Carbonic anhydrases catalyze the reversible hydration of carbon dioxide to form bicarbonate. This activity is universally required for fatty acid biosynthesis as well as for the production of a number of small molecules, pH homeostasis, and other functions. At least three different carbonic anhydrase families are known to exist, of which the \( \alpha \)-class found in humans has been studied in most detail. In the present work, we describe the structures of two of the three \( \beta \)-class carbonic anhydrases that have been identified in *Mycobacterium tuberculosis*, i.e. Rv1284 and Rv3588c. Both structures were solved by molecular replacement and then refined to resolutions of 2.0 and 1.75 Å, respectively. The active site of Rv1284 is small and almost completely shielded from solvent, whereas that of Rv3588c is larger and quite open to solution. Differences in coordination of the active site metal are also observed. In Rv3588c, an aspartic acid side chain displaces a water molecule and coordinates directly to the zinc ion, thereby closing the zinc coordination sphere and breaking the salt link to a nearby arginine that is a feature of Rv1284. The two carbonic anhydrases thus exhibit both of the metal coordination geometries that have previously been observed for structures in this family. Activity studies demonstrate that Rv3588c is a completely functional carbonic anhydrase. The apparent lack of activity of Rv1284 in the present assay system is likely exacerbated by the observed depletion of zinc in the preparation.

The World Health Organization estimates that 1.8 billion people are infected with *Mycobacterium tuberculosis*, with 8 million new cases and 2 million deaths per year. The so-called “short-term” antibiotic treatment currently available lasts 6 months and does not seem to offer much hope for eradicating the disease in the long run. In the milestone publication of the World Health Organization in 1995 (1) were able to assign functions to some 60% of the gene products. Subsequent studies have attempted to identify genes that are likely to be required for bacterial survival and infection (e.g. studies by Sassetti and Rubin (2) and Sassetti et al. (3)). We have used such investigations as the basis for identifying *M. tuberculosis* gene products that are interesting subjects for detailed structural and functional studies.

The gene Rv1284 is annotated at the Pasteur Institute TubercuList server (genolist.pasteur.fr/TubercuList) as coding for a conserved hypothetical protein of unknown function, that is, it is found in related species and presumably has an essential role in their survival, although it is not clear what function is involved. Furthermore, Rv1284 was judged to be essential by Himar1-based transposon mutagenesis in strain H37Rv (3), and its transcription was highly up-regulated under the starvation conditions used to model persistent bacteria (4). Phylogenetic studies of putative carbonic anhydrases (CAs) (EC 4.2.1.1) suggested that Rv1284 in fact encodes a \( \beta \)-CA and uncovered distant relationships to two other *M. tuberculosis* proteins, those produced by the Rv3588c gene and the C-terminal region of Rv3273 (5). Rv3588c has been identified as being required for mycobacterial growth *in vivo* by Sassetti and Rubin (2), whereas Rv3273 is not essential (2, 3). Human carbonic anhydrases belong to the \( \alpha \)-class, which has a completely different structure; both \( \alpha \)- and \( \beta \)-classes, however, are comprised of zinc-based enzymes (see Ref. 6 for a review of carbonic anhydrase structures). This difference as well as the success of drug design for the \( \alpha \)-CAs (reviewed in Ref. 7) prompted us to carry out a detailed analysis of the mycobacterial enzymes. In the present publication we report structural and functional studies of the Rv1284 and Rv3588c gene products. The results demonstrate that Rv3588c is indeed a \( \beta \)-CA and that Rv1284 is a \( \beta \)-CA-like protein.

