Structure of a hyperthermostable carbonic anhydrase identified from an active hydrothermal vent chimney

Fredslund, Folmer; Borchert, Martin S.; Poulsen, Jens-Christian Navarro; Mortensen, Steen Bennike; Perner, Mirjam; Streit, Wolfgang R.; Lo Leggio, Leila

Published in:
Enzyme and Microbial Technology

DOI:
[10.1016/j.enzmictec.2018.03.009](https://doi.org/10.1016/j.enzmictec.2018.03.009)

Publication date:
2018

Document version
Peer reviewed version

Document license:
[CC BY-NC-ND](#)

Citation for published version (APA):
Fredslund, F., Borchert, M. S., Poulsen, J-C. N., Mortensen, S. B., Perner, M., Streit, W. R., & Lo Leggio, L. (2018). Structure of a hyperthermostable carbonic anhydrase identified from an active hydrothermal vent chimney. *Enzyme and Microbial Technology*, 114, 48-54. <https://doi.org/10.1016/j.enzmictec.2018.03.009>

Structure of a hyperthermostable carbonic anhydrase identified from an active hydrothermal vent chimney

Folmer Fredslund^a, Martin S. Borchert^b, Jens-Christian N. Poulsen^a, Steen Bennike Mortensen^b,
Mirjam Perner^c, Wolfgang R. Streit^d, Leila Lo Leggio^{a*}

- 5 a- Department of Chemistry, University of Copenhagen, Universitetsparken 5, Copenhagen,
 DK-2100, Denmark
- b- Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark
- c- Molecular Biology of Microbial Consortia, University of Hamburg, Biocenter Klein
 Flottbek, Ohnhorststr. 18, 22609 Hamburg, Germany
- 10 d- Microbiology and Biotechnology, Biocenter Klein Flottbek, Ohnhorststr. 18, 22609
 Hamburg, Germany

* Corresponding author, leila@chem.ku.dk

15 **Abstract** – Carbonic anhydrases (CAs) are extremely fast enzymes, which have attracted much
interest in the past due to their medical relevance and their biotechnological potential. An α -type
CA gene was isolated from DNA derived from an active hydrothermal vent chimney, in an effort to
identify novel CAs with suitable properties for CO₂ capture. The gene product was recombinantly
produced and characterized, revealing remarkable thermostability, also in the presence of high ionic
20 strength alkaline conditions, which are used in some CO₂ capture applications. The T_m was above
90 °C under all tested conditions. The enzyme was crystallized and the structure determined by
molecular replacement, revealing a typical bacterial α -type CA non-covalent dimer, but not the
disulphide mediated tetramer observed for the hyperthermophilic homologue used for molecular
replacement, from *Thermovibrio ammonificans*. Structural comparison suggests that an increased
25 secondary structure content, increased content of charges on the surface and ionic interactions
compared to mesophilic enzymes, may be main structural sources of thermostability, as previously
suggested for the homologue from *Sulfurihydrogenibium yellowstonense*.

Keywords: carbonic anhydrase, thermostability, metagenomics, crystal structure

1. Introduction¹

Carbonic anhydrases (CAs, EC 4.2.1.1) are some of the fastest enzymes on earth, achieving k_{cat} of over 10^6 s^{-1} , which approach the diffusion limit. CAs catalyse the reversible hydration of CO_2 to produce bicarbonate ions (HCO_3^-) and protons. They are ubiquitous in all domains of life and are of significant biological, pharmaceutical and biotechnological importance [1]. In the last ten years, CAs, especially thermostable ones, have also attracted considerable attention for CO_2 capture applications [2, 3]. Depending on the structure they are classified into several classes, the most prominent and best studied family being the α -class, which contains mammalian, fungal and bacterial members and are Zn-dependent. The active site Zn is coordinated by three histidine residues.

