Natalia Díaz Torres,^a Guillermo González,^a Shyamasri Biswas,^a Kathleen M. Scott^b and Robert McKenna^{a*}

^aDepartment of Biochemistry and Molecular Biology, University of Florida, PO Box 100245, Gainesville, FL 32610, USA, and ^bDepartment of Integrated Biology, University of South Florida, Tampa, FL 33620, USA

Correspondence e-mail: rmckenna@ufl.edu

Received 4 May 2012

Accepted 7 July 2012

Preliminary X-ray crystallographic analysis of α -carbonic anhydrase from *Thiomicrospira crunogena* XCL-2

Thiomicrospira crunogena XCL-2 is a novel sulfur-oxidizing chemolithoautotroph that plays a significant role in the sustainability of deep-sea hydrothermal vent communities. This recently discovered gammaproteobacterium encodes and expresses four carbonic anhydrases (CAs) from three evolutionarily and structurally distinct CA families: an α -CA, two β -CAs and a γ -CA. In order to characterize and elucidate the physiological roles of these CAs, X-ray crystallographic structural studies have been initiated on the α -CA. The α -CA crystallized in space group *C2*. The crystals diffracted to a maximum resolution of 2.6 Å, with unit-cell parameters $a = 127.1$, $b = 102.2$, $c = 105.0$ Å, $\beta = 127.3^\circ$, and a calculated Matthews coefficient of $2.04 \text{ Å}^3 \text{ Da}^{-1}$ with four identical protein molecules in the crystallographic asymmetric unit. A preliminary solution was determined by molecular replacement with the *PHENIX AutoMR* wizard, which had an initial TFZ score of 17.9. Refinement of the structure is currently in progress.

1. Introduction

Carbonic anhydrases (CAs) are mainly zinc metalloenzymes that catalyze the interconversion of carbon dioxide and water to bicarbonate and protons (Duda & McKenna, 2006; Domsic *et al.*, 2008). They are found in virtually all living organisms and have been shown to be involved in many important and complex physiological processes, such as respiration, pH regulation in tissues and secretory processes including the formation of cerebrospinal fluid and aqueous humor in animals (Henry & Swenson, 2000; Maren, 1984; Cheg-widden *et al.*, 2000). However, CAs have also been found to play important roles in the physiological processes of more primitive organisms, including photosynthesis, CO₂ fixation and transport in plants and bacteria, making them essential to most living organisms (Shively *et al.*, 1998; Badger & Price, 1994). There are three evolutionarily and structurally distinct families of CAs: α -CAs (mostly expressed in mammals), β -CAs (found in plants, prokaryotes and fungi) and γ -CAs (expressed only in methane-producing bacteria and diatoms) (Hewett-Emmett & Tashian, 1996). δ -CA and ε -CA classes which lack significant sequence similarity to existing CAs have also been suggested, but structural studies have shown that they are distantly related to the α -CA and β -CA classes, respectively (Lane & Morel, 2000; Sawaya *et al.*, 2006).

Genomic studies of the novel deep-sea vent gammaproteobacterium *Thiomicrospira crunogena* XCL-2 revealed the coding and expression of four different CAs which span the three evolutionarily and structurally distinct CA families (Scott *et al.*, 2006; Dobrinski *et al.*, 2010). *T. crunogena* is a sulfur-oxidizing chemolithoautotroph that plays a significant role in the sustainability of deep-sea hydrothermal vent communities. Chemolithoautotrophs are defined as bacterial organisms that obtain energy from the oxidation of inorganic compounds to carry out CO₂ fixation (Dobrinski *et al.*, 2010; Khalifah *et al.*, 1977; Domsic *et al.*, 2008). Some of these bacteria, including *T. crunogena*, live in habitats in which nutrients such as dissolved inorganic carbon (DIC) are temporally and spatially limited. In addition, their environment is constantly changing, as dilute hydrothermal fluid (warm, anoxic, pH 5–8, [DIC] = 2–7 mM) interacts with bottom water (275 K, oxic, pH 8, [DIC] = 20 μ M), leading to oscillatory habitat chemistry dominance (Scott *et al.*, 2006). Interestingly

© 2012 International Union of Crystallography
All rights reserved

enough, many of these bacteria have developed a wide array of adaptations, such as carbon-concentrating mechanisms (CCMs) and the formation of microbial mats or veils to produce a stable micro-environment, which allow them to grow and thrive in these nutrient-limited habitats (Shively *et al.*, 1998). *T. crunogena*, for example, is equipped with a CCM which allows it to grow rapidly during DIC scarcity (Dobrinski *et al.*, 2005). Whole-cell studies using radio-labeled CO₂ and bicarbonate have shown that *T. crunogena* can use extracellular CO₂ and bicarbonate to increase the intracellular DIC concentration, suggesting that active transport of bicarbonate or carbon dioxide plays a role in the CCM of *T. crunogena* (Dobrinski *et al.*, 2005). Therefore, the expression of four different CAs suggests a role in bicarbonate and carbon dioxide transport as part of the CCM. To gain insight into the roles of these four CAs in the CCM, we have started by expressing, purifying and crystallizing the α -CA expressed in *T. crunogena* XCL-2.

