Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

X-ray structure of the first 'extremo-α-carbonic anhydrase', a dimeric enzyme from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1

Anna Di Fiore,[a] Clemente Capasso,[b] Viviana De Luca,[b] Simona Maria Monti,[a] Vincenzo Carginale,[b] Claudiu T. Supuran,[c] Andrea Scozzafava,[c] Carlo Pedone,[a] Mosè Rossi[b] and Giuseppina De Simone[a]*

[a]Istituto di Biostrutture e Bioimmagini – CNR, Via Mezzocannone 16, 80134 Napoli, Italy, [b]Istituto di Biochimica delle Proteine – CNR, Via P. Castellino 111, 80131 Napoli, Italy, and [c]Laboratorio di Chimica Bioinorganica, Università degli Studi di Firenze, Rm. 188, Via della Lastruccia 3, 50019 Sesto Fiorentino (Firenze), Italy

Correspondence e-mail: gdesimon@unina.it

Received 30 November 2012
Accepted 15 March 2013

PDB Reference: SspCA, 4g7a

SspCA, a novel 'extremo-α-carbonic anhydrase' isolated from the thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1, is an efficient catalyst for the hydration of CO_2 and presents exceptional thermostability. Indeed, SspCA retains a high catalytic activity even after being heated to 343–373 K for several hours. Here, the crystallographic structure of this α-carbonic anhydrase (α-CA) is reported and the factors responsible for its function at high temperature are elucidated. In particular, the study suggests that increased structural compactness, together with an increased number of charged residues on the protein surface and a greater number of ionic networks, seem to be the key factors involved in the higher thermostability of this enzyme with respect to its mesophilic homologues. These findings are of extreme importance, since they provide a structural basis for the understanding of the mechanisms responsible for thermal stability in the α-CA family for the first time. The data obtained offer a tool that can be exploited to engineer α-CAs in order to obtain enzymes with enhanced thermostability for use in the harsh conditions of the CO_2 capture and sequestration processes.

1. Introduction

The enhancement in the atmosphere of the greenhouse gas carbon dioxide (CO_2), mainly derived from the increased burning of fossil fuels, has been shown to be the main source of recent global warming (Savile & Lalonde, 2011); thus, the development of efficient technologies for CO_2 capture and sequestration is urgently required (Lee *et al.*, 2010). Various methods have been proposed in order to capture and concentrate CO_2 from combustion gases for its transit and storage. Among these, the most interesting are (i) the chemical absorption of CO_2 by alkanolamines (Greer *et al.*, 2010) and its subsequent stripping with steam and (ii) the sequestration of CO_2 as an inorganic salt, since carbonate minerals represent the largest reserve of CO_2 on earth (Mignardi *et al.*, 2011). However, these classical processes are expensive and are not convenient in all situations. To overcome these problems, in recent years there has been increasing interest in using carbonic anhydrases (CAs) in CO_2 capture and sequestration processes (Savile & Lalonde, 2011; Lee *et al.*, 2010; Fisher *et al.*, 2012).

CAs are ubiquitous metalloenzymes that catalyze the reversible hydration of carbon dioxide with the production of bicarbonate and protons (Alterio *et al.*, 2012). Five genetically distinct CA families are known to date, the α-, β-, γ-, δ- and

ζ-CAs, all of which use zinc(II), cadmium(II) or iron(II) in their active sites (Alterio *et al.*, 2012). The members of the α-class seem to be the most interesting candidates for the above-mentioned applications owing to their very high catalytic activity when compared with the enzymes of the other classes (Fisher *et al.*, 2012). However, their use generally requires stable enzymes that are able to work under the harsh conditions typical of the CO_2-capture process (*i.e.* high temperature, high ionic strength *etc.*; Mignardi *et al.*, 2011); thus, the employment of enzymes isolated from microorganisms that live at high temperatures, *i.e.* thermophiles, may represent an interesting possibility.

In recent years, the scientific community has shown an increasing interest in the use of thermophilic proteins in understanding enzyme evolution, the molecular mechanisms of protein thermostability and the upper temperature limit of enzyme function (Unsworth *et al.*, 2007). Thermophilic organisms, which can be generally classified into moderate thermophiles, extreme thermophiles and hyperthermophiles, optimally grow at temperatures which range from 323 K to more than 363 K, and their 'extremozymes' are usually active at very high temperatures (Adams, 1993; Adams & Kelly, 1995; Adams *et al.*, 1995; Gomes & Steiner, 2004) and present a number of biotechnological advantages over mesophilic or psychrophilic enzymes. Indeed, once expressed in mesophilic hosts, thermophilic enzymes preserve most of their features, being easily purified by heat treatment, showing high resistance to chemical denaturants and maintaining their enzymatic activity at high temperatures (Myers *et al.*, 1995; Kumar & Nussinov, 2001). Furthermore, as they work at high temperatures, they allow the use of higher substrate concentrations and low viscosity, have fewer risks of microbial contamination and often have higher reaction rates (Adams, 1993; Adams & Kelly, 1995; Adams *et al.*, 1995).

Recently, Capasso *et al.* (2012) identified an α-CA from the newly discovered thermophilic bacterium *Sulfurihydrogenibium yellowstonense* YO3AOP1. This genus is found in chemosynthetic bacterial communities living in hot springs at temperatures of up to 383 K. The isolated α-CA, hereafter referred to as SspCA, has exceptional thermal stability, retaining its high catalytic activity for the CO_2 hydration reaction even after being heated to 343–373 K for several hours (Capasso *et al.*, 2012). Furthermore, after immobilization within a polyurethane foam, SspCA is active and stable for up to 50 h at 373 K (Migliardini *et al.*, 2013). Phylogenetic analysis of bacterial/coral/nematode/mammalian α-CAs showed that SspCA is more closely related to other bacterial α-CAs than to nematode, coral or mammalian enzymes. This is not unexpected, but it should be mentioned that SspCA shows a catalytic activity that is very similar to that of the human cytosolic isoform hCA II and not to those of other bacterial α-CAs investigated previously (Vullo *et al.*, 2013). All of these features make SspCA a very interesting candidate for the above-mentioned biotechnological applications.

In this paper, we report the solution of the crystallographic structure of SspCA, allowing two important tasks to be accomplished. On one hand, in this study we provide important insights into the structural features of bacterial α-CAs, which have been only poorly characterized to date. Indeed, kinetic and inhibition studies have only been reported for the bacterial α-CAs from *Neisseria gonorrhoeae* (Chirică *et al.*, 1997), *N. sicca* (Adler *et al.*, 1972) and *Helicobacter pylori* (Nishimori *et al.*, 2006), whereas structural studies are only available for the α-CA from *N. gonorrhoeae* (Huang *et al.*, 1998). On the other hand, this enzyme is the first 'thermostable α-CA' to be structurally characterized, and comparison with mesophilic members of the same class allows important information on the structural determinants responsible for themostability in the α-class of the CA family to be derived. These findings may constitute a basis for the engineering of α-CAs to obtain enzymes with enhanced thermostability.