**EXPERIMENTAL PROCEDURES**

**Cloning, Protein Expression, and Purification**—The sequences correponding to the open reading frames of Rv1284 and Rv3588c were amplified by PCR from *M. tuberculosis* DNA strain H37Rv (1) with Pfu polymerase using the primer pair 5'-GTGACGTTACCCGACCTACTCGT-3' (Rv1284.forward) and 5'-TATGGTATGCTCATCTGCAAACTCAGATTGTA-3' (Rv1284.reverse) for Rv1284 and 5'-GTGACGTTACCCGACCTACTCGT-3' (Rv3588c.forward) and 5'-TCCCGGTATCCTATTTAGACCTCTGCGATGT-3' (Rv3588c.reverse) for Rv3588c. A second PCR to introduce an N-terminal His tag into each construct was performed with the same reverse primer and a new forward primer, Rv1284.his (5'-ATGGCCCATCATCATCATCATCATCTCTGTTGCTACGCTACGCAGACTACTCTG-3') and Rv3588c.his (5'-ATGGCCCATCATCATCATCATCATCTCTGTTGCTACGCTACGCAGACTACTCTG-3').

1 The abbreviations used are: CA, carbonic anhydrase; TAPS, 3-[tris(hydroxymethyl)methylamino]-2-hydroxypropanesulfonic acid; SSPE, saline/sodium phosphate/EDTA; r.m.s.d., root mean square deviation.
Carbonic Anhydrases from Mycobacterium tuberculosis

18783

ATCATCATGATGTCUCCACCAAGGATCGGGAGCCG-3, respectively, using the products from the first PCRs as templates. The PCR products were incubated with deoxynucleotide triphosphates and Taq polymerase to create an A-overhang, purified, and ligated into vector pCR®. The resulting plasmids were transformed into Top10 cells (Invitrogen) by heat shock and grown on Luria-Bertani agar plates containing 50 μg/ml ampicillin. Positive clones were transformed into E. coli BL21 (DE3) in which protein was overexpressed after induction with 0.02% arabinose at 37 °C for 2 h. Cells were harvested by centrifugation. Cell pellets were removed from the pellet with 1× SSEPE buffer (150 mM NaCl, 10 mM NaH2PO4, pH 7.5, and 1 mM EDTA) and stored at −70 °C. All clones were verified by DNA sequencing.

For purification, thawed cells were re-suspended in native lysis buffer (50 mM NaH2PO4, pH 8.3, 300 mM NaCl, 20 mM imidazole, and 10% glycerol) with 0.01 mg/ml RNase, 0.02 mg/ml DNase, and 1 mM phenylmethylsulfonfluoride and lysed using a One Shot cell disruptor (Constant Systems, Ltd). Cell-free extracts were incubated with pre-equilibrated nickel-agarose (Analytica AB) (Qiagen) slurry for 45 min at 4 °C; resin was then washed, and protein was eluted with 250 mM imidazole. The eluted protein fractions were applied to size-exclusion chromatography columns (HiLoad™ 16/60 Superdex™ 75; Amersham Biosciences) using a buffer containing 150 mM NaCl, 20 mM Tris-HCl, pH 7.5, and 5% glycerol. For Rv1284, the buffer also included 5 mM β-mercaptoethanol. The purified protein solutions were evaluated by SDS-PAGE (PhastSystem™; Amersham Biosciences). Protein samples were concentrated using Vivaspin concentrators (Vivascreen) with a molecular mass cutoff of 10 kDa. Final storage buffer was 10 mM Tris-HCl, pH 7.5, and 5 mM β-mercaptoethanol for Rv1284 and 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 5% glycerol for Rv3588c.

Assays and Metal Analysis—Esterase activity was measured as p-nitrophenylacetate hydrolysis at 25 °C, using a modification of previously published methods (8). A 5 mM stock of p-nitrophenylacetate was freshly prepared by dissolving the compound in acetone and then diluting it 25-fold with water. The uncatalyzed reaction was measured by adding 1.9 ml of 50 mM Tris-HCl, pH 7.5, to 1 ml of this substrate solution and recording the change in A254. After 3 min, 100 μM of enzyme solution was added, and the catalyzed reaction was monitored for an additional 3 min. A non-enzymatic control (using bovine serum albumin) was included in each set of measurements; bovine carbonic anhydrase II (Sigma) was used as a positive control.

