is a monomer in *E. coli*, forms a dimer in the hyperthermophile *Pyrococcus furiosus* (growth temperature 100°C; Tahirov *et al.*, 1998). The formation of intersubunit disulfides to aid stability has also been observed in the pyrrolidone carboxyl peptidase enzyme from the thermophilic archaeon *Thermococcus litoralis* (Singleton *et al.*, 1999). The monomeric rTaCA enzyme has been shown to be less thermostable when compared with the tetrameric native TaCA (Fig. 1*c*); this suggests a link between the oligomeric state of the enzyme and its thermal stability. Although disulfide bonds were not thought to exist within a cellular environment, it has now become clear from crystallographic (Guy *et al.*, 2003) and bioinformatic studies (Mallick *et al.*, 2002) that they do exist in the aerobic archaea *Aeropyrum pernix* and *Pyrobaculum aerophilum*.

When the TaCA protein was overexpressed in the presence of diamide, this forced the formation of the disulfide bonds, including the intrasubunit ones (oTaCA). The oTaCA enzyme has been shown to be more thermostable than native TaCA (Fig. 1*c*). The reported thermostabilities of the β-CA enzyme from the thermophilic archaeon *Methanobacterium thermoautotrophicum* (Smith & Ferry, 1999) and the thermophilic bacterial SspCA (Vullo *et al.*, 2013) are similar to that of the oTaCA enzyme.

3.8. The active site

The active site of TaCA is highly conserved compared with that found in other α-CA enzymes. The active-site zinc residue is held in place by three histidine residues (His112, His113 and His131). The hydrophobic CO_2-binding pocket (Val133, Val143, Leu197, Val206 and Trp209) is well conserved, as are the hydrophilic proton-shuttle residues (Tyr28, Asn85, Lys90, Thr198 and Thr199), except for a single-residue substitution at position Lys90 (TaCA numbering) which is a conserved Asn in other α-CAs. This, however, would not hinder proton transfer between the active site and the external buffer. All CAs possess a histidine residue (His87 in TaCA) that is involved in the movement of protons to and from the active site. This residue can be found in two conformations referred to as 'in' and 'out' (Domsic *et al.*, 2010). In the native and SAN inhibitor structures, His87 was modelled in the 'out' conformation, facing away from the active site. In the AZM-bound structure

His87 shows a mixed species with most of the His (occupancy 0.6) found in the 'in' conformation and with a smaller proportion in the 'out' conformation. This is consistent with the protein-shuttle system previously described in other CA structures (Domsic *et al.*, 2008). There are a number of water molecules observed in the structure, with a high-order hydrogen-bonding network observed between the zinc-bound water molecule and the proton-shuttle residue His87 in the active site.

3.9. The inhibitor binding

The binding interactions of the inhibitor AZM in the TaCA active site are similar to those reported for the NgCA and SspCA enzymes (Huang *et al.*, 1998; Di Fiore *et al.*, 2013), with the inhibitor sulfonamide group displacing the zinc-bound hydroxyl ion and a further two water molecules in the active site (Fig. 5*a*). The AZM sits in a pocket formed by Gln110, His112, His114, His131, Leu197, Thr198 and Thr199, which are all conserved in the NgCA enzyme. The sulfonamide group interacts with a zinc ion and the main-chain N atom and hydroxyl O atom of Thr198. Hydrogen bonds are observed between the N atom of the AZM ring and the hydroxyl O atom of Thr198, the acetylamido O atom of AZM and Gln110 (Fig. 5*b*).

The inhibitor SAN binds with its sulfonamide group binding in a similar conformation to that of AZM. SAN was modelled with a partial occupancy of 0.35 in both subunits of the complex structure, mainly owing to disorder of the *p*-nitrophenyl group, as the sulfonamide group appeared to have higher occupancy. The disorder seems to be owing to steric hindrance to the binding of the SAN aminobenzene ring and its inability to form hydrogen bonds to the side chains of the active-site residues; such hydrogen bonds are formed upon the binding of the inhibitor AZM.

