Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The unique structure of carbonic anhydrase αCA1 from *Chlamydomonas reinhardtii*

Kaoru Suzuki,[a] Shi-Yuan Yang,[a] Satoru Shimizu,[b] Ella Czarina Morishita,[b]‡ Jiandong Jiang,[c] Fang Zhang,[c] Md. Mominul Hoque,[d,e] Yoshiteru Sato,[b]§ Masaru Tsunoda,[c] Takeshi Sekiguchi[a] and Akio Takénaka[b,c,d]*

[a]College of Science and Engineering, Iwaki-Meisei University, Chuodai-iino, Iwaki, Fukushima 970-8551, Japan, [b]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuda, Midori-ku, Yokohama 226-8501, Japan, [c]Graduate School of Science and Engineering, Iwaki-Meisei University, Chuodai-iino, Iwaki, Fukushima 970-8551, Japan, [d]Faculty of Pharmacy, Iwaki-Meisei University, Chuodai-iino, Iwaki, Fukushima 970-8551, Japan, and [e]Department of Biochemistry and Molecular Biology, Rajshahi University, Rajshahi, Bangladesh

‡ Present address: Systems and Structural Biology Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.
§ Present address: Faculty of Sciences, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-0053, Japan.

Correspondence e-mail: atakenak@bio.titech.ac.jp

Received 4 July 2011
Accepted 13 August 2011

PDB Reference: *Cr*-αCA1, 3b1b.

Chlamydomonas reinhardtii α-type carbonic anhydrase (*Cr*-αCA1) is a dimeric enzyme that catalyses the interconversion of carbon dioxide and carbonic acid. The precursor form of *Cr*-αCA1 undergoes post-translational cleavage and N-glycosylation. Comparison of the genomic sequences of precursor *Cr*-αCA1 and other αCAs shows that *Cr*-αCA1 contains a different N-terminal sequence and two insertion sequences. A 35-residue peptide in one of the insertion sequences is deleted from the precursor during maturation. The crystal structure of the mature form of *Cr*-αCA1 has been determined at 1.88 Å resolution. Each subunit is cleaved into the long and short peptides, but they are linked together by a disulfide bond. The two subunits are linked by a disulfide bond. N-Glycosylations occur at three asparagine residues and the attached N-glycans protrude into solvent regions. The subunits consist of a core β-sheet structure composed of nine β-strands. At the centre of the β-sheet is the catalytic site, which contains a Zn atom bound to three histidine residues. The amino-acid residues around the Zn atom are highly conserved in other monomeric and dimeric αCAs. The short peptide runs near the active site and forms a hydrogen bond to the zinc-coordinated residue in the long chain, suggesting an important role for the short peptide in *Cr*-αCA1 activity.

1. Introduction

Carbonic anhydrase (CA), an enzyme that catalyses the rapid interconversion of carbon dioxide and carbonic acid, is widely distributed in all kingdoms of life (Chegwidden, 2000; Tripp *et al.*, 2001; Mitra *et al.*, 2004). CAs are grouped into three classes (αCA, βCA and γCA) depending on architectural differences (Hewett-Emmett & Tashian, 1996), but they have a similar function and most of them contain a Zn atom in their active site. Therefore, it has been suggested that the three classes have evolved independently from different ancestor genes to create a similar enzyme active site (Mitra *et al.*, 2004; Hewett-Emmett & Tashian, 1996).

CAs form various subunit assemblies. Originally, αCAs were found as monomeric enzymes from humans, animals, algae and prokaryotes (Chegwidden, 2000; Tripp *et al.*, 2001), whereas the αCA from the unicellular green alga *Chlamydomonas reinhardtii* is dimeric (Fujiwara *et al.*, 1990). Recently, many monomeric, dimeric and higher oligomeric states of αCAs have been reported (Whittington *et al.*, 2001; Hilvo *et al.*, 2008; Alterioa *et al.*, 2009; Cuesta-Seijo *et al.*, 2011). γCAs are trimeric (Alber & Ferry, 1996), while βCAs are generally dimeric but can also be tetrameric, hexameric or octameric

upon interaction of dimers (Covarrubias *et al.*, 2006; Strop *et al.*, 2001; Kimber & Pai, 2000; Guilloton *et al.*, 1992; Smith & Ferry, 1999; Hiltonen *et al.*, 1998; Rumeau *et al.*, 1996; Tobin, 1970; Kisiel & Graf, 1972; Kandel *et al.*, 1978).

In addition, many CA isozymes belonging to the different classes have been found in the various tissues of organisms (Chegwidden, 2000). For instance, three *Cr*-αCA isozymes and five *Cr*-βCA isozymes have been identified in *Chlamydomonas reinhardtii*. *Cr*-αCA1 and *Cr*-αCA2 coexist[1] in the periplasm (Coleman & Grossman, 1984; Kimpel *et al.*, 1983; Yagawa *et al.*, 1986) and *Cr*-αCA3 and *Cr*-βCA8 are found in the chloroplast and mitochondria (Karlsson *et al.*, 1998; van Hunnik *et al.*, 2001; Villarejo *et al.*, 2002), while *Cr*-βCA6 is localized in the thylakoid lumen (Parker *et al.*, 2008). *Cr*-αCA1 was first identified (Coleman & Grossman, 1984; Yang *et al.*, 1985), followed by *Cr*-αCA2, and the two αCAs were shown to have complementary functions. In the presence of light, the major CA under low-CO_2 conditions is *Cr*-αCA1, while that under high-CO_2 conditions is *Cr*-αCA2 (Fujiwara *et al.*, 1990).