2. Materials and methods

The coding sequence for the α -CA was obtained from the NCBI Reference Sequence collection (NCBI RefSeq YP_391811.2). The full-length sequence of the α -CA consists of 315 amino-acid residues, including a signal peptide at the N-terminus as indicated by the *SignalP* software (Petersen *et al.*, 2011) that indicates a periplasmic location for the α -CA. A pET SUMO plasmid construct containing the *T. crunogena* α -CA gene was provided by KMS. PCR and cloning protocols have been reported previously (Dobrinski *et al.*, 2010). This plasmid was transformed into *Escherichia coli* BL21 (DE3) cells for high-yield protein expression. Cultures were incubated at 310 K in 2 $\times$ YT broth containing 50 μ g ml⁻¹ kanamycin at 200 rev min⁻¹ until an OD₆₀₀ of 0.6–1.0 was attained, at which point protein expression was induced for 18 h at 291 K by the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. The cells were harvested at 277 K by centrifugation at 4000 rev min⁻¹ for 10 min and were cooled to 253 K. The cell pellets were

resuspended in lysis buffer (0.1 M Tris-HCl pH 9.0, 0.2 M Na₂SO₄) with 1 mg ml⁻¹ lysozyme and 50 μ g ml⁻¹ DNase I (Invitrogen) and were lysed at 277 K with stirring overnight. Samples were then centrifuged for 70 min at 15 000 rev min⁻¹ (Dobrinski *et al.*, 2010; Domsic *et al.*, 2008).

The cell lysate was loaded onto a *p*-aminobenzenesulfonamide (*p*-AMBS, a carbonic anhydrase inhibitor) affinity column equilibrated with ten column volumes of lysis buffer (Khalifah *et al.*, 1977). Following binding of the α -CA, the column was washed with 20 column volumes of lysis buffer at pH 9.0 followed by 20 column volumes of lysis buffer at pH 7.0. The α -CA was eluted with elution buffer (50 mM Tris-HCl pH 7.8, 0.4 M sodium azide), buffer-exchanged and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore, Billerica, Massachusetts, USA) to remove the sodium azide. The final step in the purification of the α -CA required a second round of *p*-AMBS affinity chromatography. The α -CA samples were pooled, desalted and concentrated using an Amicon centrifugal filter; the protein concentration was determined by UV-Vis spectroscopy at 280 nm using a molar extinction coefficient of 42 985 M⁻¹ cm⁻¹ calculated from the amino-acid sequence (Gill & von Hippel, 1989) and the purity was assessed by SDS-PAGE.

Initial crystallization screening was performed in 96-well Intelli-Plate sitting-drop vapor-diffusion crystallization plates (Art Robbins Instruments, Sunnyvale, California, USA) using a Gryphon protein-crystallization system (Art Robbins Instruments, Sunnyvale, California, USA) with four commercial screens: Crystal Screen, Crystal Screen 2, PEG/Ion and PEG/Ion 2 (Hampton Research, Aliso Viejo, California, USA). Drops consisting of protein solution at 9 mg ml⁻¹ (in 50 mM Tris-HCl pH 7.8, 100 mM NaCl) and precipitant solution at two different ratios (1:1 and 2:3 protein:precipitant solution) were equilibrated at 290 K against a reservoir consisting of 60 μ l crystallization solution. Small protein crystals were obtained in 2%(v/v) Tacsimate pH 4.0, 0.1 M sodium acetate trihydrate pH 4.6,

Figure 1
Crystals of α -CA from *T. crunogena* grown in 2%(v/v) Tacsimate pH 4.0, 0.1 M sodium acetate trihydrate pH 4.6, 16%(w/v) PEG 3350 at 290 K using the hanging-drop vapor-diffusion method. The crystal dimensions are approximately 0.05 $\times$ 0.05 $\times$ 0.01 mm.

Figure 2
X-ray diffraction image of a crystal of α -CA from *T. crunogena*. The image was collected on beamline F1 at CHESS with a crystal-to-detector distance of 300 mm, a wavelength of 0.9177 Å, a 1° oscillation angle and an exposure time of 60 s. The open circle represents 2.6 Å resolution.

Table 1

Data-collection statistics and processing parameters.

Values in parentheses are for the highest resolution shell.

Beamline	F1, CHESS
Detector	ADSC Q270 CCD
Crystal-to-detector distance (mm)	300
Wavelength (Å)	0.9177
Temperature (K)	100
Resolution range (Å)	20.00–2.60 (2.69–2.60)
Space group	C2
Unit-cell parameters (Å, °)	$a = 127.1$, $b = 102.2$, $c = 105.0$, $\beta = 127.3$
No. of measured reflections	120808
No. of unique reflections	32793
Multiplicity	3.7 (3.5)
Completeness (%)	99.8 (99.1)
Mean $I/\sigma(I)$	9.2 (2.5)
$R_{\text{merge}}^{\dagger}$ (%)	10.0 (40.5)
V_M (Å ³ Da ^{−1})	2.04

$\dagger R_{\text{merge}}$ is defined as $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl) \times 100$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl) \rangle$ is the average intensity for this reflection; the summation is over all intensities.