Although the α -CAs have traditionally been considered monomeric enzymes, some mammalian, fungal and bacterial members have been shown to be dimeric [1, 4]. The dimerization interfaces are not conserved across all dimeric α -class CAs which have been structurally characterized. However, within the bacterial members of the class, conservation of the dimerization interface has been observed as first noted when comparing the α -CA from *Neisseria gonorrhoeae* [5] and the α -CA (SspCA) from *Sulfurihydrogenibium yellowstonense* YO3AOP1 [6]. SspCA is also the most thermostable CA known, able to retain activity at 100°C for 3 h [6]. Another structurally characterized CA from *Thermovibrio ammonificans* (TaCA), shows a similar dimeric structure, and additionally a cysteine mediated tetramer [7]. Recently, another bacterial enzyme, SazCA from *Sulfurihydrogenibium azorense*, has been identified as the most efficient CA known to date [8] and

¹ Abbreviations: CA, carbonic anhydrase; SspCA, carbonic anhydrase from *Sulfurihydrogenibium yellowstonense* YO3AOP1; TaCA, carbonic anhydrase from *Thermovibrio ammonificans*; SazCA, carbonic anhydrase from *Sulfurihydrogenibium azorense*; PMCA, carbonic anhydrase from *Persephonella marina* DSM 14350; LOGACA, carbonic anhydrase from the Logatchev hydrothermal field metagenome; DSC, differential scanning calorimetry; PDB, protein data bank; PEG, polyethylene glycol; PEGMME, polyethylene glycol monomethyl ether; TSA, thermal shift assay

its structure has been determined [9]. SazCA is more than 60 % identical to SspCA and forms, as the other bacterial CAs characterized so far [7], non-covalent dimers. Though SazCA is hyperthermostable, it is not so to the same extent as SspCA. Thus attempts have been made to transfer the properties of SazCA to SspCA by site directed mutagenesis, but have to date not
55 succeeded [10]. Through a comparison of SspCA to human mesophilic counterparts, several structural determinants of thermostability were suggested by [6]: an increase in secondary structure elements and ‘compactness’; an increase in charged residues contribution to surface area; increased long range ionic interactions and ionic networks. For TaCA [7], the main structural feature leading to thermostability was identified as the formation of tetramers through 4 ion pairs and 2 disulphide
60 bridges. TaCA in its tetrameric form retains 90% activity upon incubation at 70 °C for 1 hour.

In an effort to identify novel CAs with suitable properties for CO₂ capture, we have cloned, expressed, structurally and functionally characterized a CA from a metagenome derived from an active chimney structure emanating hot, reduced hydrothermal fluids.

2. Materials and methods

65 2.1. Sample collection and sample description

Parts of an active hydrothermal vent chimney were collected from the Logatchev hydrothermal field located on the Mid-Atlantic Ridge at 14°45’N and 44°58’W at a water depth of 3000 m. The sample was retrieved during a dive with the remotely operated vehicle Kiel ROV 6000 (GEOMAR, Kiel) during the cruise HYDROMAR VII (January/February 2009). On the inside of the chimney
70 conditions can be characterized as extremely hot (temperatures of the emanating fluids at the chimney outlet measured 350°C) and highly reduced [11], while on the outside of the chimney ambient seawater conditions prevail, i.e. oxygen is present and the temperature is 4°C. Hence, a steep chemical and thermal gradient exists within the chimney (from inside to outside). The hydrothermal fluids are generally enriched in hydrogen and methane due to serpentinization

75 reactions. After sample recovery the chimney piece was immediately frozen at -80°C until further processing.

2.2. DNA extraction

A fragment of the frozen chimney sample was pestled to fine powder and the fine powder was suspended in 2 mL washing solution (800 mM NaCl, 100 mM EDTA pH 8.0) and incubated for 1 h
80 on ice. Then 2 mL TE-sucrose buffer (10 mM Tris, 1 mM EDTA, 20% (w/v) sucrose, pH 8.0) were added and the sample was incubated in an ice bath for 15 min. After the addition of 2 mL cell lysis buffer (10 mM Tris, 1 mM EDTA, 10 mg/mL lysozyme, 1 mg/mL RNase A, pH 8.0) the sample was incubated at 37°C for 1 h. Then 1 mL Proteinase K (1 mg/mL in 5% (w/v) sacrosyl) was added and the reaction was again incubated at 37°C for 1 h. The DNA was extracted with
85 phenol/chloroform (1:1). To precipitate the DNA from the upper phase, 2.5 vol ethanol (99%, vol/vol) and 0.1 vol 3 M sodium acetate (pH 5.5) were added. The tube was inverted and incubated at -20°C for at least 1 h. Then the DNA was sedimented at 13,000 g and 4°C for 20 min and the supernatant was decanted carefully. The DNA pellet was washed twice with 300 uL ice-cold ethanol (70%, vol/vol) each followed by a centrifugation of 2 min at 4°C and 13,000 g. After
90 discarding the supernatant, the pellet was air-dried. The DNA was then dissolved in an appropriate volume of sterile water. The DNA was subjected to Illumina sequencing (300 bp PE run). Read count was 12,141,468 with 2,888,817,382 bp. The N50 was 4,941 and the raw coverage /x 52.5456. The assembled length was 54,977,361 bp on 17,000 contigs. The CA gene (GenBank entry MG547709) was discovered in the metagenome by homology searches. The metagenome CA
95 protein sequence displayed 83% identity to bacterial CA from *Persephonella* sp. IF05-L8 (NCBI Reference Sequence WP_029521561) and 82% identity to CA from *Persephonella marina* EX-H1 or *Persephonella marina* DSM 14350 (PMCA, NCBI sequence reference WP_015898908 and sequence from patent WO 2012/025577[12]).