After extensive and intensive studies, a number of structures of human αCA isozymes with no modifications have been reported. However, *Cr*-αCA1 differs from these in that it undergoes post-translational glycosylation at selected asparagine residues and cleavage into two peptides (Ishida *et al.*, 1993). These modifications may be essential for some biological processes. In the present study, we performed a comparative sequence analysis of the *Cr*-αCA1 gene and other αCA genes. It is noteworthy that the deletion of 35 residues between residues 306 and 340 in precursor *Cr*-αCA1 does not occur in the corresponding regions of other genes and that the other αCAs, with the exceptions of human αCA1 (Kondo *et al.*, 1987) and rat αCA4 (Waheed *et al.*, 1992), are not known to be glycosylated. We propose that αCAs can be further divided into three subclasses according to the types of modifications. In addition, we have confirmed that *Cr*-αCA1 exists as a dimeric enzyme as reported, but that it behaves, in part, as higher oligomeric states do. In order to obtain insight into the structural and the functional contributions of N-glycosylation, peptide cleavage and higher assembly formation, we performed X-ray crystallographic analysis of the mature form of *Cr*-αCA1 and determined its structure at 1.88 Å resolution.

2. Materials and methods

2.1. Sequence alignment

The amino-acid sequence of *Cr*-αCA1 and those of αCAs from other organisms were retrieved from the KEGG database (http://www.genome.jp/kegg/) by sequence-homology searching using the program *BLAST* (http://www.ddbj.nig.ac.jp/index-e.html). To analyze their primary-structure relationship, the selected sequences were compared using the program *ClustalX* (Thompson *et al.*, 1997).

[1] The two corresponding genes are aligned in tandem in the genome (Kondo *et al.*, 1987).

2.2. Crystallization, X-ray data collection and structure determination

Mature *Cr*-αCA1 was isolated and purified from cultivated wild-type *C. reinhardtii* cells and crystallized according to the method described previously (Suzuki *et al.*, 2010). For the X-ray experiment, a *Cr*-αCA1 crystal was soaked for 30 s in reservoir solution containing 20% glycerol and flash-frozen in a CryoLoop (Hampton Research). Diffraction data sets were collected on the AR-NW12 beamline at the Photon Factory (Tsukuba, Japan). To apply multiple anomalous dispersion (MAD/SAD) methods for phase estimation, the anomalous scattering factors (f' and f'') of the Zn atoms bound in the crystal were derived from XAFS data obtained prior to data collection using the program *CHOOCH* (Evans & Pettifer, 2001). Three wavelengths were chosen and used for data collection: 1.28254 Å (peak), 1.28297 Å (edge) and 1.00 Å (remote). The diffraction patterns were recorded at 100 K on a Quantum 210 CCD detector positioned 86.3, 86.3 and 129.8 mm from the crystal for the peak, edge and remote wavelengths, respectively. Each frame was obtained with an exposure time of 10 s and an oscillation step of 1° over the range 0–180°. Bragg spots were indexed and intensities were estimated by integrating around the spots. The diffraction data were processed at 2.10, 2.00 and 1.88 Å resolution for the peak, edge and remote data, respectively, although weak diffraction spots were observed beyond these limits. Intensity data were then scaled and merged using the *HKL*-2000 package (Otwinowski & Minor, 1997). These data were further converted to structure-factor amplitudes using *TRUNCATE* from the *CCP*4 suite (Winn *et al.*, 2011).

The MAD method was applied to solve the phase problem with the programs *SHARP*/*autoSHARP* (Vonrhein *et al.*, 2007) and *ARP*/*wARP* (Perrakis *et al.*, 1999) using the MAD data sets. All of the atomic parameters of the constructed structural model were iteratively refined using the program *REFMAC*5 (Murshudov *et al.*, 2011) and the graphics program *Coot* (Emsley & Cowtan, 2004). The structure was validated with a Ramachandran plot using the program *PROCHECK* (Laskowski *et al.*, 1993).

2.3. Molecular-size estimation

The molecular size of purified *Cr*-αCA1 was estimated from the final gel-filtration purification step using commercially available low- and high-molecular-weight gel-filtration calibration kits (GE Healthcare). A broad-range marker (APRO Science) was used to calibrate the molecular weights. Native PAGE was performed using a Compact PAGE (Atto Co. AE-7300) with 10%(*w*/*v*) acrylamide gel (Atto Co. C10L). Protein solution containing 20 m*M* Tris–HCl pH 8.0, 1 mM EDTA, 0.1% BPB and 10% glycerol was loaded, with the running conditions being a constant electric current of 20 mA and 150 V for about 90 min. The gel was calibrated using the High-Molecular Weight Calibration kit for Native Electrophoresis (GE Healthcare). Protein bands were stained with Coomassie Brilliant Blue R-350.

Table 1
Statistics of the observed intensities, crystallographic data and structure refinement.

Values in parentheses are for the last shell.