2. Materials and methods

2.1. Construct preparation, protein expression and purification

GeneArt (Invitrogen), which specializes in gene synthesis, designed a synthetic *Sulfurihydrogenibium* sp. gene encoding SspCA lacking the signal peptide (the first 20 amino acids of the peptide sequence) and containing *Nde*I and *Xho*I sites at the 5′ and the 3′ ends, respectively. The resulting plasmid was amplified into *Escherichia coli* DH5α cells. The *Sulfurihydrogenibium* sp. DNA fragments were separated on 1% agarose gel. The recovered *Sulfurihydrogenibium* sp. gene and linearized expression vector (pET15-b) were ligated by T4 DNA ligase to form the expression vector pET15-b/Ssp. Competent *E. coli* BL21 (DE3) cells were transformed with pET15-b/Ssp, grown at 310 K and induced with 1 m*M* IPTG. After growth for an additional 5 h, the cells were harvested and disrupted by sonication at 277 K. Following centrifugation, the cell extract was heated at 363 K for 30 min and centrifuged. The supernatant was loaded onto HIS-Select HF Nickel Affinity Gel (Sigma–Aldrich) and the protein was eluted with 250 m*M* imidazole. At this stage of purification the enzyme was at least 95% pure. For further details of enzyme expression and purification, see Capasso *et al.* (2012).

2.2. Assay for carbonic anhydrase using CO_2 as a substrate

The CA activity assay was a modification of the procedure described by Chirică *et al.* (1997). The assay was based on monitoring the pH variation arising from the catalyzed conversion of CO_2 to bicarbonate and protons. For details, see Capasso *et al.* (2012).

2.3. Effect of temperature on the stability of SspCA, hCA I and hCA II

In order to compare the stabilities of the α-CAs considered in the present paper at different temperatures, enzymes at a concentration of 3 μg ml^{-1} in 10 m*M* Tris–HCl pH 8.3 were incubated at 313, 323, 333, 343, 353, 363 and 373 K for different times (30 and 180 min). Enzyme aliquots (30 ng) were withdrawn at appropriate times and the residual activity was measured at 273 K using CO_2 as a substrate.

Table 1
Crystal parameters, data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P2_1$
Unit-cell parameters	
a (Å)	48.50
b (Å)	68.36
c (Å)	74.69
β (°)	106.03
Data-collection statistics	
Resolution (Å)	20.00–1.80 (1.86–1.80)
Total reflections	127441
Unique reflections	42233 (3520)
Multiplicity	3.0 (1.6)
Completeness (%)	97.4 (81.2)
$R_{\rm merge}$†	0.036 (0.140)
Mean $I/\sigma(I)$	23.34 (4.78)
Refinement statistics	
Resolution (Å)	20.00–1.80
$R_{\rm cryst}$† (%)	16.8
$R_{\rm free}$‡ (%)	20.3
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.007
Bond angles (°)	1.4
No. of protein atoms	3698
No. of inhibitor atoms	26
No. of water molecules	564
Average B factor (Å^2)	
All atoms	18
Protein atoms	17
Inhibitor atoms	14
Water molecules	27
Ramachandran analysis, residues in (%)	
Most favoured regions	90.9
Additionally allowed region	8.6
Generously allowed regions	0.5
Disallowed regions	0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the mean value for this reflection. ‡ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. $R_{\rm free}$ is calculated in the same manner as $R_{\rm cryst}$, except that it uses 5% of the data which were omitted from refinement.

2.4. Circular dichroism (CD)

CD spectra of hCA I, hCA II and SspCA were recorded using a Jasco J-715 spectropolarimeter equipped with a Peltier temperature-control system (Model PTC-423-S; Ascione *et al.*, 2012). Thermal unfolding curves were determined by recording the molar ellipticity at 210 nm using a scanning rate of 1 K min^{-1} in the temperature ranges 293–353 K for hCA I and hCA II and 293–378 K for SspCA. Experiments were performed in 10 m*M* phosphate buffer pH 7.5 for hCA I and hCA II and 100 m*M* bis-tris, 200 m*M* $MgCl_2$ pH 6.0 for SspCA. The protein concentrations were 0.6, 0.4 and 0.5 mg ml^{-1} for hCA I, hCA II and SspCA, respectively.

2.5. Multiple-angle light-scattering (MALS) analysis

The quaternary structure of recombinant SspCA was investigated by combining size-exclusion chromatography (SEC) with multiple-angle light scattering (MALS; Ascione *et al.*, 2012). Experiments were run at 0.5 ml min^{-1} in 100 m*M* bis-tris, 200 m*M* $MgCl_2$ pH 6.0; 150 µg enzyme was loaded onto a Superdex 10/30 size-exclusion column (GE Healthcare) connected to an ÄKTApurifier FPLC system, which in turn was connected to a refractive-index detector (Shodex RI 101) followed by a miniDAWN Treos light-scattering instrument (Wyatt Technology, USA). Data were processed using *ASTRA* 5.3.4.14 software (Wyatt Technology).

2.6. Crystallization and X-ray data collection

Crystallization experiments of SspCA were performed using both the unbound form and a complex with acetazolamide (AZM), which is a strong α-CA inhibitor. The SspCA–AZM complex was prepared by adding a fivefold molar excess of the inhibitor to 7.5 mg ml^{-1} protein solution in 20 m*M* Tris–HCl pH 8.3.

The search for initial crystallization conditions was performed using Crystal Screen, Crystal Screen 2 and Index from Hampton Research (Jancarik & Kim, 1991; Cudney *et al.*, 1994). The wells contained 500 µl precipitant solution and the drops were composed of 1 µl reservoir solution and 1 µl protein–complex solution at a concentration of 7.5 mg ml^{-1}. Crystals were only obtained in the case of the complex, using a reservoir solution consisting of 25%(*w*/*v*) polyethylene glycol 3350, 0.2 *M* magnesium chloride, 0.1 *M* bis-tris pH 5.5. Several parameters such as the buffer composition, pH and protein concentration were varied in order to improve the crystal quality. Optimal conditions for crystallization were achieved at the same complex concentration using a reservoir solution consisting of 20%(*w*/*v*) polyethylene glycol 3350, 0.2 *M* magnesium chloride, 0.1 *M* bis-tris pH 5.5 at 298 K.

Crystals appeared in the drops within 48 h and grew to maximum dimensions of 0.2 × 0.3 × 0.3 mm in about one week. A complete data set was collected to 1.80 Å resolution from a single crystal at a temperature of 100 K using a copper rotating-anode generator developed by Rigaku equipped with a Rigaku Saturn CCD detector. Prior to cryogenic cooling, the crystals were transferred into precipitant solution with the addition of 15%(*w*/*v*) glycerol. Data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals belonged to space group $P2_1$, with unit-cell parameters a = 48.50, b = 68.36, c = 74.69 Å, β = 106.03°. The Matthews coefficient ($V_{\rm M}$ = 2.29 Å^3 Da^{-1}) indicated that the crystallographic asymmetric unit contained two molecules, with a corresponding solvent content of 46.3%. The data-collection statistics are reported in Table 1.

2.7. Structure determination and refinement

The structure of SspCA was solved by the molecular-replacement technique with the program *AMoRe* (Navaza, 1994) using the crystallographic structure of α-CA from *N. gonorrhoeae* (NgCA; PDB entry 1kop; Huang *et al.*, 1998) as a model. The rotation and translation functions were calculated using data between 15.0 and 3.5 Å resolution. A one-body translation search using the centred-overlap function on the first 50 rotation solutions led to a single solution with a correlation coefficient of 0.163 and an R factor of 0.524. An n-body translation search carried out with the phased translation function, including a Patterson correlation refinement before each n-body translation search, led to the second

molecule being found. This improved the correlation coefficient and R factor to 0.314 and 0.486, respectively.