The Rv3588c solution containing enzyme (final protein concentration, 2.5 mg/ml) mixed with 2 μl of reservoir solution containing 0.1 μM sodium thiosulfate and 0.2 μM EDTA was used to prepare the sample for crystallization. The preparation was added to the mother liquor and then dialyzed against a similar buffer/indicator mixture (A) (pH 8.4–9.0, containing 136 mM sodium sulfate to maintain a relatively constant ionic strength). The reaction was initiated by addition of 0.1-s intervals using a Beckmann DU-600 spectrophotometer with time t = 0 coinciding with manual addition of substrate. The reaction medium was prepared by adding 1 μl of 50 mM Tris-HCl, pH 7.5, to 1 ml of this substrate solution and recording the change in A254. After 3 min, 100 μM of enzyme solution was added, and the catalyzed reaction was monitored for an additional 3 min. A non-enzymatic control (using bovine serum albumin) was included in each set of measurements; bovine carbonic anhydrase II (Sigma) was used as a positive control.

Carbonic anhydrase activity was assayed by measuring changes in pH during the reaction with a dye indicator method (9). Assays were again performed at 25 °C. The buffer/indicator pairs used were (a) 25 mM HEPES plus 100 μM phenol red (pH 7.5, A550) and (b) 25 mM TAPS plus 100 μM m-cresol purple (pH 8.4–9.0, A580); both solutions also contained 100 mM sodium sulfate to maintain a relatively constant ionic strength. The reaction was initiated by addition of 0.5 ml of CO2-saturated water to 0.5 ml of a 2× buffer/indicator solution containing enzyme (final protein concentration, 2.5 μM). Non-enzymatic and positive controls were included as described for the esterase assay. Data were acquired at the appropriate wavelength at 0.1-s intervals using a Beckmann DU-600 spectrophotometer with time t = 0 coinciding with manual addition of substrate.

Crystallization, Data Collection, and Processing—Rv1284 crystals were grown by the hanging-drop vapor-diffusion method (10) at 20 °C. The crystallization drop setup consisted of 2 μl of protein sample (14.5 mg/ml) mixed with 2 μl of reservoir solution containing 0.1 μM sodium thiosulfate and 0.2 μM EDTA and then dialyzed against a similar buffer/indicator mixture containing 100 mM sodium sulfate and 50% glycerol. Crystals appeared after 1 day. Prior to data collection, the crystal was transferred to a cryoprotectant solution comprising reservoir solution with 30% glycerol and then immediately flash-frozen in liquid nitrogen. X-ray data were collected at 100 K on beamline ID14-1 of the European Synchrotron Radiation Facility in Grenoble, France. Diffraction data were processed using MOSFLM (11) and reduced and scaled in SCALa (12) using the CCP4 program suite (13).

Rv3588c was crystallized using the sitting-drop vapor diffusion method. Final conditions for crystallization were as follows: 1 μl of protein solution (34 mg/ml) was mixed with 1 μl of reservoir solution containing 200 mM MgCl2, 35% polyethylene glycol 400, and 100 mM HEPES at pH 7.0. Crystals grew to 0.1–0.2 mm in diameter in 5–8 days and could be flash-frozen directly from the mother liquor. Diffraction data were collected at 100 K at beamline I111 at the MAXII synchrotron in Lund, Sweden. The data were processed and scaled using Denzo and Scalepack (14).