	Peak	Edge	Remote
Data collection			
Wavelength (Å)	1.28254	1.28297	1.00
Resolution (Å)	50–2.00 (2.07–2.00)	50–2.10 (2.18–2.10)	50–1.88 (1.95–1.88)
Space group	$P6_5$	$P6_5$	$P6_5$
Unit-cell parameters (Å)	$a = b = 134.3$, $c = 120.2$	$a = b = 134.4$, $c = 120.2$	$a = b = 134.2$, $c = 120.1$
Observed reflections	885376	762127	1055288
Unique reflections	82859	71757	99548
Completeness (%)	99.9 (100)	99.9 (100)	99.9 (100)
Multiplicity	10.7 (10.2)	10.6 (10.1)	10.6 (10.0)
$\langle I/\sigma(I)\rangle$	57.5 (10.0)	57.6 (11.5)	60.0 (8.7)
$R_{\rm merge}$† (%)	8.1 (30.2)	8.1 (27.1)	8.2 (31.4)
Structure refinement			
Resolution (Å)			41.3–1.88
Reflections used			98307
R‡ (%)			18.9
$R_{\rm free}$§ (%)			21.8
R.m.s.d. bond lengths (Å)			0.02
R.m.s.d. angles (°)			1.5
Protein atoms			4896
Water molecules			888
SO_4^{2-} ions			3
Zn^{2+} ions			2

† $R_{\rm merge} = 100 \times \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement of the intensity of reflection hkl and $\langle I(hkl)\rangle$ is its mean value. ‡ R factor $= 100 \times \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $|F_{\rm obs}|$ and $|F_{\rm calc}|$ are the observed and calculated structure-factor amplitudes, respectively. § Calculated using a data set consisting of a random 10% of the observations that were not included throughout refinement (Brünger, 1992).

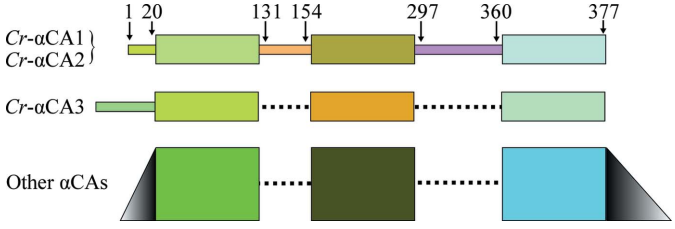

Figure 1
A schematic drawing of the sequences of *C. reinhardtii* αCAs and those of other αCAs. For a detailed alignment, see Supplementary Fig. 2. The numbering for *Cr*-αCA1 is indicated.

2.4. PDB deposition

The coordinates and structure factors have been deposited in the PDB with accession code 3b1b.

3. Results

3.1. Amino-acid sequence of *Cr*-αCA1

A multiple alignment of amino-acid sequences of αCAs retrieved from genomic databases reveals remarkable differences in the sequences of αCAs from the *Chlamydomonas* genus (Fig. 1 and Supplementary Fig. 2[2]). Only the precursor forms of *Cr*-αCA1 and *Cr*-αCA2 contain the two insertion sequences (residues 131–154 and 297–342). The second insertion sequences of precursor *Cr*-αCA1 and *Cr*-αCA2 are highly homologous to each other. It has been shown that a 35-residue peptide consisting of residues 306–340 is deleted from the precursor during maturation (Ishida *et al.*, 1993). Apart from the insertion sequences, *Cr*-αCA1 and *Cr*-αCA2 have the same sequence of 20 N-terminal residues, which differs from those of *Cr*-αCA3 and αCAs from other organisms.

3.2. Quality of X-ray analysis

Statistics of the observed intensities, crystallographic data and structure refinement are given in Table 1. X-ray diffraction patterns showed the space group to be $P6_1$ or $P6_5$ and the number of dimers in the asymmetric unit was estimated to be one (Suzuki *et al.*, 2010). Two Zn atoms were identified at a significant level, from which the best solution, with a correlation coefficient of 0.322 between F_o and F_c in the case of space group $P6_5$, was derived. An electron-density map calculated with the estimated phase angles was modified by the solvent-flattening technique, with the solvent content being fixed at 61%. The final correlation coefficient in ($|E^2|$) was 0.872. After structure model building, more than 88% of the residues of the protein were successfully assigned.

The atomic parameters of the crystal structure were refined using the programs *REFMAC*5 (Murshudov *et al.*, 2011) and *Coot*, which was also used to assign additional atoms. An electron-density map (Fig. 2) suggested the existence of glycans attached to the carbamoyl groups of Asn101, Asn135 and Asn297 in both subunits of the dimer. Since the types of

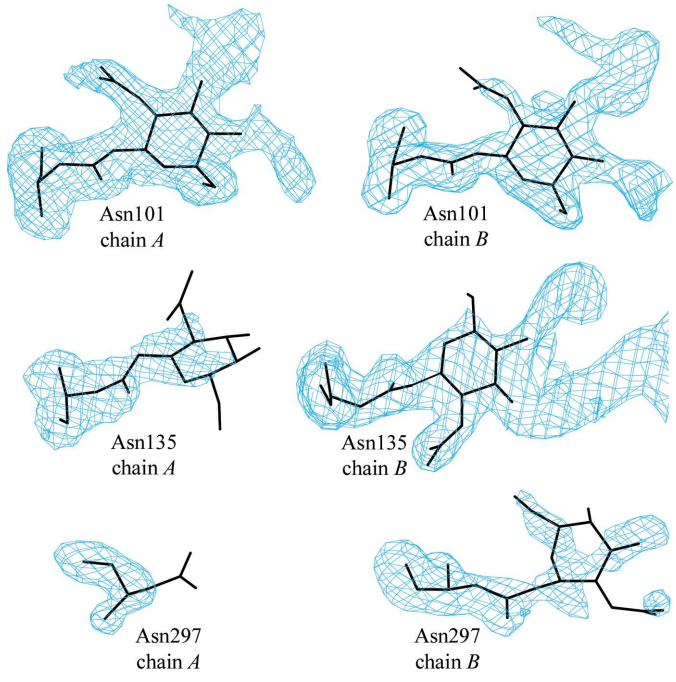

Figure 2
Local $|F_o| - |F_c|$ OMIT maps (contoured at the 3σ level) calculated without *N*-acetylglucosamine for the Asn101, Asn135 and Asn297 residues of the two subunits (chains *A* and *B*).