Refinement of the structure was performed with *CNS* (Brünger *et al.*, 1998) and model building was performed with *O* (Jones *et al.*, 1991). The initial refinement cycles were performed with the application of strict noncrystallographic symmetry constraints, which were gradually released. Inspection of the electron-density maps at various stages of crystallographic refinement showed clear evidence for one AZM inhibitor molecule bound to the active site of the enzyme. Introduction of an AZM molecule with full occupancy into the coordinate set was followed by several cycles of addition of water molecules, manual rebuilding, and positional and thermal B-factor refinement. Solvent molecules were only included in the model if they displayed electron density higher than 3σ in the $F_o - F_c$ maps and if they were at a hydrogen-bonding distance from the appropriate atoms. After their introduction into the coordinate set, they were retained if they presented a peak above 1.0σ in the $2F_o - F_c$ maps. The final model contained 3698 non-H atoms, 564 solvent molecules and 26 inhibitor atoms and had R_{free} and R_{cryst} values of 0.203 and 0.168, respectively, in the 20.00–1.80 Å resolution range.

All residues were well defined in the electron-density maps, except for the N-terminal region that includes the His tag and residue His1 in chain *A*. The refined model presents good geometry, with root-mean-square deviations (r.m.s.d.s) from ideal bond lengths and angles of 0.007 Å and 1.4°, respectively. The structure has good stereochemistry as tested using *PROCHECK* (Laskowski *et al.*, 1993). The most favoured and additionally allowed regions of the Ramachandran plot contained 90.9 and 8.6%, respectively, of the nonglycine residues. Data-refinement statistics are reported in Table 1. Coordinates and structure factors have been deposited in the Protein Data Bank (accession code 4g7a).

Figs. 4, 5, 6, 8 and 9 were generated using *PyMOL* (Schrödinger; http://www.pymol.org).

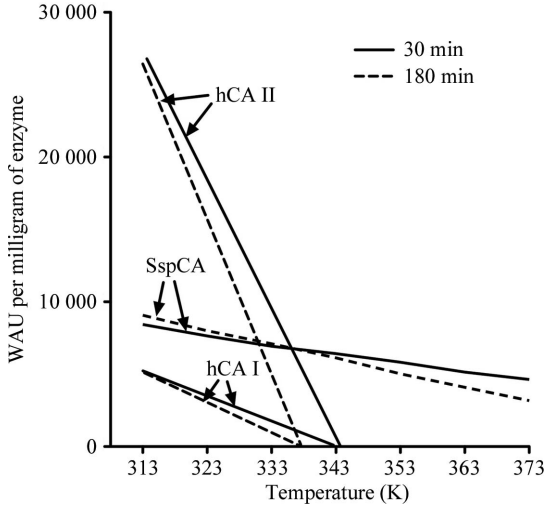


Figure 1
Thermostability of SspCA, hCA I and hCA II. The enzymes were incubated for 30 min (continuous lines) and 180 min (dotted lines) at the temperatures indicated on the x axis and were assayed using CO_2 as a substrate. Each point is the mean ± standard error of three independent determinations.

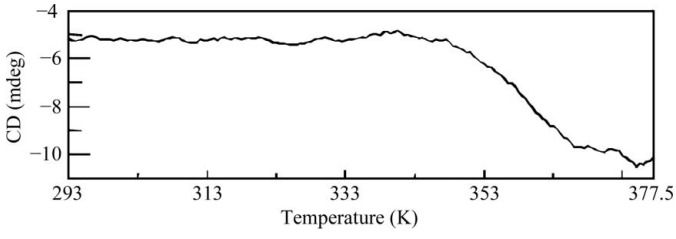


Figure 2
Thermal denaturation of SspCA monitored by far-UV CD at 210 nm.

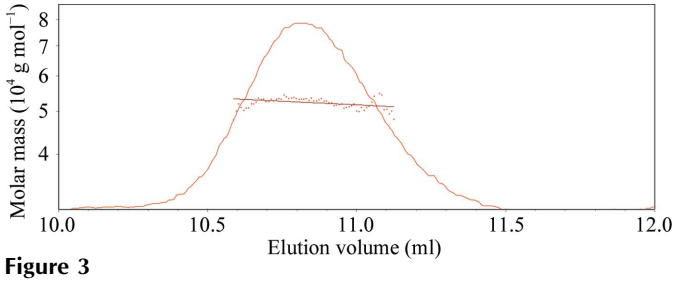


Figure 3
Analytical SEC–MALS of SspCA.

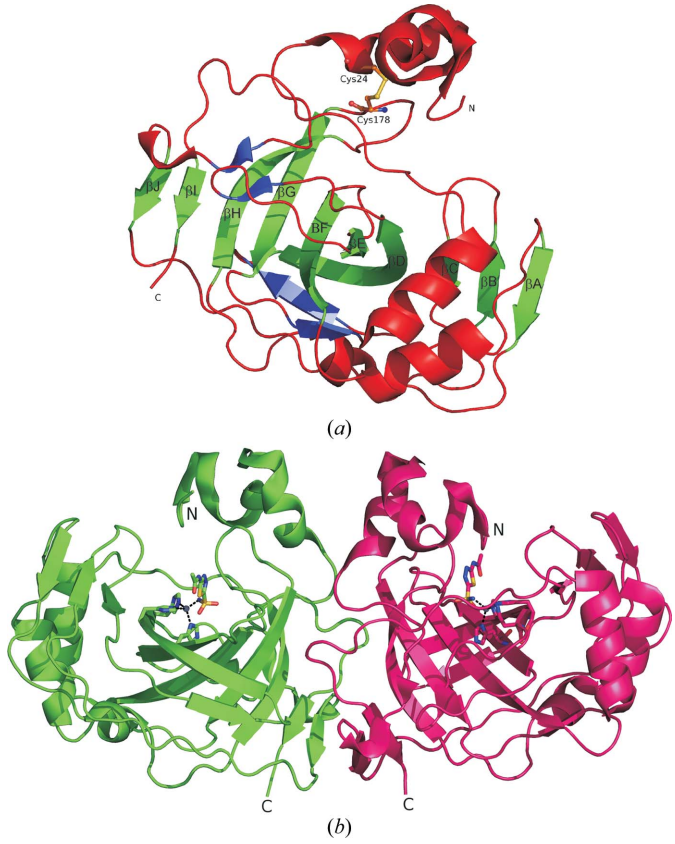


Figure 4
(*a*) Ribbon representation of the overall fold of SspCA, with the ten-stranded central β-sheet coloured green, additional β-strands coloured blue and helix regions coloured red. (*b*) SspCA dimer structure, with one monomer shown in green and the other in magenta. Zn^{2+}-ion coordination and the inhibitor AZM are also shown.