16% (v/v) PEG 3350 (condition No. 31 from PEG/Ion 2) after 14 d at 290 K. The crystals were cryoprotected using the precipitant solution with 20% glycerol prior to cooling in liquid nitrogen.

Diffraction data were obtained using an ADSC Q270 CCD detector on beamline F1 at the Cornell High Energy Synchrotron Source (CHESS), New York, USA. Data sets were collected at a crystal-to-detector distance of 300 mm using a wavelength of 0.9177 Å and a 1° oscillation angle with an exposure time of 60 s per image over 180 frames. The data were integrated and scaled to 2.6 Å resolution with the *HKL-2000* software (Otwinowski & Minor, 1997). The *AutoMR* wizard from the *PHENIX* software package (Adams *et al.*, 2010) was used for molecular replacement with human CA II (PDB entry 3ks3; Avvaru *et al.*, 2010) as the starting model, which has 30% sequence identity to *T. crunogena* α -CA.

3. Results and discussion

The α -CA expressed by *T. crunogena* was successfully overexpressed and isolated with a typical yield of 10 mg per litre of bacterial culture. Small rectangular crystals were formed after 14 d in a variety of crystallization conditions containing PEG 3350 in the pH range 4.0–7.4 (Fig. 1). Diffraction data to 2.6 Å resolution were collected at the CHESS synchrotron (Fig. 2). The crystals were shown to belong to space group C2, with unit-cell parameters $a = 127.1$, $b = 102.2$, $c = 105.0$ Å, $\beta = 127.3^\circ$ and with an R_{merge} of 10.0%. Initial data processing with *HKL-2000* (Otwinowski & Minor, 1997) also gave the possibility of space group F222 with unit-cell parameters $a = 102.2$, $b = 127.0$, $c = 167.2$ Å, but with an R_{merge} of ~50%, which led to the assignment of C2 as the space group. Data-collection statistics and processing parameters are summarized in Table 1.

Considering the C2 space group and assuming a protein molecular mass of 33 217 Da (with the signal sequence removed), a Matthews coefficient (V_M) of 2.04 Å³ Da^{−1} was calculated assuming the presence of four molecules in the crystallographic asymmetric unit. However, *T. crunogena* α -CA is dimeric in solution, with a molecular mass of 66 434 Da; we therefore expect that the four molecules in the asymmetric unit are actually two *T. crunogena* α -CA dimers.

We are currently determining the structure of *T. crunogena* α -CA by molecular replacement using human CA II (PDB entry 3ks3; Avvaru *et al.*, 2010) as the starting search model. A unique solution comprised of four identical protein molecules with a TFZ score of 17.9 was obtained using the *PHENIX AutoMR* wizard using all data. TFZ scores above 8 usually indicate that the structure has been solved (McCoy *et al.*, 2007). Structure refinement is now in progress.

The authors would like to thank the staff at the F1 beamline, CHESS, Cornell University and Kathy Dedrick for assistance in obtaining beam time and during X-ray diffraction data collection. We also thank Greg Stupp for taking the optical photograph of the *T. crunogena* α -CA crystals.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Avvaru, B. S., Kim, C. U., Sippel, K. H., Gruner, S. M., Agbandje-McKenna, M., Silverman, D. N. & McKenna, R. (2010). *Biochemistry*, **49**, 249–251.
- Badger, M. R. & Price, G. D. (1994). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 369–392.
- Cheggwidden, W. R., Carter, N. D. & Edwards, Y. H. (2000). *The Carbonic Anhydrases: New Horizons*. Switzerland: Birkhäuser.
- Dobriniski, K. P., Boller, A. J. & Scott, K. M. (2010). *Appl. Environ. Microbiol.* **76**, 3561–3567.
- Dobriniski, K. P., Longo, D. L. & Scott, K. M. (2005). *J. Bacteriol.* **187**, 5761–5766.
- 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.
- Duda, D. M. & McKenna, R. (2006). *Handbook of Metalloproteins*. New York: John Wiley & Sons.
- Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
- Henry, R. P. & Swenson, E. R. (2000). *Respir. Physiol.* **121**, 1–12.
- Hewett-Emmett, D. & Tashian, R. E. (1996). *Mol. Phylogenet. Evol.* **5**, 50–77.
- Khalifah, R. G., Strader, D. J., Bryant, S. H. & Gibson, S. M. (1977). *Biochemistry*, **16**, 2241–2247.
- Lane, T. W. & Morel, F. M. (2000). *Plant Physiol.* **123**, 345–352.
- Maren, T. H. (1984). *Ann. N. Y. Acad. Sci.* **429**, 568–579.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.
- Sawaya, M. R., Cannon, G. C., Heinhorst, S., Tanaka, S., Williams, E. B., Yeates, T. O. & Kerfeld, C. A. (2006). *J. Biol. Chem.* **281**, 7546–7555.
- Scott, K. M. *et al.* (2006). *PLoS Biol.* **4**, e383.
- Shively, J. M., van Keulen, G. & Meijer, W. G. (1998). *Annu. Rev. Microbiol.* **52**, 191–230.