[2] Supplementary material has been deposited in the IUCr electronic archive (Reference: GX5191). Services for accessing this material are described at the back of the journal.

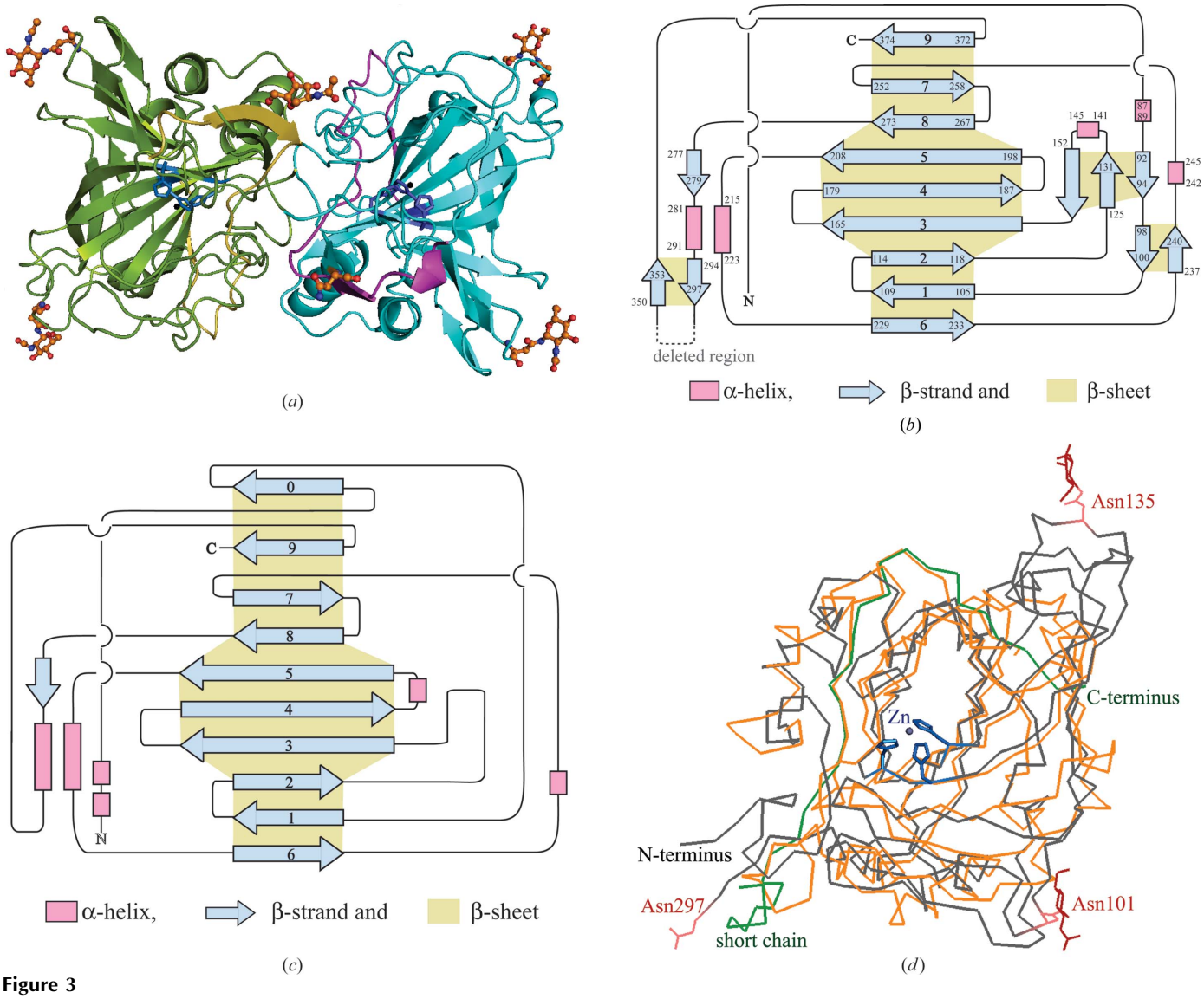

Figure 3
Overall structure of *Cr*-αCA1. (*a*) A ribbon drawing of the dimeric structure. The long and short peptides are coloured green and yellow in one subunit and cyan and magenta in the other. The topology of the secondary structures is shown for (*b*) the *Cr*-αCA1 subunit and (*c*) human αCA2. The numbering is indicated for *Cr*-αCA1. A comparison of the C^{α} structures of *Cr*-αCA1 (grey and green) and human αCA2 (orange; PDB entry 1ca2) is shown in (*d*). The C^{α} structures of a *Cr*-αCA1 subunit and human αCA2 were superimposed on each other using the program *Swiss-PdbViewer*. The r.m.s.d. was 4.03 Å for 120 equivalent C^{α} atoms.

glycans were not clear as the corresponding densities were not well resolved, they were assumed to be *N*-acetylglucosamine. Three sulfate ions and 888 water molecules were also assigned. The final R and $R_{\rm free}$ (Brünger, 1992) values were 18.9% and 21.8%, respectively. In the Ramachandran plot, 469 (89.8%), 49 (9.4%), two (0.4%) and two (0.4%) residues were in the most favoured, allowed, generally allowed and disallowed geometries, respectively.