2.3. Protein production

100 A synthetic CA (referred hereon as LOGACA) gene corresponding to the identified sequence was cloned into *Bacillus subtilis* in the same expression vector as previously used for PMCA[12]. The native CA signal peptide was exchanged with a *Bacillus* signal peptide for recombinant expression in *B. subtilis*, leaving the mature peptide of LOGACA after predicted cleavage with Signal Peptide peptidase with N-terminal residues 'GGVGHW'. The recombinant CA protein expression was
105 performed in rich media supplemented with 6 mg/L chloramphenicol and successful expression and purification using ion-exchange chromatography (SP-Sepharose) was monitored by SDS-PAGE analysis. It was determined that there was very high CA activity in the culture broth solution, measured according to [13]. Purified LOGACA was stored in 50 mM MES pH 6, 100 mM NaCl at a protein concentration of 5.2 mg/mL.

110 2.4. Biochemical characterization

The thermostability of the recombinant LOGACA was determined by thermal shift assay (TSA) in a Roche Lightcycler 480 II machine running Roche LightCycler 480 software (release 1.5.0 SP4) as follows: TSA were run with enzyme samples diluted to 0.3 mg/ml in assay buffers: 0.1 M succinic acid, 0.1 M HEPES, 0.1 M CHES, 0.1 M CAPS, 0.15 M KCl, 1 mM CaCl₂, 0.01 % Triton
115 X100, pH adjusted to 5.0, 7.5 and 10.0 respectively. SYPRO Orange dye (Life Technologies S6650) was diluted 101x in deionized water. 10 µl diluted enzyme sample + plus 10 µl assay buffer + plus 10 µl dye were mixed in wells of TSA assay plates (Light-Cycler 480 Multiwell plate 96, white, Roche and covered with optic seal LightCycler 480 Sealing foil, Roche). Protein melting analysis was conducted at 25-99 °C at 200 °C/h. The LOGACA and a PMCA reference sample for
120 comparison with [14] were analysed in duplicate and the data used to determine T_m, defined as the midpoint value of the protein melting curves.

The thermostability of LOGACA and PMCA was also determined by Differential Scanning Calorimetry (DSC) at high ionic strength, because certain high ionic strength alkaline solvents are used in CO₂ capture applications. Samples were diluted to approximately 1 mg/ml in 1.5 M Glycine buffer at pH 8, 9 or 10 and the thermal midpoint (T_m) determined by scanning from 20 – 120°C at 200°C per hour.

The specific activity of LOGACA determined according to [13] was compared to PMCA in order to assess industrial applicability. Esterase activity with *p*-nitrophenyl acetate (*p*-NPA) was measured with a continuous assay in 96-well format. 15 µL dilution of the substrate in 2-propanol was added 285 µl enzyme in 0.1 M phosphate buffer pH 7.5 (total enzyme concentration 7.6 µM) at room temperature and the reaction monitored by following A_{405} . Both LOGACA and PMCA followed Michaelis-Menten kinetics. K_m , V_{max} and k_{cat} could be calculated using a simple web-tool (<http://www.ic50.tk/kmvmax.html>). Activity at high pHs could not be measured accurately due to an extremely high background due to fast non-catalyzed reaction. An apparent ϵ_{405} of 14720 M⁻¹ was used for pNP in the microtiter plate assay.