3.10. Comparison of TaCA with other carbonic anhydrases

As bacterial and mammalian α-CAs are significantly different and have been thoroughly compared previously (Huang *et al.*, 1998), the TaCA comparison will be limited to other bacterial CAs. The monomers of TaCA and NgCA superimpose closely with an r.m.s.d. of <0.81 Å for 221 C^{α} atoms. Superimposition of the TaCA enzyme with SspCA (PDB entry 4g7a; Di Fiore *et al.*, 2013) across all 220 C^{α} residues gives an estimated r.m.s.d. of 0.95 Å, with all key active-site residues being highly conserved. As most bacterial α-CAs appear to be exported into the periplasmic space, the intrasubunit disulfide is highly conserved within this group, as it makes an important contribution to stability. The structure of TaCA is consistent with the idea of a conserved dimer in bacterial α-CAs; however, while NgCA has hydrogen bonding consistent with an antiparallel β-sheet observed in TaCA on the subunit interface, SspCA does not form this β-sheet. The C-terminus is two residues longer in SspCA than in TaCA, which has the same position of the C-terminus as NgCA. Lys65 has a conserved equivalent in SspCA but not in NgCA. A *BLASTP* search (Altschul *et al.*, 1990) against a nonredundant protein-sequence database revealed that even the closest TaCA homologues (the α-CAs from *Persephonella marina*, *Caminibacter mediatlanticus* and *Thiocystis violascens*) do not have a cysteine residue in a position equivalent to Cys67 in TaCA. This implies that the tetrameric structure observed in TaCA is unique since it depends on this cysteine residue to form the stabilizing disulfide.

3.11. Structural basis for the absence of p-NpA activity of the TaCA enzyme

All α-CAs reported to date have significant esterase activity towards the substrate p-NpA (Höst *et al.*, 2006; Supuran, 2011) although variations in this secondary activity have been reported for the SazCA enzyme, which has a fivefold lower activity than the SspCA enzyme and a 20-fold lower activity than BCAII (Akdemir *et al.*, 2013). The complete absence of p-NpA esterase activity observed for the TaCA enzyme is novel. This seems surprising as the active-site residues are very conserved within the α-CA family. The absence of esterase activity does not seem to be related to the quaternary structure of TaCA, as monomeric rTaCA does not display any esterase activity.

The binding of the p-NpA molecule in the TaCA active site was modelled with an orientation of the nitrophenyl group consistent with that of aminobenzene ring in the SAN inhibitor complex. For the esterase reaction to proceed, the acetyl group of p-NpA has to be positioned close to the hydrophobic patch in the TaCA active site, where it would have an unfavourable steric clash with the side chains of Val143 and Trp208. The thermal movement of the protein scaffold could cause minor displacements of these side chains, allowing the p-NpA to be positioned for catalysis. Different rigidity of the protein scaffolds may be responsible for the reported variations in esterase activity. In the TaCA structure there is an ion-pair 'latch' formed by residues Arg207, Asp102 and Arg192 on the opposite side of the β-sheet to the active site. This appears to stabilize the secondary structure of the TaCA monomer in the region of the hydrophobic patch, thereby preventing movement of the scaffold to allow hydrolysis of p-NpA. This ionic network is not observed in the related α-CAs and is proposed to be responsible for the absence of p-NpA esterase activity of TaCA.