3.3. Architectural features of *Cr*-αCA1

Cr-αCA1 is a dimeric enzyme composed of two identical subunits which are related by noncrystallographic twofold symmetry[3] as shown in Fig. 3(*a*). The two subunits are associated with each other through interactions between the loop regions. There are two regions in each subunit that were not seen in the electron-density map: residues 1–20 and 298–344. Hereafter, the peptides consisting of residues 21–297 and 345–377 in each subunit will be referred to as the long and short peptides, respectively. Each subunit contains a core β-sheet structure composed of nine β-strands flanked on either side by small α-helices, β-strands and numerous loops (Fig. 3*b*). To stabilize this tertiary structure, two S—S bonds are formed between Cys61 and Cys264 and between Cys194 and Cys198 as shown in Fig. 4. An additional S—S bond is formed between Cys296 and Cys351 to link the long and short peptides. Furthermore, the two Cys21 residues form an S—S bond to stabilize the dimer formation. A similar example of dimer formation by an S—S bond is observed between the two subunits of human αCA9 (Alterio *et al.*, 2009), but the S—S

[3] The r.m.s.d. value between the corresponding Cα atoms is 0.23 Å when the two subunits are superimposed.

bond site is different. Glycosylation sites were confirmed at Asn101, Asn135 and Asn297 in both subunits, indicating that the glycans are located on the protein surface and protrude into the solvent region.

3.4. The active site of *Cr*-αCA1

The core β-sheet is mostly folded in the left-handed screw sense, with several loops making a tunnel Fig. 3(*d*). At the centre of the tunnel, a Zn atom is bound in a tetrahedral coordination by the N^{η} atoms of three histidine residues (His163, His165 and His182) and a water molecule (Fig. 5*a*). The water molecule is further hydrogen bonded to the hydroxyl group of Thr260, suggesting that it is trapped as an activated water molecule. At the right side of the Zn atom and beside the water molecule, the hydrophobic residues Val184, Val201, Trp270, Phe199, Leu259 and Leu268 make a substrate-binding pocket in which a network of water molecules can be found (Fig. 5*a*). The short peptide can be found running on the left side of the Zn atom. The main-chain carbonyl group of the Phe360 residue in the short peptide forms a hydrogen bond to the imidazole N^{δ} atom of the His165 residue in the long peptide. This interaction appears to facilitate the interaction between the imidazole N^{η} atom of His165 and the Zn atom. In addition, the main-chain amide N atom of Arg361 in the short peptide forms a hydrogen bond to the carboxyl group of Glu169 in the long peptide and this carboxyl group in turn forms a hydrogen bond to the catalytic Thr260 residue, which supports the water molecule bound to the Zn atom.

4. Discussion

A phylogenetic tree (see Supplementary Fig. 1) generated by *ClustalX* using all known αCAs, including those that were putatively assigned from genomic sequences, shows that *Cr*-αCA1, *Cr*-αCA2 and *Cr*-αCA3 belong to the same group as αCAs from *Pectobacterium carotovorum*, *Klebsiella pneumoniae*, *Neisseria gonorrhoeae*, *Nostoc* sp. PCC7120, *Nicotiana langsdorfii* and *Dunaliella salina*, which is one of two groups branched from the original stem. In this group, *Cr*-αCA1 and *Cr*-αCA2 form a small branch and *Cr*-αCA3 is isolated in another small branch. *D. salina* αCA, which has a long C-terminal extension, is also isolated in a different small branch. Closer examination shows that the primary structures of *Cr*-αCA1 and *Cr*-αCA2 differ from those of the other αCAs in the same group in the following aspects. Firstly, there are two insertion sequences in *Cr*-αCA1 and *Cr*-αCA2. Secondly, the 20 N-terminal residues of *Cr*-αCA1 and *Cr*-αCA2 are the same (MARTGALLLVALALAGCAQA). *Cr*-αCA3 does not contain the insertion sequences and has a longer N-terminal sequence that differs from those of *Cr*-αCA1 and *Cr*-αCA2. However, the remaining regions of *Cr*-αCA3 are similar to those of *Cr*-αCA1 and *Cr*-αCA2. These features may have been retained by *C. reinhardtii* from its plant/animal ancestor (Merchant *et al.*, 2007).

Based on our comparative sequence analysis, we propose that αCAs can be divided into three subclasses, α_{I}, α_{II} and α_{III}, corresponding to CAs that are unmodified, glycosylated and both glycosylated and cleaved, respectively. *Cr*-αCA1 and *Cr*-αCA2 fall into subclass α_{III} since they undergo two post-translational modifications: glycosylation at the three asparagine residues of the subunit and cleavage into the long and short peptides (Fig. 4). Glycosylation could take place in order to protect *Cr*-αCA1 and *Cr*-αCA2 from proteolytic attack or for some other physiological function.