Data collection statistics for both crystal forms are shown in Table I. Structure Determination and Refinement—The Rv1284 structure was solved by molecular replacement in AMoRe (15) using a dimer of the carbonic anhydrase from Methanobacterium thermoautotrophicum (Protein Data Bank code 1G5C) (16) as a template for the search model. All residues with sequence identity were kept unchanged, whereas other non-alanine and non-glycine residues were mutated to serine using SEAMAN (17). The first 25 residues of both molecules in the pellicle were removed from the model. Because a strong translation peak appeared in the native Patterson map, the same top rotation solution was used to locate a second dimer. Rigid body refinement was then performed in REFMACS (18) with each molecule defined as a rigid unit. NCS phased refinement in CNSref (within the CCP4 interface) (19) was used to generate an averaged map. Additional features not present in the search model and consistent with the M. tuberculosis sequence appeared in the averaged electron density. By alternating NCS phased refinement, model rebuilding in O (20), and data extension to 2 Å, the complete M. tuberculosis sequence could eventually be placed in density. NCS restraints were then released, and the refinement proceeded with alternating cycles of REFMACS and manual rebuilding of the four separate subunits. Each dimer was used to define a separate TLS group (21) for the final cycles of refinement.

Rv3588c was solved by molecular replacement using CNS (22). The search model was constructed from the two structures deposited in the Protein Data Bank having the highest sequence identity: Protein Data Bank code 1H6O (23) with 33.7% identity in a 166-amino acid segment and Protein Data Bank code 1DDZ (24) with 34.6% identity in a 182-amino acid segment. The final search model consisted of four polypeptide chains containing 136 residues in both enantiomorphic space groups, P4212 and P4212, which were then compared to the 2-Å resolution map. The best solution was found in P4212. Packing analysis showed that the top solution produced a correct dimer with a symmetry-related molecule and plausible crystal packing. Rigid body refinement to 3Å resolution was followed by alternating cycles of REFMACS refinement and manual rebuilding. During refinement, density incoherence became apparent in the active site of residues 1–3, 32–41, and 207.

The Rv1284 structure appeared in the averaged electron density. By alternating NCS phased refinement, model rebuilding in O (20), and data extension to 2 Å, the complete M. tuberculosis sequence could eventually be placed in density. NCS restraints were then released, and the refinement proceeded with alternating cycles of REFMACS and manual rebuilding of the four separate subunits. Each dimer was used to define a separate TLS group (21) for the final cycles of refinement.

RESULTS

Overall Structures of Rv1284 and Rv3588c—His-tagged full-length Rv1284 and Rv3588c were expressed, purified, and crystallized, and their structures were determined by molecular replacement. Both structures have been refined to 2 Å resolution or better, with the statistics reported in Table I. The Rv1284 crystal contains two dimers in the asymmetric unit, each of which is part of a tetramer formed via a crystallographic 2-fold axis. Rv3588c, on the other hand, contains just a single subunit that forms a dimer by the interaction around a crystallographic 2-fold axis. The Rv1284 structure shows continuous electron density for the entire sequence, i.e. from residue 1 to residue 348. Rv3588c, however, has no density for residues 1–3, 32–41, and 207.

Both structures have the β-CA all β-f anticlinal from sequence analysis. Each subunit consists of a five stranded β-sheet, in which strands 1–4 are parallel and ordered 2-1-3-4. The fifth strand is anti-parallel and connected to the fourth by a short reverse turn. The connecting loops are of different length and structure in the two proteins. The β1-β2 loop is short and irregular in both cases. The β2-β3 connection includes a regular α-helix (α2) packing on the surface of the sheet. The β3-β4 loop is much longer, containing two helices in Rv1284 and four helices in Rv3588c. One of these helices (α4 in Rv1284) packs on the sheet near β4 in both proteins. The
remaining portion of the loop stretches across the surface of the dimer to interact with the 2-fold related helix linking β2 and β3. In Rv1284, the first α-helix clearly packs on the sheet of the other subunit, as was observed in three of the four known β₇CA structures (16, 23, 24, 26) (see Fig. 1A). Because of the density break in Rv3588c we cannot rule out the possibility that its first helix packs in the same manner, but the positioning of the abutting residues suggests that it is not domain-swapped (Fig. 1B). The subunits of Rv1284 and Rv3588c can be superimposed such that 106 Cα pairs have an r.m.s.d. of 1.45 Å (27), consistent with their structure-based sequence identity of ~15%. The largest structural variation is observed in the β3-β4 loop, as described above.