4. Conclusions

The TaCA enzyme has been shown to be highly thermostable, while its activity is comparable to that of the BCA1 enzyme at room temperature. The characterization of TaCA suggests that its thermostability is dependent on the formation of intrasubunit and intersubunit disulfide bonds. The unique TaCA tetrameric structure is held together by the formation of an unusual core assembly formed by two intersubunit disulfides and four lysine residues from each subunit which are involved in intersubunit ionic interactions. The structure of this core in the centre of the tetramer prevents the reduction of the intersubunit disulfides. The conserved intrasubunit

disulfides contribute to the stability of the enzyme in the environment of the periplasm in the natural *Thermovibrio* host. This intrasubunit disulfide is only partially formed in the cytosol of the *E. coli* expression host. The oTaCA enzyme, which has both intrasubunit and intersubunit disulfides intact and is proposed to be similar to the state of TaCA *in vivo,* shows high thermostability. The monomeric rTaCA enzyme which has all of the disulfide bonds broken shows a significant reduction in thermal stability.

Unlike other characterized α-CAs, the TaCA enzyme does not display esterase activity towards p-NpA. In the TaCA structure described in this paper there is an ion-pair 'latch' formed by residues Arg207, Asp102 and Arg192 on the opposite side of the β-sheet to the active site. This ionic network is not observed in the related α-CAs. In TaCA it is proposed that this 'latch' stabilizes the secondary structure of the monomer and restricts the movement of the protein in the region of the side chains of Val143 and Trp208, preventing the binding of the p-NpA substrate in a suitable position for catalysis.

This study has shown that the TaCA enzyme has a higher oligomeric state than previously reported α-CAs and shows novel features of thermostable adaptation which differ significantly from those described for SspCA, the other thermophilic bacterial α-CA with a known structure. The thermostability and high activity of TaCA make it an excellent candidate for industrial applications for CO_2 capture.

The authors would like to thank the Diamond Light Source for access to beamlines I02, I03 and I04-1 (proposal No. MX6851) and the beamline staff scientists. This work was supported by funding from Statoil ASA. MI is grateful to the University of Exeter and for the BBSRC-funded ERA-IB grant BB/L002035/1. The Wellcome Trust, EU, BBSRC and EPSRC supported funding of JAL's laboratory.

References

Akdemir, A., Vullo, D., De Luca, V., Scozzafava, A., Carginale, V., Rossi, M., Supuran, C. T. & Capasso, C. (2013). *Bioorg. Med. Chem. Lett.* **23**, 1087–1090.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Andersen, J. F., Sanders, D. A., Gasdaska, J. R., Weichsel, A., Powis, G. & Montfort, W. R. (1997). *Biochemistry*, **36**, 13979–13988.

Armstrong, J. M., Myers, D. V., Verpoorte, J. A. & Edsall, J. T. (1966). *J. Biol. Chem.* **241**, 5137–5149.

Chirică, L. C., Elleby, B., Jonsson, B. H. & Lindskog, S. (1997). *Eur. J. Biochem.* **244**, 755–760.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

De Luca, V., Vullo, D., Scozzafava, A., Carginale, V., Rossi, M., Supuran, C. T. & Capasso, C. (2013). *Bioorg. Med. Chem.* **21**, 1465–1469.

Di Fiore, A., Capasso, C., De Luca, V., Monti, S. M., Carginale, V., Supuran, C. T., Scozzafava, A., Pedone, C., Rossi, M. & De Simone, G. (2013). *Acta Cryst.* D**69**, 1150–1159.

Domsic, J. F., Avvaru, B. S., Kim, C. U., Gruner, S. M., Agbandje-McKenna, M., Silverman, D. N. & McKenna, R. (2008). *J. Biol. Chem.* **283**, 30766–30771.

Domsic, J. F. & McKenna, R. (2010). *Biochim. Biophys. Acta*, **1804**, 326–331.

Domsic, J. F., Williams, W., Fisher, S. Z., Tu, C., Agbandje-McKenna, M., Silverman, D. N. & McKenna, R. (2010). *Biochemistry*, **49**, 6394–6399.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Fath, S., Bauer, A. P., Liss, M., Spriestersbach, A., Maertens, B., Hahn, P., Ludwig, C., Schäfer, F., Graf, M. & Wagner, R. (2011). *PLoS One*, **6**, e17596.