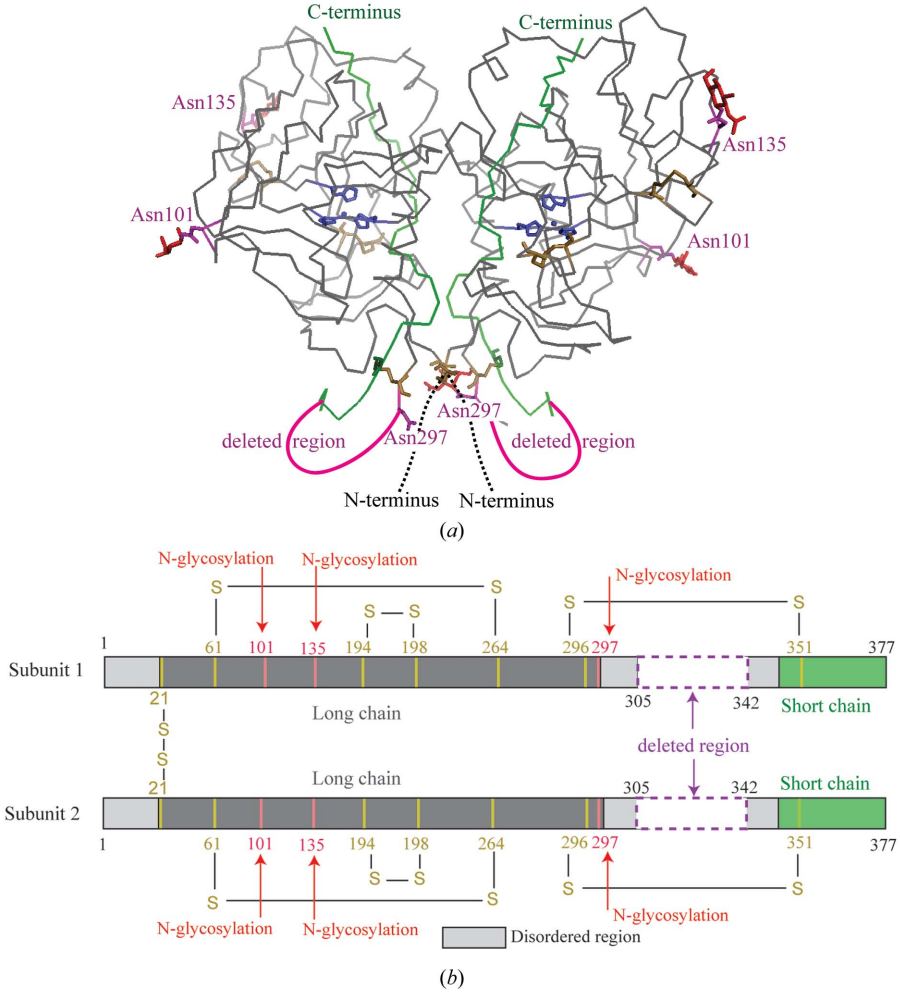

Figure 4
A summary of the structural features in (*a*) the tertiary and (*b*) the primary structure of *Cr*-αCA1.

The insertion sequence is cleaved at two sites: one between residues 305 and 306 and the other between residues 340 and 341 (Fig. 6). The sequences flanking the N-termini of the two cleavage sites are similar in *Cr*-αCA1 and *Cr*-αCA2, but those flanking the C-termini are different, suggesting that different proteases are involved in the cleavage of the two sites. *Cr*-αCA1s may be cleaved by proteases that thrive under low-CO_2 conditions, whereas *Cr*-αCA2s may be targeted by proteases that require CO_2 (Fujiwara *et al.*, 1990).

The deleted peptide in the insertion sequence of *Cr*-αCA1 and the corresponding region in the insertion sequence of *Cr*-αCA2 contain the peculiar sequence HHHHHRR-LLHNHAHLEEVPAATSEPKHYFRRVM (residues 307–339 in *Cr*-αCA1). Histidine-rich sequences usually bind to metals, haems or organic ligands. Thus, it might be possible that the histidine-rich deleted peptide of *Cr*-αCA1 binds to an as yet unknown ligand to perform a biological function.

Many X-ray structures of α_I CAs have previously been reported. Fig. 3(*c*) shows the topology of the secondary structure of human αCA2 (PDB entry 1ca2; Eriksson *et al.*, 1988; Avvaru *et al.*, 2010) as an example of an α_I CA enzyme. The core β-sheets of *Cr*-αCA1 (Fig. 3*b*) and human αCA2 have almost the same topology. One noticeable difference between the two proteins is that in human αCA2 an additional β-sheet (β0) is present parallel to the β9 strand. Moreover, the regions flanking the core β-sheets of *Cr*-αCA1 and human αCA2 are quite different in that those of *Cr*-αCA1 have more secondary structure compared with those of human αCA2 (Fig. 3*d*).

The stereochemical arrangement of the essential residues in the active site of *Cr*-αCA1 is very similar to those in αCAs of other organisms (Fig. 5*b*). Carbon dioxide could bind to the hydrophobic pocket, which is occupied by the water network near the Zn atom in the present structure. His163 and His182, which are coordinated to the Zn atom, form hydrogen bonds to Gln161 and Glu180, respectively (see Supplementary Fig. 3), as observed in other uncleaved αCAs (Eriksson *et al.*, 1988; Avvaru *et al.*, 2010). Glu169 indirectly interacts with the Zn atom through mediation of Thr260 and a water molecule, whereas His165 directly interacts with the Zn atom. The Phe360 and Arg361 residues on the short peptide form hydrogen bonds to His165 and Glu169, respectively, in the long peptide. It is possible that the C-terminal region of the precursor form of *Cr*-αCA1 corresponding to the short peptide in the mature form does not interact with the N-terminal region corresponding to the long peptide in the mature form. It has previously been shown that association of the long and short peptides is essential for catalytic activity (Ishida *et al.*, 1993). Moreover, the short peptide has a sequence homologous to that in human αCA which facilitates the interaction of CO_2 within the active site (Sheridan & Allen, 1981). We propose that the cleavage of precursor *Cr*-αCA1 into the mature form is essential for the proper positioning of the active-site residues that enables the enzyme to perform its catalytic function.