2.8. Surface calculations and analysis of noncovalent interactions

Ionic, hydrophobic, aromatic–aromatic, aromatic–sulfur and cation–π interactions were calculated using the program *PIC* (*Protein Interactions Calculator*; Tina *et al.*, 2007), while hydrogen bonds were calculated using *HBPLUS* (McDonald & Thornton, 1994) with the following default parameters: maximum D–A distance of 3.9 Å and maximum H–A distance of 2.5 Å; the minimum angle for D–H–A, D–A–AA and H–A–AA was 90°. In the calculation of ionic interactions, histidine residues coordinating the zinc ion were considered to be neutral. Accessible surface areas were calculated using the program *GRASP* (Nicholls *et al.*, 1991) with a 1.4 Å probe. Arg, Lys, Asp, Glu and His were defined as charged residues, Gly, Ser, Thr, Tyr, Gln, Asn and Cys as polar residues and Ala, Leu, Ile, Trp, Pro, Phe, Met and Val as hydrophobic residues. The secondary-structure percentage was calculated using *PROMOTIF* (Hutchinson & Thornton, 1996).

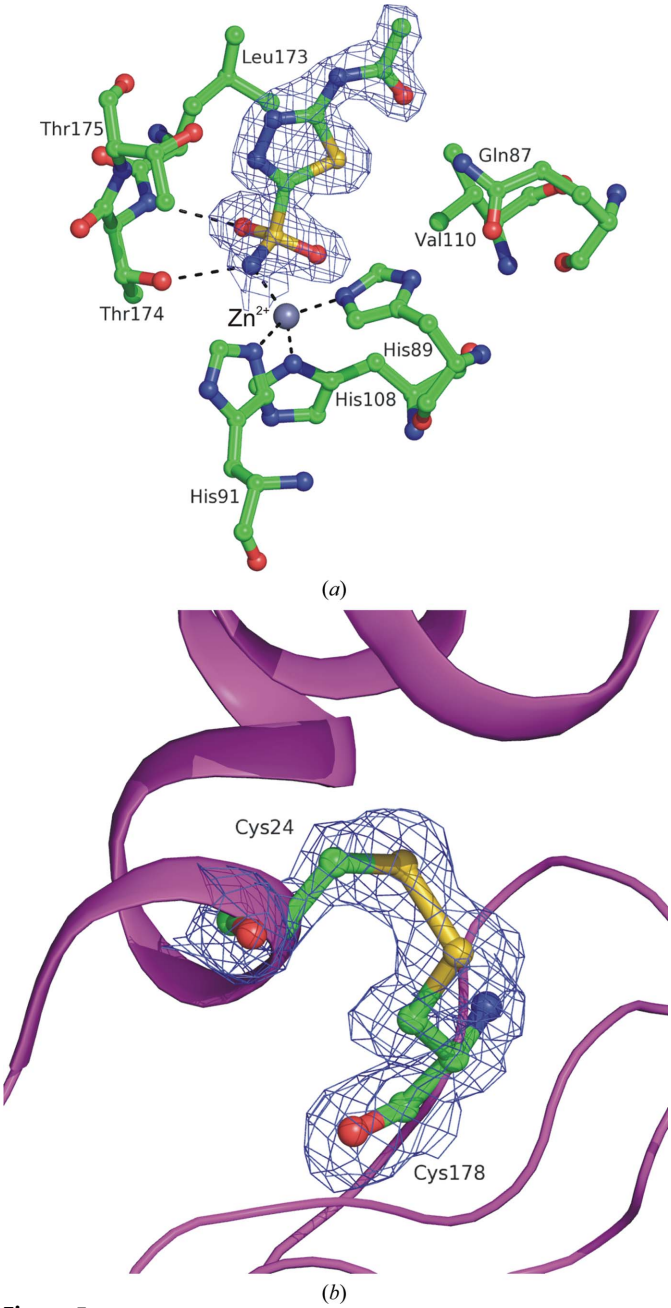


Figure 5
(*a*) Active-site region of the SspCA–AZM complex. The inhibitor is shown in association with a σ_A-weighted $|2F_o - F_c|$ simulated-annealing OMIT map (at 1.0σ). Hydrogen bonds, van der Waals interactions (distance of <4.0 Å) and the active-site Zn^{2+}-ion coordination are also shown. (*b*) σ_A-weighted $|2F_o - F_c|$ simulated-annealing OMIT map (at 1.0σ) for the SspCA intramolecular disulfide bridge.

3. Results and discussion

3.1. Catalytic activity and thermostability

The CO_2 hydratase activity of SspCA was measured after incubation at different temperatures and compared with those of two other mammalian α-CAs: hCA I and hCA II (Fig. 1).

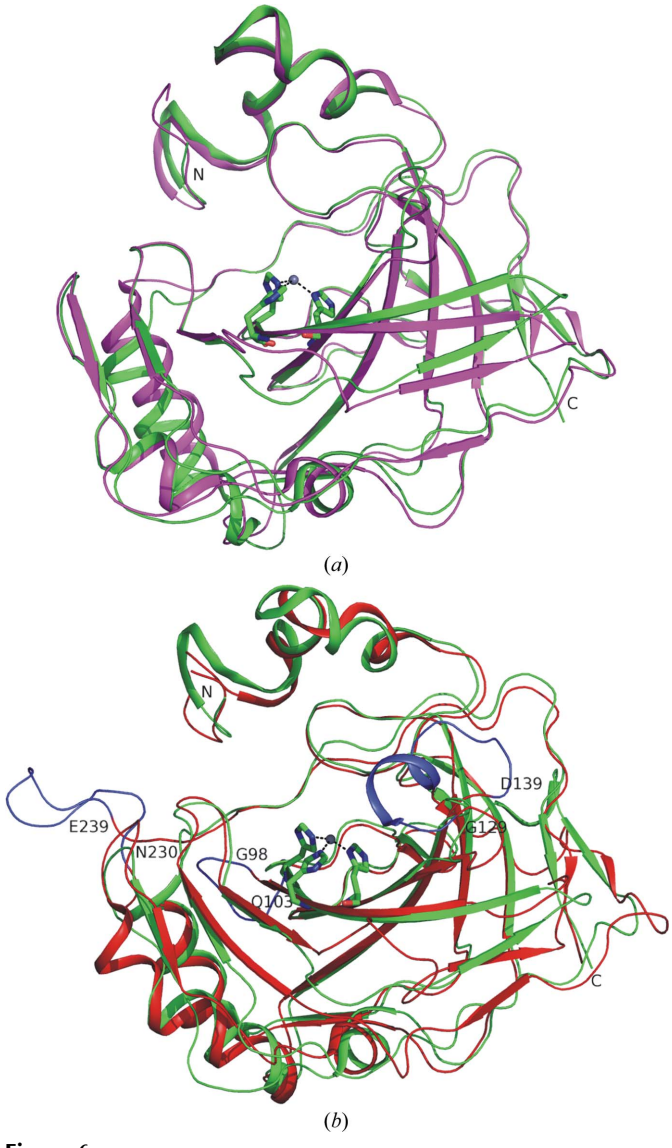


Figure 6
Structural superposition of SspCA (green) with (*a*) NgCA (magenta) and (*b*) hCA II (red). The insertions in hCA II with respect to SspCA are shown in blue.

Interestingly, while the mammalian enzymes were inactivated at temperatures higher than 343 K for all incubation times, SspCA was still active after incubation at 373 K for 3 h, highlighting a greater resistance to high temperature with respect to its mammalian homologues.