2.5. Crystallization and data collection

Initial crystallization attempts were carried out with an Oryx 8 liquid handling robot (Douglas Instruments Ltd) in MRC 2-well sitting drop trays, using 0.3 µL drops (protein:reservoir ratio of 1:1 and 3:1) and using the JCSG+ and Morpheus screens (Molecular Dimensions, Newman et al. 2005). Crystals were obtained in 20% w/v PEGMME 500, 10%w/v PEG 20000, 0.1M MES/Imidazole pH 6.5, diffracting to around 4Å in space group $P4_12_12$ (the crystals were thin needles with dimensions around 20x20x300 µm³, length varying between droplets), but 2D grid optimization of these and three other initial crystallization hits were unsuccessful. Additive screening revealed that crystal morphology improved when the protein sample was preincubated with 10 mM of the reducing agent TCEP (*tris*(2-carboxyethyl)phosphine) and screening using standard screens was restarted with the

new sample. The most promising new crystallization conditions was optimized for pH and the final crystallization conditions was with a precipitant of 20% w/v Ethylene glycol , 10% w/v PEG 8000 , 0.1 M Citric acid pH 4.0. The optimized crystals were obtained by sitting drop vapour diffusion with 2+2 μ l droplets (protein:reservoir, with some drops being added 1-7 μ l water during setup) at room temperature with a protein concentration of 5.2 mg/ml, pre-incubated for one hour with 10 mM TCEP. These crystals were approximately 40x40x70 μ m³ and they diffracted to 2.56Å. Data was collected at the MAX-lab beamline I911-3 [15] with a total of 200° of data with 0.5° per frame. Processing and scaling was performed with *XDS/XSCALE* [16] using *xdsapp* [17]. The crystal belonged to space group *P1*, with 6 molecules in the asymmetric unit. Processing and crystal data are given in Table 1.

2.6. Structure determination and refinement

Structure determination and refinement was carried out using the PHENIX suite [18]. The search model was created starting with the structure of TaCA (PDB 4C3T), having 57% sequence identity with the target [7], which was edited with *phenix.sculptor*. Molecular replacement with *Phaser* [19] found 6 molecules with a TF Z-score of 19.7 indicated a very clear solution. The resulting model was then iteratively rebuilt with *COOT* [20] and refined using *phenix.refine* [21], initially with NCS restraints that were removed during the final stages of refinement. Refinement and validation statistics are given in Table 1.

2.7. Structure analysis and representation

The structure was analysed with *qtpisa* [22] and *PyMOL* [23]. Hydrogen bonds were calculated with *HBPLUS* [24], secondary structure was assigned and calculated with *DSSP* [25] using the *DSSP* and *Stride* plugin for *PyMOL* [26]. The solvent accessible surface area was calculated within *PyMOL* using the *get_area* command with *dot_solvent* on and *dot_density* set to 4. Possible ion pairs were calculated within *PyMOL* using side chain oxygen atoms from Asp/Glu residues as

170 donors and nitrogen atoms from His/Arg/Lys as acceptors. Charged residues were defined as Arg, Lys, Asp, Glu and His. Polar residues were defined as Gly, Ser, Thr, Tyr, Gln, Asn and Cys. Hydrophobic residues were defined as Ala, Leu, Ile, Trp, Pro, Phe, Met and Val. Aromatic residues were defined as Trp, Phe, Tyr and His.

2.8. Coordinates

175 The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession code 6EKI.

3. Results and discussion

3.1. Activity and Thermostability

LOGACA had comparable activity as PMCA (1.4 times the specific activity according to the
180 Wilbur-assay described in [13]. Esterase activity was measured using *p*-NPA as substrate for both LOGACA and PMCA as a reference. The K_m values measured were 0.25 ± 0.06 mM for LOGACA and 0.39 ± 0.05 mM for PMCA. These values are somewhat lower than previously reported for PMCA, but comparable to the ones reported for other marine CAs which are in the submillimolar range [14]. The apparent V_{max} under the given conditions were 0.32 ± 0.034 $\mu\text{mol pNP mg}^{-1}\text{min}^{-1}$
185 for LOGACA and 0.42 ± 0.026 $\mu\text{mol pNP mg}^{-1}\text{min}^{-1}$ for PMCA, which again are comparable to previously reported values. The calculated k_{cat}/K_m for LOGACA was $499 \text{ M}^{-1} \text{ s}^{-1}$ and $407 \text{ M}^{-1} \text{ s}^{-1}$ for PMCA, while for human CA II the reported k_{cat}/K_m for pNPA is $2080 \text{ M}^{-1} \text{ s}^{-1}$ [27]. In summary, the purified LOGACA has comparable esterase activity to other marine and non marine carbonic anhydrases.