Cr-αCA1 is a dimeric protein, as shown in Fig. 3(*a*). Although dimers have been reported for αCAs from human

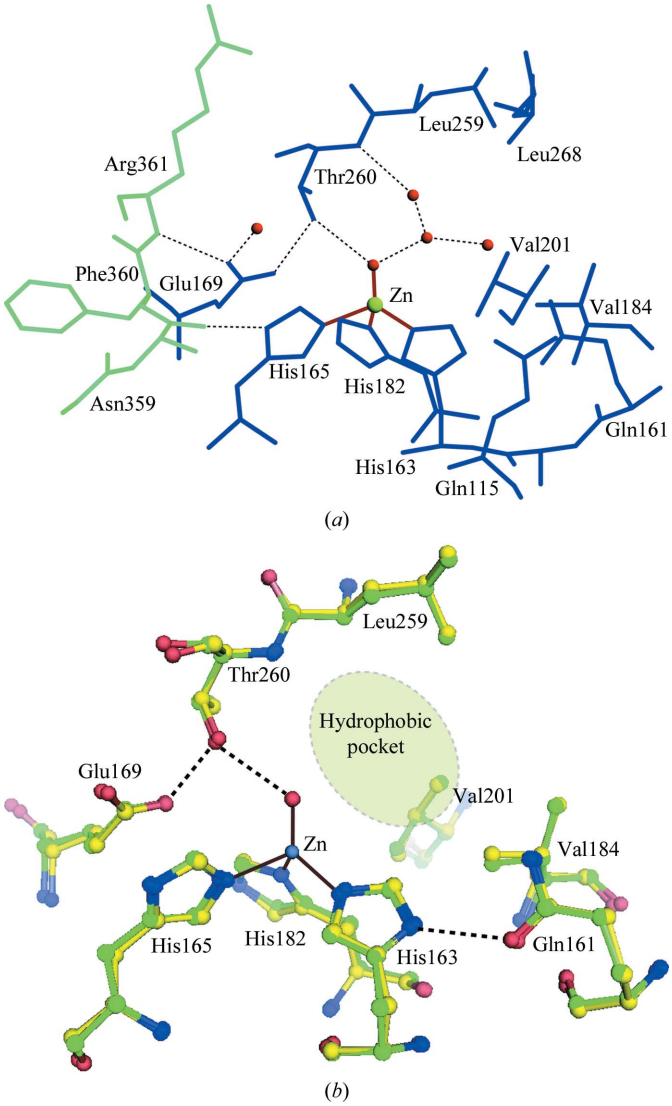

Figure 5
Structure of the active site of *Cr*-αCA1. A Zn atom bound to three histidine residues traps a water molecule, which is bound to the hydroxyl group of Thr260 (*a*). Three additional water molecules are accommodated in the hydrophobic pocket. Running on the left side is the short peptide (coloured light blue). The carbonyl O atom of Phe360 of the short peptide forms a hydrogen bond to His165. In addition, the amide N atom of Arg361, also in the short peptide, forms a hydrogen bond to Glu169, which is hydrogen bonded to Thr260. The superimposition of *Cr*-αCA1 (green) and human αCA2 (yellow) in (*b*) shows that the amino-acid residues in the active site are well conserved.

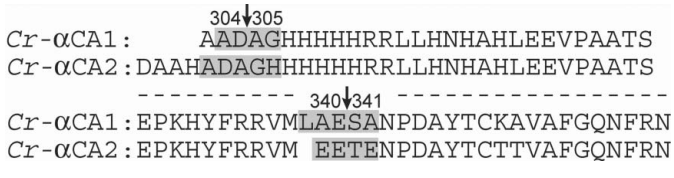

Figure 6
Amino-acid sequences of the second insertion sequence in the precursor forms of *Cr*-αCA1 and *Cr*-αCA2. Arrows indicate the cleavage sites of *Cr*-αCA1. The cleavage sites of *Cr*-αCA2 are currently unknown.

[αCA9 (PDB entry 3iai; Hilvo *et al.*, 2008; Alterio *et al.*, 2009) and αCA12 (PDB entries 1jcz and 1jd0; Whittington *et al.*, 2001)] and *Aspergillus oryzae* (αCA; PDB entry 3q31; Cuesta-Seijo *et al.*, 2011), the dimerization interfaces between the two subunits differ from that of *Cr*-αCA1. *Cr*-αCA1 is unique in that its two subunits are linked together by a disulfide bond, differing from other dimers.

Mature *Cr*-αCA1 shows a molecular weight of 79.6 kDa (almost twice the calculated molecular weight of a subunit) as estimated by gel filtration. The *Cr*-αCA1 structure shows that one disulfide bond links the long and short peptides and another links two long peptides. In the final purification, the elution pattern showed two peaks, suggesting the possibility that *Cr*-αCA1 exists as a single dimeric enzyme as well as its dimer (a tetramer) (see Supplementary Fig. 4). In Supplementary Fig. 4(*b*), the native PAGE of sample *A* shows a band (marked *a*) corresponding to the molecular size of a dimer of dimers (tetramer). However, there is no band for the single dimeric enzyme. When sample *A* was stored for three months (referred to as sample *B*), band *a* disappeared and a new band (marked *c*) appeared at the position corresponding to a trimer of dimers (hexamer). Such oligomerization properties have been observed for βCAs (Tripp *et al.*, 2001), but the structural architecture of βCAs is quite different from that of *Cr*-αCA1. In the crystal, the *Cr*-αCA1 dimers are in contact with each other according to the crystallographic 6_5 symmetry and are stacked along the *c* axis. There are two molecular interaction modes in the lateral and vertical directions in the unit cell. Under certain physiological conditions *Cr*-αCA1 dimers form higher oligomeric states, perhaps by these interaction modes.