Changes in the far-UV CD signal at 210 nm were used to follow the thermal unfolding of SspCA, hCA I and hCA II. Because of the very high thermal stability of SspCA, it was not possible to obtain a complete thermal unfolding curve as the instrumental setup was unable to work at temperatures higher than 378 K. Surprisingly, a midpoint transition was observed at a temperature of 360.5 K with an increased negative ellipticity, suggesting an increase in β-structure, either by intra-subunit formation of β-sheets or β-turns or by the formation of aggregates *via* interchain cross-β interactions (Fig. 2; Dao *et al.*, 2011). However, this event was reversible as SspCA recovered its native spectrum after 10 min at room temperature (data not shown), thus confirming its thermally resistant nature. In contrast, hCA I and hCA II showed T_m values of 337.0 and 333.5 K, respectively.

3.2. Structural studies

Nonreducing SDS–PAGE coupled with light-scattering experiments revealed a noncovalent dimeric arrangement for SspCA (Fig. 3). Crystallization experiments were performed on the recombinant enzyme both in the unbound form and in an inhibitor-bound form. Acetazolamide (AZM), which is a strong α-CA inhibitor, was used for this purpose (Alterio *et al.*, 2012). Crystals were only obtained for the SspCA–AZM complex, using the hanging-drop vapour-diffusion method with PEG 3350 as a precipitant. They belonged to space group $P2_1$ and contained two molecules per asymmetric unit. The structure was solved by molecular replacement using the crystallographic coordinates of α-CA from *N. gonorrhoeae* (NgCA; PDB entry 1kop; 45.6% sequence identity) as the starting model (Huang *et al.*, 1998). The refinement statistics are summarized in Table 1.

The two independent molecules within the asymmetric unit overlapped well. Superimposition of the C$^\alpha$ positions of all 224 residues gave an r.m.s.d. value of 0.37 Å. For this reason, the subsequent discussion is based on one arbitrarily chosen molecule unless stated otherwise. SspCA appears as a compact globular protein of approximate dimensions 41 × 41 × 34 Å. Its structure presents the typical fold of human α-CAs characterized by a central ten-stranded β-sheet (Fig. 4*a*). As observed in the structures of other α-CAs (Alterio *et al.*, 2012), the active site is located in a deep conical cavity which extends from the protein surface to the centre of the molecule. The catalytic zinc ion is located at the bottom of this cavity and is tetrahedrally coordinated by three histidine residues (His89, His91 and His108) and by the N atom of the sulfonamide moiety of the inhibitor (Figs. 4*b* and 5*a*).

An intramolecular disulfide bond between Cys24 and Cys178 is present in this structure (Figs. 4*a* and 5*b*). Interestingly, this disulfide bridge is conserved in several other members of the human CA family such as CA IV

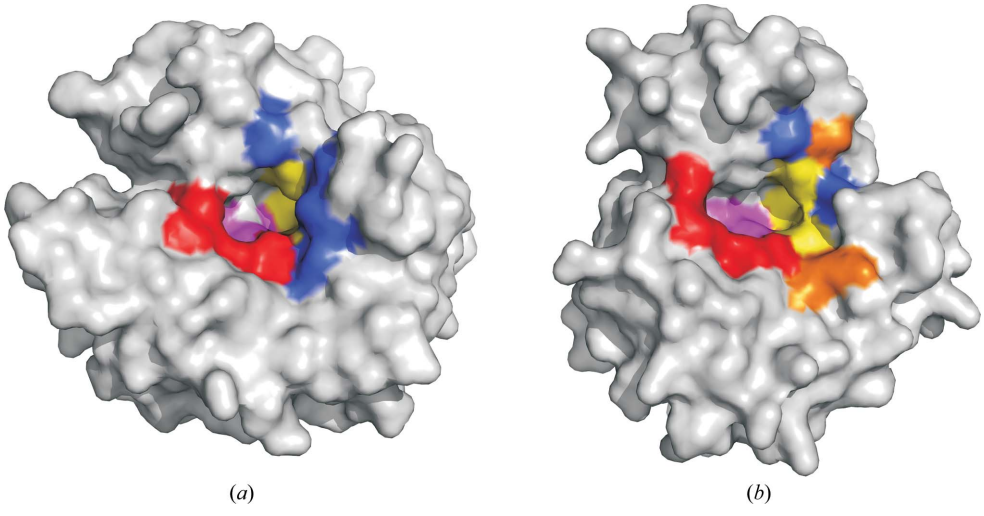
```
SspCA   2   -EWSYEGEKGPEHWAQLKPEFFWCKL-KNQSPINIDKKYKVKAN-LPKLNLYYKTAKESEVVNNGHTIQ
NgCA    5   THWGYTGHDSPESWGNLSEEFRLCSTGKNQSPVNITE--TVSGK-LPAIKVNYKP-SMVDVENNGHTIQ
hCA II  4    HWGYGKHNGPEHWHKD---FPIAKG-ERQSPVDIDTHTAKYDPSLKPLSVSYDQATSLRILNNGHAFN

SspCA  68   INIKE---DNTLNY--LGEKYQLKQFHFHT------PSEHTIEKKSYPLEIHFVHKTEDG---------
NgCA   70   VNYPEG--GNTLTV--NGRTYTLKQFHFHV------PSENQIKGRTFPMEAHFVHLDENK---------
hCA II 68   VEFDDSQDKAVLKGGPLDGTYRLIQFHFHWGSLDGQGSEHTVDKKKYAAELHLVHWNTKYGDFGKAVQQ
                                    * *                 *

SspCA  117  --KILVVGVMAKLGKTNKELDKILNVA--PAEE-GEKILDKNLNLNNLIPKDKRYMTYSGSLTTPPCTE
NgCA   120  --QPLVLAVLYEAGKTNGRLSSIWNVM--PMTA-GKVKLNQPFDASTLLPKRLKYYRFAGSLTTPPCTE
hCA II 138  PDGLAVLGIFLKVGSAKPGLQKVVDVLDSIKTKGKSADF-TNFDPRGLLPESLDYWTYPGSLTTPPLLE

SspCA  181  GVRWIVLKKPISISKQQLEKLKSV----------MVNPNNRPVQEINSRWIIEGF
NgCA   184  GVSWLVLKTYDHIDQAQAEKFTRA----------VGSENNRPVQPLNARVVIE--
hCA II 206  CVTWIVLKEPISVSSEQVLKFRKLNFNGEGEPEELMVDNWRPAQPLKNRQIKASF
```

Figure 7
Structure-based sequence alignment of SspCA with NgCA (PDB entry 1kop; Huang *et al.*, 1998) and hCA II (PDB entry 1ca2; Eriksson *et al.*, 1988). α-Helices, 3_{10}-helices and β-strands are highlighted in green, cyan and yellow, respectively. Catalytic histidines are marked with asterisks.



(*a*) (*b*)

Figure 8
Surface representation of (*a*) hCA II and (*b*) SspCA showing the active-site cavities of the two enzymes. In both enzymes catalytic histidines are coloured magenta and the CO_2-binding pocket is coloured yellow, while hydrophilic and hydrophobic regions which surround the catalytic site are coloured red and blue, respectively. Hydrophobic residues of hCA II which are not conserved in SspCA are coloured orange.