190 Using TSA, the T_m for LOGACA was 91°C under the chosen conditions, while for PMCA it was 87°C . In high Glycine concentration buffer, although PMCA has a higher T_m relative to LOGACA at pH 8.0 and 9.0, at pH 10.0 the PMCA thermal profile indicates precipitation of the protein in this buffer system (an exotherm was obtained instead of a melting endotherm), whereas the thermal

195 profile of LOGACA is improved at pH 10.0 relative to pH 8.0 and 9.0 (Table 2) and no indication
of precipitation was observed across the range pH 8-10. Thus, the data indicates that the LOGACA,
while having comparable activity, is even more thermally stable at $\text{pH} \geq 10$ relative to PMCA,
which might prove advantageous for CO_2 capture applications that typically operate across alkaline
pH ranges [3,12,14]. SspCA has previously been reported (at pH 6.0) to have a reversible transition
around 87 °C, and could not be irreversibly denatured below 105 °C (the highest temperature that
200 could be experimentally tested) [6]. Thus LOGACA is among the most thermostable α -CAs
characterized to date, especially at high pH. The latter observation is somewhat surprising, as the
environment of the hydrothermal vent is acidic.

3.2. Structure description

The structure of LOGACA was solved to a resolution of 2.56Å in space group *P1*, with 3
205 homodimers in the asymmetric unit. Each non-covalent homodimer has an interaction surface of
1000 Å². The six monomers in the asymmetric unit are very similar with RMSD values ranging
from 0.26Å (for the A/C pair) to 0.53Å (for the D/E pair) calculated with CEAlign in PyMOL for
the 216 equivalent amino acid residues. The first 4 residues of the mature protein were not visible in
the experimental electron density and have not been modelled, but the rest of the residues are
210 visible. A central 10 stranded β -sheet is the core of the monomeric structure with the catalytic Zn
ion of the active site coordinated by 3 histidine residue atoms (His117-NE2, His119-NE2 and
His136-ND1, see Figure 1A) at distances in the ranges 2.06–2.18Å (His117-NE2), 2.08–2.28Å
(His119-NE2), and 2.11–2.25Å (His136-ND1). His92 corresponds to the His fulfilling the function
of proton shuttle in other α -CAs. Thr202 has its OG1 at a distance of 3.4–3.7Å from the Zn ion. A
215 corresponding Thr is in human CA II considered to be involved in catalysis by coordinating the Zn-
bound water molecule [28]. The electron density around the Zn in the LOGACA structure does not
however fully support modelling the Zn-bound water in the same position, but rather with a water

molecule positioned nearer the CO₂ binding pocket [29], where Thr202 cannot coordinate it. However the resolution of the LOGACA structure is rather limited compared to the human CA II structure available, and thus no mechanistic conclusions should be drawn from the differences in coordination around the LOGACA Zn.

3.3. Comparison to other thermostable bacterial CAs

A sequence alignment of LOGACA and sequences of other CAs referred to in this article is shown in Figure 2. The structure of LOGACA was compared to TaCA, the molecular replacement model, and the two other hyperthermostable bacterial α -CAs for which structures are known. Chain A of each structure was used for comparison. The final refined model of LOGACA has an RMSD of 0.84Å compared to TaCA (over 216 residues, calculated using CEalign in PyMOL) with a sequence identity/similarity of 57.4%/71.1%. LOGACA has an RMSD of 1.23Å compared to SspCA (over 216 residues, identity/similarity 45.1%/65.0%) and an RMSD of 1.11Å compared to SazCA (also over 216 residues, identity/similarity 50.4%/65.9%). Thus LOGACA is structurally most similar to TaCA among the structurally characterized thermostable bacterial CAs. However, the disulphides forming the tetramer in TaCA and important for its thermostability, are not found in LOGACA. The area around the His fulfilling the function of proton shuttle in bacterial α -CAs has previously been discussed [9] while trying to understand the structural basis of the extreme catalytic efficiency of SazCA. SazCA was identified to have two histidine residues (His2 and His207) that could influence the pK_a of the proton shuttle histidine. In the structures of LOGACA, SspCA and TaCa His2 from SazCA is present in LOGACA and TaCA, although the position of the sidechains are further away from the proton shuttle histidine, and a histidine corresponding to His207 is not present in the other enzymes. Especially in absence of comparative activity measurements with all enzymes under identical conditions, it is difficult to comment on the effect of the environment around the proton-shuttle.