Fisher, Z., Boone, C. D., Biswas, S. M., Venkatakrishnan, B., Aggarwal, M., Tu, C., Agbandje-McKenna, M., Silverman, D. & McKenna, R. (2012). *Protein Eng. Des. Sel.* **25**, 347–355.

Gai, C. S., Lu, J., Brigham, C. J., Bernardi, A. C. & Sinskey, A. J. (2014). *AMB Express*, **4**, 2.

Guy, J. E., Isupov, M. N. & Littlechild, J. A. (2003). *Acta Cryst.* D**59**, 174–176.

Höst, G., Mårtensson, L. G. & Jonsson, B. H. (2006). *Biochim. Biophys. Acta*, **1764**, 1601–1606.

Huang, S., Xue, Y., Sauer-Eriksson, E., Chirica, L., Lindskog, S. & Jonsson, B. H. (1998). *J. Mol. Biol.* **283**, 301–310.

James, P. (2010). PhD thesis. Exeter University, England.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Khalifah, R. G. & Edsall, J. T. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 172–176.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. & Murshudov, G. N. (2012). *Acta Cryst.* D**68**, 431–440.

Liljas, A., Kannan, K. K., Bergstén, P. C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S. & Petef, M. (1972). *Nature New Biol.* **235**, 131–137.

Lindskog, S. (1997). *Pharmacol. Ther.* **74**, 1–20.

Lloyd, M. D., Pederick, R. L., Natesh, R., Woo, L. W. L., Purohit, A., Reed, M. J., Acharya, K. R. & Potter, B. V. L. (2005). *Biochem. J.* **385**, 715–720.

Mallick, P., Boutz, D. R., Eisenberg, D. & Yeates, T. O. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 9679–9684.

Mathew, M. & Palenik, G. J. (1974). *J. Chem. Soc. Perkin Trans. 2*, pp. 532–536.

Maupin, C. M. & Voth, G. A. (2007). *Biochemistry*, **46**, 2938–2947.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.

Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.

Savile, C. K. & Lalonde, J. J. (2011). *Curr. Opin. Biotechnol.* **22**, 818–823.

Silverman, D. N. & Lindskog, S. (1988). *Acc. Chem. Res.* **21**, 30–36.

Singleton, M., Isupov, M. & Littlechild, J. (1999). *Structure*, **7**, 237–244.

Smith, K. S. & Ferry, J. G. (1999). *J. Bacteriol.* **181**, 6247–6253.

Smith, K. S., Jakubzick, C., Whittam, T. S. & Ferry, J. G. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 15184–15189.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Sundaram, V., Rumbolo, P., Grubb, J., Strisciuglio, P. & Sly, W. S. (1986). *Am. J. Hum. Genet.* **38**, 125–136.

Supuran, C. T. (2011). *Front. Pharmacol.* **2**, 34.

Supuran, C. T., Scozzafava, A. & Casini, A. (2003). *Med. Res. Rev.* **23**, 146–189.

Tahirov, T. H., Oki, H., Tsukihara, T., Ogasahara, K., Yutani, K., Ogata, K., Izu, Y., Tsunasawa, S. & Kato, I. (1998). *J. Mol. Biol.* **284**, 101–124.

Tripp, B. C., Smith, K. & Ferry, J. G. (2001). *J. Biol. Chem.* **276**, 48615–48618.
Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B. H. & Lindskog, S. (1989). *Biochemistry*, **28**, 7913–7918.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Vetriani, C., Speck, M. D., Ellor, S. V., Lutz, R. A. & Starovoytov, V. (2004). *Int. J. Syst. Evol. Microbiol.* **54**, 175–181.
Vullo, D., De Luca, V., Scozzafava, A., Carginale, V., Rossi, M., Supuran, C. T. & Capasso, C. (2013). *Bioorg. Med. Chem.* **21**, 1534–1538.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* D**69**, 1260–1273.