In the present study, we have described the structural features of *Cr*-αCA1 and suggested a possible function for the short peptide in the mature form. Its role would be clarified by examining the activity and performing X-ray analysis of the precursor *Cr*-αCA1 protein. The crystal structure of the precursor will also confirm whether the N-terminal 20-amino-acid signal peptide adopts an autonomous structure or interacts with the insertion sequences.

This work was supported in part by Grants-in-Aid for the Protein3000 Research Program from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We thank S. Kuramitsu for organizing the research group in the program and N. Igarashi and S. Wakatsuki for facilities and help during data collection.

References

Alber, B. E. & Ferry, J. G. (1996). *J. Bacteriol.* **178**, 3270–3274.

Alterio, V., Hilvo, M., Fiore, A. D., Supuran, C. T., Pan, P., Parkkila, S., Scaloni, A., Pastorek, J., Pastorekova, S., Pedone, C., Scozzafava, A., Monti, S. M. & Simone, G. D. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 16233–16238.

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.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Chegwidden, W. R. (2000). *The Carbonic Anhydrases: New Horizons.* Basel: Birkhäuser.

Coleman, J. R. & Grossman, A. R. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6049–6053.

Covarrubias, A. S., Bergfors, T., Jones, T. A. & Högbom, M. (2006). *J. Biol. Chem.* **281**, 4993–4999.

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.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Eriksson, A. E., Jones, T. A. & Liljas, A. (1988). *Proteins*, **4**, 274–282.

Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.

Fujiwara, S., Fukuzawa, H., Tachiki, A. & Miyachi, S. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 9779–9783.

Guilloton, M. B., Korte, J. J., Lamblin, A. F., Fuchs, J. A. & Anderson, P. M. (1992). *J. Biol. Chem.* **267**, 3731–3734.

Hewett-Emmett, D. & Tashian, R. E. (1996). *Mol. Phylogenet. Evol.* **5**, 50–77.

Hiltonen, T., Björkbacka, H., Forsman, C., Clarke, A. K. & Samuelsson, G. (1998). *Plant Physiol.* **117**, 1341–1349.

Hilvo, M. *et al.* (2008). *J. Biol. Chem.* **283**, 27799–27809.

Hunnik, E. van, Livne, A., Pogenberg, V., Spijkerman, E., van den Ende, H., Mendoza, E. G., Sültemeyer, D. & de Leeuw, J. W. (2001). *Planta*, **212**, 454–459.

Ishida, S., Muto, S. & Miyachi, S. (1993). *Eur. J. Biochem.* **214**, 9–16.

Kandel, M., Gornall, A. G., Cybulsky, D. L. & Kandel, S. I. (1978). *J. Biol. Chem.* **253**, 679–685.

Karlsson, J., Clarke, A. K., Chen, Z.-Y., Hugghins, S. Y., Park, Y.-I., Husic, H. D., Moroney, J. V. & Samuelsson, G. (1998). *EMBO J.* **17**, 1208–1216.

Kimber, M. S. & Pai, E. F. (2000). *EMBO J.* **19**, 1407–1418.

Kimpel, D. L., Togasaki, R. K. & Miyachi, S. (1983). *Plant Cell Physiol.* **24**, 255–259.

Kisiel, W. & Graf, G. (1972). *Phytochemistry*, **11**, 113–117.

Kondo, T., Murakami, K., Ohtsuka, Y., Tsuji, M., Gasa, S., Taniguchi, N. & Kawakami, Y. (1987). *Clin. Chim. Acta*, **166**, 227–236.

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

Merchant, S. S. *et al.* (2007). *Science*, **318**, 245–250.

Mitra, M., Lato, S. M., Ynalvez, R. A., Xiao, Y. & Moroney, J. V. (2004). *Plant Physiol.* **135**, 173–182.

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

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Parker, M. S., Mock, T. & Armbrust, E. V. (2008). *Annu. Rev. Genet.* **42**, 619–645.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rumeau, D., Cuiné, S., Fina, L., Gault, N., Nicole, M. & Peltier, G. (1996). *Planta*, **199**, 79–88.

Sheridan, R. P. & Allen, L. C. (1981). *J. Am. Chem. Soc.* **103**, 1544–1550.

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

Strop, P., Smith, K. S., Iverson, T. M., Ferry, J. G. & Rees, D. C. (2001). *J. Biol. Chem.* **276**, 10299–10305.

Suzuki, K., Shimizu, S., Juan, E. C. M., Miyamoto, T., Fang, Z., Hoque, Md. M., Sato, Y., Tsunoda, M., Sekiguchi, T., Takénaka, A. & Yang, S.-Y. (2010). *Acta Cryst.* F**66**, 1082–1085.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). *Nucleic Acids Res.* **24**, 4876–4882.

Tobin, A. J. (1970). *J. Biol. Chem.* **245**, 2656–2666.

Tripp, B. C., Smith, K. & Ferry, J. G. (2001). *J. Biol. Chem.* **276**, 48615–48618.

Villarejo, A., Shutova, T., Moskvin, O., Forssén, M., Klimov, V. V. & Samuelsson, G. (2002). *EMBO J.* **21**, 1930–1938.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Waheed, A., Zhu, X. L., Sly, W. S., Wetzel, P. & Gros, G. (1992). *Arch. Biochem. Biophys.* **294**, 550–556.
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.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yagawa, Y., Aizawa, K., Yang, S.-Y. & Miyachi, S. (1986). *Plant Cell Physiol.* **27**, 215–221.
Yang, S.-Y., Tsuzuki, M. & Miyachi, S. (1985). *Plant Cell Physiol.* **26**, 25–34.