The structural features of LOGACA, TaCA, SspCA and SazCA that can correlate with thermostability have been analysed similarly as in [6] and compared (Table 3). To give a meaningful comparison of structural features, the dimer of TaCA corresponding to the other bacterial CAs dimers was used, though the covalent tetramer is the likely biological unit for TaCA [7]. The trends which were seen for SspCA when compared to mesophilic carbonic anhydrases are also seen for LOGACA, which has even higher secondary structure content, charged surface residues, and comparable or higher number of residues involved in ion pairs than SspCA, SazCA and TaCA. These seem therefore general thermostabilizing features of bacterial α -CAs.

250

Acknowledgments

We thank the captain and crews of the RV MS Merian as well as the ROV Kiel 6000 (GEOMAR, Kiel) for helping us to obtain the deep-sea vent sample. We wish to thank MAXLAB, Sweden and the European Synchrotron Radiation Facility (ESRF), France, for synchrotron beamtime and assistance. LLL is a member of ISBUC Integrative Structural Biology at the University of Copenhagen (www.isbuc.ku.dk).

Funding

The work was partially supported by grants from the Deutsche Forschungsgemeinschaft (DFG) priority program 1144 "From Mantle to Ocean: Energy-, Material- and Life-cycles at Spreading Axes". Travel to synchrotrons was supported by the Danish Ministry of Higher Education and Science through the Instrument Center DANSCATT and the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement 283570). Part of the work was funded by Novozymes A/S. This research did not receive any other specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

265

References

- [1] C.T. Supuran, Structure and function of carbonic anhydrases, *Biochem. J.*, 473 (2016) 2023-2032.
- [2] C.D. Boone, S. Gill, A. Habibzadegan, R. McKenna, Carbonic anhydrase: an efficient enzyme
270 with possible global implications, *Int. J. Chem. Eng.*, 2013 (2013) 6.
- [3] A. Di Fiore, V. Alterio, S.M. Monti, G. De Simone, K. D'Ambrosio, Thermostable carbonic anhydrases in biotechnological applications, *Int. J. Mol. Sci.*, 16 (2015) 15456-15480.
- [4] J.A. Cuesta-Seijo, M.S. Borchert, J.C. Navarro-Poulsen, K.M. Schnorr, S.B. Mortensen, L. Lo
275 Leggio, Structure of a dimeric fungal alpha-type carbonic anhydrase, *FEBS Lett.*, 585 (2011) 1042-1048.
- [5] S.H. Huang, Y.F. Xue, E. Sauer-Eriksson, L. Chirica, S. Lindskog, B.H. Jonsson, Crystal structure of carbonic anhydrase from *Neisseria gonorrhoeae* and its complex with the inhibitor acetazolamide, *J. Mol. Biol.*, 283 (1998) 301-310.
- [6] A. Di Fiore, C. Capasso, V. De Luca, S.M. Monti, V. Carginale, C.T. Supuran, A. Scozzafava,
280 C. Pedone, M. Rossi, G. De Simone, X-ray structure of the first 'extremo- α -carbonic anhydrase', a dimeric enzyme from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1, *Acta Crystallogr. D*, 69 (2013) 1150-1159.
- [7] P. James, M.N. Isupov, C. Sayer, V. Saneei, S. Berg, M. Lioliou, H.K. Kotlar, J.A. Littlechild,
285 The structure of a tetrameric alpha-carbonic anhydrase from *Thermovibrio ammonificans* reveals a core formed around intermolecular disulfides that contribute to its thermostability, *Acta Crystallogr. D*, 70 (2014) 2607-2618.
- [8] V. De Luca, D. Vullo, A. Scozzafava, V. Carginale, M. Rossi, C.T. Supuran, C. Capasso, An α -carbonic anhydrase from the thermophilic bacterium *Sulphurihydrogenibium azorense* is the fastest enzyme known for the CO₂ hydration reaction, *Bioorg. Med. Chem.*, 21 (2013) 1465-1469.