(Stams *et al.*, 1996), CA VI (Pilka *et al.*, 2012), CA IX (Alterio *et al.*, 2009), CA XII (Whittington *et al.*, 2001) and CA XIV (Whittington *et al.*, 2004), and in four other α-CAs for which crystal structures have been solved, namely NgCA (Huang *et al.*, 1998), CA II from *Dunaliella salina* (DsCA; Premkumar *et al.*, 2005), αCA1 from *Chlamydomonas reinhardtii* (CrCA; Suzuki *et al.*, 2011) and α-CA from *Aspergillus oryzae* (AoCA; Cuesta-Seijo *et al.*, 2011).

As mentioned previously, NgCA is the only other bacterial α-CA with a known three-dimensional structure. Structural comparison between SspCA and NgCA shows a substantial degree of three-dimensional similarity as expected on the basis of their sequence identity. Indeed, the r.m.s.d. value calculated for the superimposition of the corresponding 219 C^{α} atoms is 1.2 Å. Structural superposition of the two proteins reveals that the majority of the secondary-structural elements are conserved and no significant insertions or deletions are observed (Figs. 6*a* and 7). In contrast, structural comparison between SspCA and hCA II (Eriksson *et al.*, 1988), which was chosen as a representative member of the human α-CA family, revealed some important differences. In particular, SspCA is much more compact than hCA II (Fig. 6*b*). Although the central ten-stranded β-sheet as well as most of the remaining secondary-structural elements are conserved in the two proteins, three large insertions are present in the human enzyme with respect to the bacterial enzyme (Figs. 6*b* and 7). These insertions correspond to the protein regions 98–103, 230–239 and 129–139 (hCA II numbering). The first two of these insertions are located on the protein surface far from the active site and thus probably do not influence either the catalytic activity or the binding of inhibitors and activators, whereas the insertion 129–139 is located on the upper edge of the active site and contains the helix 130–135, which has been reported to be crucial for ligand recognition in several studies

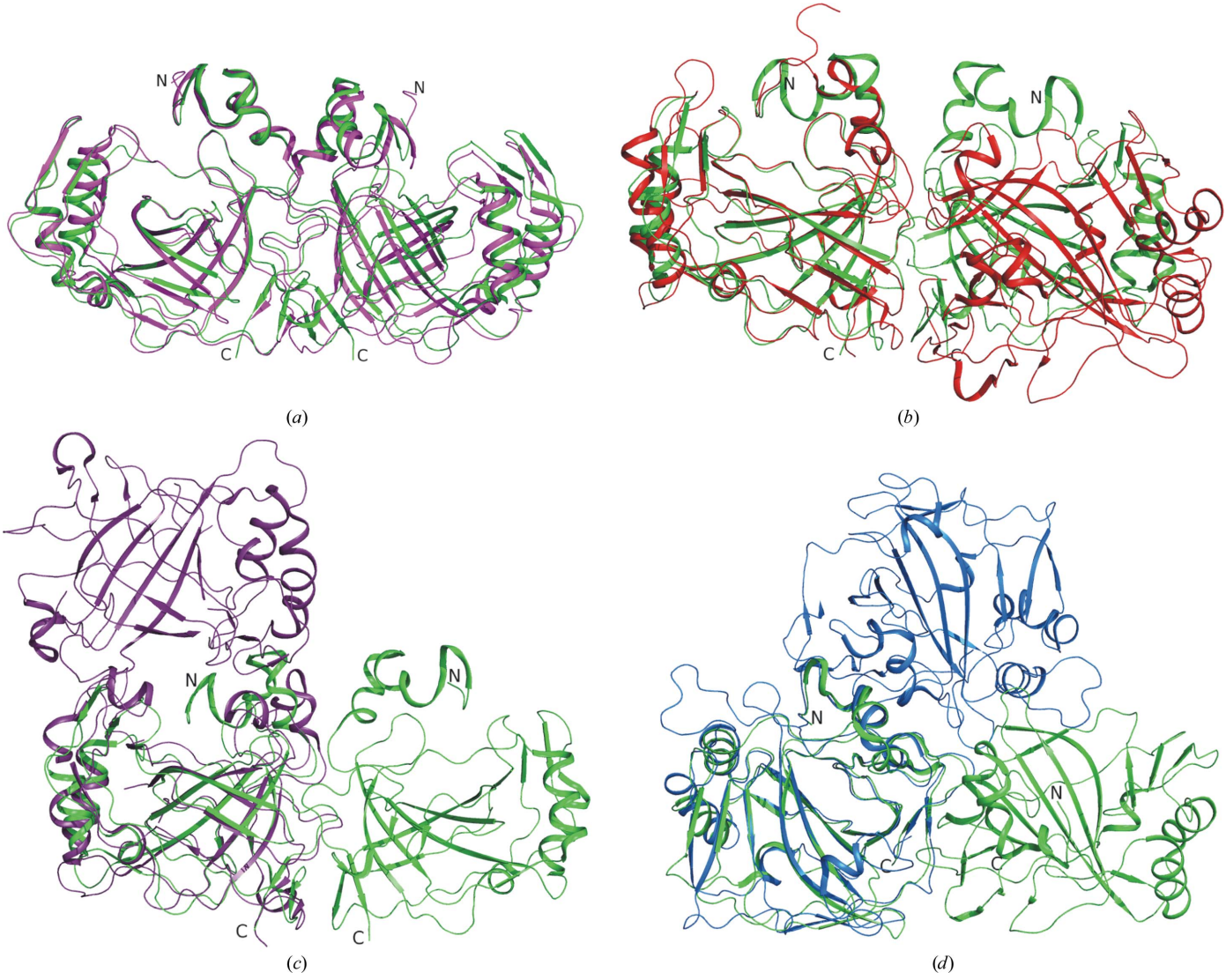


Figure 9
Structural superposition of the dimeric structure of SspCA (green) with selected α-CA dimeric structures: (*a*) NgCA (magenta), (*b*) hCA IX (red), (*c*) hCA VI (violet) and (*d*) hCA XII (blue). The NgCA dimer was generated by running the *PISA* program (http://www.ebi.ac.uk) on the crystallographic coordinates of PDB entry 1kop and choosing the most stable potential dimer.

Table 2
Comparison of the molecular compactness, hydrophilicity of the molecular surface and noncovalent interactions of SspCA, hCA I and hCA II.