Active Sites—The active site of a β₇CA lies near a switch point at the C-terminal edge of its parallel β-sheet, as has been observed for other α/β proteins (28). Although neither of the present crystallization reagents included metal ions, both structures show strong electron density at the position where one would normally expect a zinc ion in an active β₇CA enzyme. In each case we have modeled this density as a zinc atom, although the peak height in Rv3588c is twice that in Rv1284, although it is not coordinated to the metal ion. The Rv1284 active site pocket is small (accessible surface volume, ~7 Å³) and almost completely shielded from solvent; its shape is a perfect match for that of the observed thiocyanate. It contains only one other ionizable side chain, His54, from the second subunit.

In Rv3588c, the active site residue Asp53 displaces the water molecule and coordinates directly to the zinc ion, thus breaking a potential salt link to Arg55 (Fig. 2, B and C). The two M. tuberculosis β-CAs, therefore, exhibit both of the coordination variants reported for β-CAs elsewhere (16, 23, 24, 26). There is an additional hydrogen-bonding group in the active site cavity of Rv3588c, Tyr88⁹, which is positioned 6.7 Å from the Asp53 ligand; this residue is structurally conserved in three of the four other structures. The Rv3588c active site is much more open than that of Rv1284 and directly accessible to the solvent. This difference may be primarily due to the lack of an ordered loop between the N-terminal helix and β1.

In Rv3588c the greatest deviations from tetrahedral geometry at the zinc atom involve the aspartyl ligand, with one angle deviating from the ideal by 18°. The geometry of the protein ligands in Rv1284 appears more strained; the angle subtended at the metal by the Sγ atoms again deviates by ~18° from the ideal, and the ion is also ~0.7 Å out of the plane of the imidazole ring.

Comparison to Other β-CAs—Four β₇CA structures have previously been reported in the literature. They can be aligned to the Rv1284 dimer with 220–240 matching Cα pairs having r.m.s.d. values of 1.4–1.6 Å using our standard alignment criteria (27) (see Table II). The sequence identity of the aligned residues is very low (~15%), except in the case of the M. thermoautotrophicum enzyme, to which it has 29% identity. By
contrast, similar alignments to Rv3588c indicate that it is much more similar to the other three CA structures (Table III).

The four earlier \( \beta \)-CA structures are superimposed on Rv1284 in Fig. 3. The comparison shows that the \( \beta \)-sheet and two of the \( \alpha \)-helices that pack on it (in the \( \beta_2/\beta_3 \) and \( \beta_3/\beta_4 \) loops) are well conserved, although there is a great deal of variation in other regions. One major difference involves the N-terminal \( \alpha \)-helix. In Rv1284, as well as the red alga (Protein Data Bank code 1DDZ) (24), plant (Protein Data Bank code 1EKJ) (26), \( E. \) coli (Protein Data Bank code 1I6P) (23) enzymes, this helix packs on the sheet of the other subunit, although the superimposed coordinates do not form a tight cluster. In the enzyme from \( M. \) thermoautotrophicum (Protein Data Bank code 1G5C) (16), the equivalent helix packs on the sheet of other dimers in the crystal. All of the structures show good agreement within \( \beta_1 \) as far as the conserved aspartic acid (Asp37 in Rv1284 numbering), \( \beta_2 \), the helical portion of the

| CA source               | Nr* | Nc | Ir (%) | r.m.s. distance | Z    | PDB code |
|-------------------------|-----|----|--------|----------------|------|----------|
| \( M. \) thermoautotrophicum | 316 | 239| 29.1    | 1.382          | 18.3 | 1G5C     |
| \( P. \) sativum         | 423 | 235| 16.0    | 1.559          | 14.0 | 1EKJ     |
| \( E. \) coli            | 428 | 233| 14.7    | 1.375          | 13.9 | 1I6P     |
| \( M. \) tuberculosis, Rv3588c | 386 | 238| 13.5    | 1.812          | 12.5 | 1YM3     |
| \( P. \) purpureum       | 481 | 222| 14.7    | 1.584          | 11.8 | 1DDZ     |