- 290 [9] G. De Simone, S.M. Monti, V. Alterio, M. Buonanno, V. De Luca, M. Rossi, V. Carginale, C.T. Supuran, C. Capasso, A. Di Fiore, Crystal structure of the most catalytically effective carbonic anhydrase enzyme known, SazCA from the thermophilic bacterium *Sulfurihydrogenibium azorense*, Bioorg. Med. Chem. Lett., 25 (2015) 2002-2006.
- [10] V. De Luca, S. Del Prete, V. Carginale, D. Vullo, C.T. Supuran, C. Capasso, A failed tentative
295 to design a super carbonic anhydrase having the biochemical properties of the most thermostable CA (SspCA) and the fastest (SazCA) enzymes, J. Enzyme Inhib. Med. Chem., 30 (2015) 989-994.
- [11] M. Perner, J. Kuever, R. Seifert, T. Pape, A. Koschinsky, K. Schmidt, H. Strauss, J.F. Imhoff, The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15 degrees N on the Mid-Atlantic Ridge, FEMS Microbiol. Ecol., 61 (2007) 97-109.
- 300 [12] M. Borchert, Heat stable *Persephonella* carbonic anhydrases and their uses, Patent WO2012025577, 2012.
- [13] K.M. Wilbur, N.G. Anderson, Electrometric and colorimetric determination of carbonic anhydrase, J. Biol. Chem., 176 (1948) 147-154.
- [14] B.K. Kanth, S.Y. Jun, S. Kumari, S.P. Pack, Highly thermostable carbonic anhydrase from
305 *Persephonella marina* EX-H1: Its expression and characterization for CO₂-sequestration applications, Process Biochemistry, 49 (2014) 2114-2121.
- [15] T. Ursby, J. Unge, R. Appio, D.T. Logan, F. Fredslund, C. Svensson, K. Larsson, A. Labrador, M. Thunnissen, The macromolecular crystallography beamline I911-3 at the MAX IV laboratory, J. Synchrotron Radiat., 20 (2013) 648-653.
- 310 [16] W. Kabsch, XDS, Acta Crystallogr. D, 66 (2010) 125-132.
- [17] K.M. Sparta, M. Krug, U. Heinemann, U. Mueller, M.S. Weiss, XDSAPP2.0, J. Appl. Crystallogr., 49 (2016) 1085-1092.

- [18] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr. D*, 66 (2010) 213-221.
- [19] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, *J. Appl. Crystallogr.*, 40 (2007) 658-674.
- [20] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. D*, 60 (2004) 2126-2132.
- [21] P.V. Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr. D*, 68 (2012) 352-367.
- [22] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.*, 372 (2007) 774-797.
- [23] Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.7.7.0, 2010.
- [24] I.K. McDonald, J.M. Thornton, Satisfying hydrogen-bonding potential in proteins, *J. Mol. Biol.*, 238 (1994) 777-793.
- [25] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolym.*, 22 (1983) 2577-2637.
- [26] H. Zhu, DSSP & Stride plugin for PyMOL, unpublished (Biotechnology Center (BIOTEC), TU Dresden, (2011).
- [27] G. Höst, L.-G. Mårtensson, B.-H. Jonsson, Redesign of human carbonic anhydrase II for increased esterase activity and specificity towards esters with long acyl chains, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1764 (2006) 1601-1606.

[28] A. Liljas, Carbonic anhydrase under pressure, *IUCrJ*, 5 (2018) 4-5.

[29] J.F. Domsic, B.S. Avvaru, C.U. Kim, S.M. Gruner, M. Agbandje-McKenna, D.N. Silverman, R. McKenna, Entrapment of carbon dioxide in the active site of carbonic anhydrase II, *J. Biol.*

340 *Chem.*, 283 (2008) 30766-30771.

Table and figure legends

Table 1 - Data collection, processing and refinement. Values for the outer shell are given in parentheses.

Table 2 - Melting temperature (T_m) of LOGACA and PMCA in 1.5 M Glycine at pH 8.0, 9.0 or 10.0 determined by DSC.

Table 3 - Comparison of structural features for LOGACA, TaCA, SspCA and SazCA that might influence thermostability.

Figure 1 - The structure of LOGACA and comparison with other carbonic anhydrases. (A) The structure of one monomer of LOGACA with a close-up of the active site. His92 (blue) functions as a proton shuttle and residues Val138, Val148, Leu201, Val210 and Trp212 (orange) compose the CO₂ binding pocket, here with a water molecule. The Zn ion is coordinated by His117, His119 and His136 (green). (B) Cartoon representation of the homodimeric structural arrangement of LOGACA (green loops, yellow beta-strands, red alpha-helices) aligned with the non-covalently bound equivalent dimer from TaCA (PDB 4C3T) for comparison. (C) Ribbon representation of LOGACA with SspCA and SazCA.