	SspCA†	hCA I	hCA II
PDB code	4g7a	1hcb‡	1ca2§
No. of residues in monomer	225	260	258
Charged residues¶ (%)	31	24	29
Polar residues¶ (%)	32	37	32
Hydrophobic residues¶ (%)	37	39	39
Aromatic residues¶ (%)	12	14	15
Oligomeric state	Dimer	Monomer	Monomer
Accessible surface area (Å^2)			
Total	20771	11591	11676
Charged residues††	12134 (58%)	4726 (41%)	6012 (52%)
Polar residues††	5141 (25%)	3985 (34%)	3009 (26%)
Hydrophobic residues††	3495 (17%)	2880 (25%)	2655 (23%)
Secondary-structure content (%)	50.4	44.6	43.7
α-Helices	9.4	8.5	6.6
β-Strands	33.0	28.3	28.9
3_{10}-helices	8.0	7.8	8.2
Hydrogen bonds‡‡	399 (0.9)	232 (0.9)	230 (0.9)
Main chain–main chain‡‡	202 (0.4)	122 (0.5)	120 (0.5)
Main chain–side chain‡‡	132 (0.3)	71 (0.3)	76 (0.3)
Side chain–side chain‡‡	65 (0.1)	39 (0.2)	34 (0.1)
Ion pairs (4 Å)‡‡	15 (0.03)	7 (0.03)	11 (0.04)
Ion pairs (6 Å)‡‡	40 (0.09)	16 (0.06)	19 (0.07)
No. of ionic networks (4 Å)			
Two-membered	4/5/9§§	5	11
Three-membered	0/1/1§§	1	0
Four-membered	0/0/0§§	0	0
Five-membered	1/0/1§§	0	0
No. of ionic networks (6 Å)			
Two-membered	5/8/12§§	5	13
Three-membered	3/3/4§§	4	3
Four-membered	0/0/1§§	1	0
Five-membered	0/0/0§§	0	0
Six-membered	0/1/2§§	0	0
Seven-membered	1/0/1§§	0	0
Hydrophobic interactions (5 Å)‡‡	422 (0.9)	200 (0.8)	228 (0.9)
Aromatic–aromatic interactions‡‡	23 (0.05)	16 (0.06)	16 (0.06)
Aromatic–sulfur interactions within 5.3 Å‡‡	4 (0.009)	2 (0.008)	1 (0.004)
Cation–π interactions (6 Å)‡‡	15 (0.03)	6 (0.02)	7 (0.03)

† All calculations for SspCA were performed by considering the dimeric structure except where stated otherwise. ‡ Kumar & Kannan (1994). § Eriksson *et al.* (1988). ¶ Charged residues: Arg, Lys, Asp, Glu and His. Polar residues: Gly, Ser, Thr, Tyr, Gln, Asn and Cys. Hydrophobic residues: Ala, Leu, Ile, Trp, Pro, Phe, Met and Val. Aromatic residues: Trp, Phe, Tyr and His. †† Values in parentheses represent the percentage of accessible surface area contributed by the specified residues. ‡‡ Values in parentheses represent the number of hydrogen bonds, ion pairs and noncovalent interactions normalized to the number of protein residues, considering a dimer for SspCA and a monomer for hCA I and hCA II. §§ Monomer *A*/monomer *B*/dimer.

of the human α-CA family (Alterio *et al.*, 2012). The absence of this region in SspCA makes its active-site cavity larger than that of hCA II (Fig. 8) and also those of other α-CAs in general. Limiting the comparison to the active-site region, several other important differences exist between SspCA and hCA II. The active-site cavity of hCA II, as in all human CA isoforms, consists of two very different regions: one delimited essentially by hydrophilic residues and the other by hydrophobic residues (Sjöblom *et al.*, 2009; Domsic *et al.*, 2008; Alterio *et al.*, 2012). Both regions have been reported to be important for catalysis; indeed, the hydrophilic region allows the formation of a well ordered solvent network which assists the rate-limiting step of the catalytic reaction, *i.e.* the proton-transfer reaction from the zinc-bound water molecule to the bulk solvent (Tu *et al.*, 1989; Silverman & McKenna, 2007; Roy & Taraphder, 2007; Fisher *et al.*, 2007; Alterio *et al.*, 2012). In contrast, the hydrophobic region is involved in CO_2 binding and in ligand recognition (Sjöblom *et al.*, 2009; Domsic *et al.*, 2008; Alterio *et al.*, 2012). Comparison of these regions in SspCA and hCA II shows that whereas the catalytic residues, the hydrophilic region and all of the residues involved in CO_2 binding are conserved in the bacterial enzyme with respect to hCA II, the hydrophobic region involved in inhibitor recognition is less extended in SspCA owing to the lack of helix 130–135 and the substitution of some hydrophobic residues by hydrophilic residues (Figs. 7 and 8). Together, this information can be used for the rational design of SspCA activators.

One of the most intriguing features of the SspCA structure is related to its quaternary structure. Indeed, the two independent molecules in the asymmetric unit form a dimer that is characterized by a large interface area that extends over 2300 Å^2 (Fig. 4*b*) and that is stabilized by several hydrogen bonds and hydrophobic interactions. The observations that both active sites are completely accessible to the substrate and that this dimeric arrangement is also shown by light-scattering experiments strongly suggest that SspCA acts as a physiological dimer. This is a very interesting result considering that α-CAs were long considered to be monomeric enzymes. However, in recent years this conviction has been surmounted by the observation of dimeric arrangements for several members of the α-class. In particular, NgCA (Huang *et al.*, 1998), CrCA (Suzuki *et al.*, 2011) and AoCA (Cuesta-Seijo *et al.*, 2011) and three human α-CAs, namely hCA VI (Pilka *et al.*, 2012), hCA IX (Alterio *et al.*, 2009) and hCA XII (Whittington *et al.*, 2001), have been reported to be dimeric. However, a structural comparison showed that the all of these α-CAs present a dimeric structure that is quite different from that of SspCA, with the exception of NgCA, the dimer of which is quite superimposable on that of SspCA (Fig. 9). This evidence suggests that this dimeric arrangement is typical of bacterial α-CAs; however, further experimental studies are necessary in order to prove this hypothesis.

3.3. Structural determinants of thermostability

Many efforts have been made in recent years to develop strategies for enzyme stabilization by comparing the high-resolution X-ray structures of thermostable proteins with

those of their mesophilic counterparts. Such studies have highlighted that protein stability commonly depends on a combination of factors which are related to each other (Matsui & Harata, 2007). In particular, a larger hydrophobic core, increased compactness of the protein, a greater number of ion pairs and hydrogen bonds, and an increased number of charged residues on the protein surface have been found to be the main factors responsible for protein thermostability. However, the contribution of each of these factors varies depending on the protein family (Matsui & Harata, 2007; Haney *et al.*, 1999). SspCA is the first 'extremo-α-CA' to be structurally characterized. In this paper, we have shown the capability of SspCA to remain active after heating to 373 K, in contrast to its mammalian homologues hCA I and hCA II, which were completely inactivated after heating to a temperature above 343 K. Thus, the availability of structural data for these three proteins provided us with the opportunity to identify the possible structural determinants responsible for thermostability in this enzyme family. A number of parameters which have been reported in the literature to be important for themostability were compared between SspCA, hCA I and hCA II (Table 2). Analysis of these parameters showed that a higher content of secondary-structural elements (enhanced compactness), associated with an increase in the accessible surface area contributed by charged residues, are the main features that distinguish SspCA from its mesophilic homologues. Surprisingly, the greater number of charged residues in SspCA did not correspond to a greater number of ion pairs compared with hCA I and hCA II when a cutoff distance of 4.0 Å was used. However, using a cutoff of 6.0 Å an increase in the number of ion pairs in SspCA compared with hCA I and hCA II was observed. This observation suggests that long-range ionic interactions might be important for the thermal stability of SspCA. When ionic networks are also taken into account, other important considerations can be made. Indeed, using cutoff distances of both 4.0 and 6.0 Å, SspCA presents a greater number of ionic networks formed by larger numbers of residues compared with hCA I and hCA II (see Table 2). Ionic networks are generally considered to be more favourable than the equivalent number of isolated ion pairs (Yip *et al.*, 1998); thus, this could be the strategy adopted by SspCA to increase its thermal stability. Mutagenesis studies are currently under way in order to test this hypothesis.

In conclusion, we have reported the first X-ray structure of a thermostable α-CA. Two main notable findings are outlined from this study: (i) bacterial α-CAs are likely to present a conserved dimeric structure and (ii) high structural compactness, together with an increased number of charged residues on the protein surface and a greater number of ionic networks, seem to be the key factors involved in the thermostability of this enzyme. These findings are of extreme importance since they provide a structural basis for the understanding of the mechanisms responsible for thermal stability in the α-class of the carbonic anhydrase family for the first time. Thus, they may help in engineering highly active α-CAs such as hCA II in order to obtain mutated enzymes with enhanced thermostability to be used under the harsh conditions of the biotechnological processes described in §1.

References

Adams, M. W. W. (1993). *Annu. Rev. Microbiol.* **47**, 627–658.
Adams, M. W. W. & Kelly, R. M. (1995). *Chem. Eng. News*, **73**, 32–42.
Adams, M. W. W., Perler, F. B. & Kelly, R. M. (1995). *Biotechnology*, **13**, 662–668.
Adler, L., Brundell, J., Falkbring, S. O. & Nyman, P. O. (1972). *Biochim. Biophys. Acta*, **284**, 298–310.
Alterio, V., Di Fiore, A., D'Ambrosio, K., Supuran, C. T. & De Simone, G. (2012). *Chem. Rev.* **112**, 4421–4468.
Alterio, V., Hilvo, M., Di Fiore, A., Supuran, C. T., Pan, P., Parkkila, S., Scaloni, A., Pastorek, J., Pastorekova, S., Pedone, C., Scozzafava, A., Monti, S. M. & De Simone, G. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 16233–16238.
Ascione, G., de Pascale, D., De Santi, C., Pedone, C., Dathan, N. A. & Monti, S. M. (2012). *Biochem. Biophys. Res. Commun.* **420**, 542–546.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Capasso, C., De Luca, V., Carginale, V., Cannio, R. & Rossi, M. (2012). *J. Enzyme Inhib. Med. Chem.* **27**, 892–897.
Chirică, L. C., Elleby, B., Jonsson, B. H. & Lindskog, S. (1997). *Eur. J. Biochem.* **244**, 755–760.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Cuesta-Seijo, J. A., Borchert, M. S., Navarro-Poulsen, J. C., Schnorr, K. M., Mortensen, S. B. & Lo Leggio, L. (2011). *FEBS Lett.* **585**, 1042–1048.
Dao, K. K., Pey, A. L., Gjerde, A. U., Teigen, K., Byeon, I.-J. L., Døskeland, S. O., Gronenborn, A. M. & Martinez, A. (2011). *PLoS One*, **6**, e17602.
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.
Eriksson, A. E., Jones, T. A. & Liljas, A. (1988). *Proteins*, **4**, 274–282.
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.
Fisher, S. Z., Maupin, C. M., Budayova-Spano, M., Govindasamy, L., Tu, C., Agbandje-McKenna, M., Silverman, D. N., Voth, G. A. & McKenna, R. (2007). *Biochemistry*, **46**, 2930–2937.
Gomes, J. & Steiner, W. (2004). *Food Technol. Biotechnol.* **42**, 223–235.
Greer, T., Bedelbayev, A., Igreja, J. M., Gomes, J. F. & Lie, B. (2010). *Environ. Technol.* **31**, 107–115.
Haney, P. J., Stees, M. & Konisky, J. (1999). *J. Biol. Chem.* **274**, 28453–28458.
Huang, S., Xue, Y., Sauer-Eriksson, E., Chirica, L., Lindskog, S. & Jonsson, B. H. (1998). *J. Mol. Biol.* **283**, 301–310.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kumar, V. & Kannan, K. K. (1994). *J. Mol. Biol.* **241**, 226–232.
Kumar, S. & Nussinov, R. (2001). *Cell. Mol. Life Sci.* **58**, 1216–1233.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, S.-W., Park, S.-B., Jeong, S.-K., Lim, K. -S., Lee, S.-H. & Trachtenberg, M. C. (2010). *Micron*, **41**, 273–282.
Matsui, I. & Harata, K. (2007). *FEBS J.* **274**, 4012–4022.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.

Migliardini, F., De Luca, V., Carginale, V., Rossi, M., Corbo, P., Supuran, C. T. & Capasso, C. (2013). *J. Enzyme Inhib. Med. Chem.*, doi:10.3109/14756366.2012.761608.
Mignardi, S., De Vito, C., Ferrini, V. & Martin, R. F. (2011). *J. Hazard. Mater.* **191**, 49–55.
Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). *Protein Sci.* **4**, 2138–2148.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Nishimori, I., Minakuchi, T., Morimoto, K., Sano, S., Onishi, S., Takeuchi, H., Vullo, D., Scozzafava, A. & Supuran, C. T. (2006). *J. Med. Chem.* **49**, 2117–2126.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pilka, E. S., Kochan, G., Oppermann, U. & Yue, W. W. (2012). *Biochem. Biophys. Res. Commun.* **419**, 485–489.
Premkumar, L., Greenblatt, H. M., Bageshwar, U. K., Savchenko, T., Gokhman, I., Sussman, J. L. & Zamir, A. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 7493–7498.
Roy, A. & Taraphder, S. (2007). *J. Phys. Chem. B*, **111**, 10563–10576.
Savile, C. K. & Lalonde, J. J. (2011). *Curr. Opin. Biotechnol.* **22**, 818–823.
Silverman, D. N. & McKenna, R. (2007). *Acc. Chem. Res.* **40**, 669–675.
Sjöblom, B., Polentarutti, M. & Djinovic Carugo, K. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 10609–10613.
Stams, T., Nair, S. K., Okuyama, T., Waheed, A., Sly, W. S. & Christianson, D. W. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13589–13594.
Suzuki, K., Yang, S.-Y., Shimizu, S., Morishita, E. C., Jiang, J., Zhang, F., Hoque, M. M., Sato, Y., Tsunoda, M., Sekiguchi, T. & Takénaka, A. (2011). *Acta Cryst.* D**67**, 894–901.
Tina, K. G., Bhadra, R. & Srinivasan, N. (2007). *Nucleic Acids Res.* **35**, W473–W476.
Tu, C., Silverman, D. N., Forsman, C., Jonsson, B. H. & Lindskog, S. (1989). *Biochemistry*, **28**, 7913–7918.
Unsworth, L. D., van der Oost, J. & Koutsopoulos, S. (2007). *FEBS J.* **274**, 4044–4056.
Vullo, D., De Luca, V. D., Scozzafava, A., Carginale, V., Rossi, M., Supuran, C. T. & Capasso, C. (2013). *Bioorg. Med. Chem.* **15**, 1534–1538.
Whittington, D. A., Grubb, J. H., Waheed, A., Shah, G. N., Sly, W. S. & Christianson, D. W. (2004). *J. Biol. Chem.* **279**, 7223–7228.
Whittington, D. A., Waheed, A., Ulmasov, B., Shah, G. N., Grubb, J. H., Sly, W. S. & Christianson, D. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 9545–9550.
Yip, K. S. P., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). *Eur. J. Biochem.* **255**, 336–346.