Figure 2 - Structure-based sequence alignment between LOGACA and other bacterial carbonic anhydrases, NgCA (*Neisseria gonorrhoea* carbonic anhydrase), TaCA, SspCA and SazCA. The sequence of PMCA is also added to the alignment since this CA was used as reference in the biochemical studies.

Table 1.

Diffraction source	I911-3 at MAXlab
Wavelength (Å)	1.06883
Temperature (K)	100
Detector	MAR MOSAIC 225
Crystal-detector distance (mm)	238.36
Rotation range per image (°)	0.5
Total rotation range (°)	200
Exposure time per image (s)	
Space group	<i>P</i> 1
a, b, c (Å)	46.32 71.28 103.34
α , β , γ (°)	88.38 82.84 84.22
Mosaicity (°)	0.382
Resolution range (Å)	45.73-2.56 (2.71-2.56)
Total No. of reflections	89089
No. of unique reflections	40713
Completeness (%)	96.3 (94.8)
Multiplicity	2.19
$\langle I/\sigma(I) \rangle$	11.06 (1.87)
R_{meas} [%]	8.9 (60.7)
$CC_{1/2}$	99.6 (81.4)

Completeness [%]	96.3 (94.8)
Overall B factor from Wilson plot ($\text{\AA}^2$)	45.83
Isa	28.15
No. of reflections, working set	40699 (3938)
No. of reflections, test set	2350 (228)
Final R_{cryst}	0.1685 (0.2842)
Final R_{free}	0.2379 (0.3784)
No. of non-H atoms	
Protein	10734
Ion	6
Water	76
Total	10816
R.m.s deviations	
Bonds ($\text{\AA}$)	0.016
Angles ($^\circ$)	1.62
Average B factors ($\text{\AA}^2$)	48.91
Protein	48.98
Ion	43.25
Water	38.94
Ramachandran plot	

Favoured regions (%)	96.4
----------------------	------

Additionally allowed (%)	3.5
--------------------------	-----

Outliers (%)	0.1
--------------	-----

Table 2.

	LOGACA	PMCA
T_m at pH 8.0	92.2 °C	104.3 °C
T_m at pH 9.0	91.1 °C	104.1 °C
T_m at pH 10.0	103.6 °C	Precipitation

Table 3.

Comparison of structural features for LOGACA, TaCA, SspCA and SazCA that might influence thermostability

Protein	LOGACA	SspCA	SazCA	TaCA
PDB code	6eki (this work)	4g7a	4x5s	4c3t
No. of residues in monomer	222	225	224	225
Charged residues (%) (Arg, Lys, Asp, Glu and His)	28.8	31	27.2	29.3
Polar residues (%) (Gly, Ser, Thr, Tyr, Tyr, Gln, Asn and Cys)	31.6	32	34.4	30.2
Hydrophobic residues (%) (Ala, Leu, Ile, Trp, Pro, Met and Val)	39.5	37	38.4	40.4
Aromatic residues (%) (Trp, Phe, Tyr and His)	12.6	12	15.6	15.5
Oligomeric state	Dimer	Dimer	Dimer	Dimer
Accessible surface area ($\text{\AA}^2$)				
Total	19631	20771	19429	20188
Charged residues	12039 (61%)	12134 (58%)	9926 (51%)	11223 (56%)
Polar residues	4255 (22%)	5141 (25%)	5758 (30%)	4879 (24%)
Hydrophobic residues	3331 (17%)	3495 (17%)	3743 (19%)	4084 (20%)
Secondary-structure content (%)	57.7	50.4	47.6	49.8

α -Helices	12.6	9.4	9.8	9.8
β -Strands	39.2	33	30.2	32.9
3_{10} -helices	5.9	8	7.6	7.1
Hydrogen bonds	367	398	412	428
Main chain-main chain	212	202	211	214
Main chain-side chain	109	131	141	156
Side chain-side chain	46	65	60	58
Residues in ion pairs (4 Å cut-off)				
His/Lys/Arg	24	21	21	24
Asp/Glu	18	15	16	18

Figure 2.