\* Nr, number of residues in the dimer; Nc, numbers of atoms within a 3.5-A cutoff; Ir, sequence identity in conserved region/number of residues in Rv1284; r.m.s. distance, root mean square distance to Rv1284; Z, Dali Z-score; PDB, Protein Data Bank.

\( \beta_2/\beta_3 \) connection, and \( \beta_3 \). As described for the two \( M. \) tuberculosis proteins, large variations are seen in the loop region between \( \beta_3 \) and \( \beta_4 \). The C-terminal end of this long loop is, however, tightly clustered on the helix that stacks on strand \( \beta_4 \) (\( \alpha_4 \) in Rv1284). The reverse turn between strands \( \beta_4 \) and \( \beta_5 \) is

| CA source                      | Nr* | Nc | Ir (%) | r.m.s. distance | Z    | PDB code |
|--------------------------------|-----|----|--------|----------------|------|----------|
| \( E. \) coli                  | 428 | 276| 20.7    | 1.474          | 18.8 | 1I6P     |
| \( P. \) sativum                | 423 | 280| 18.5    | 1.725          | 18.6 | 1EKJ     |
| \( P. \) purpureum              | 481 | 305| 22.2    | 1.480          | 17.5 | 1DDZ     |
| \( M. \) thermoautotrophicum   | 316 | 208| 14.0    | 1.609          | 12.7 | 1G5C     |

\* Nr, number of residues in the dimer; Nc, numbers of atoms within a 3.5-A cutoff; Ir, identical residues in conserved region/number of residues in Rv3588c; r.m.s. distance, root mean square distance to Rv3588c; Z, Dali Z-score; PDB, Protein Data Bank.
Carbonic Anhydrases from Mycobacterium tuberculosis

Fig. 3. Structural comparisons. A Co-based superposition of Rv1284 and Rv3588c dimers with other β-CAs is shown. Rv1284 has rainbow coloring going from red to blue, whereas the subunits of the molecular dimer of other structures are colored cyan and gold. One active site of Rv1284 is included as a reference point, in which the modeled zinc is illustrated as a magenta sphere.

---

Conserved in length but shows some spatial fanning. Some pairs of structures show similarities that are not present in the entire family. For example, the β2-β3 loops in the structures of Protein Data Bank codes 1I6P and 1DDZ are fully conserved, whereas in the other family members, only the helix-turn-β3 unit is similar.

In two of the previously published structures (those from P. sativum and M. thermoautotrophicum), three protein side chains and a water molecule coordinate the active site zinc atom; these exhibit geometries that are very similar to that observed in Rv1284. The pea enzyme, in particular, shows similar deviations from tetrahedral geometry at the zinc atom. The other two structures (from E. coli and red algae) exhibit the aspartyl residue coordination observed for Rv3588c. In other respects, the coordination is similar, although other residues lining the active site pocket are not conserved. The most constrained entrance to the active site is found in Rv1284, primarily due to side chains from the loop connecting the first helix and strand, particularly Met24 and Pro25. The thiocyanate ion observed in Rv1284 overlaps with the acetate ion found in the active site of the P. sativum enzyme.

Only thirteen residues are completely conserved in alignments based on the common structural core (Fig. 4). Five of these are the observed metal ligands and the neighboring arginine residue. Outside this core, the sequence alignments must be treated with caution, largely because of the structural variations already described.

The first conserved region in the sequence contains two metal ligands, as well as the nearby arginine. Here, the side chains of Ala38/Ser54 (Rv1284/Rv3588c numbering, respectively) point in opposite directions (with Cα separated by ~5 Å), as a result of the rearrangements that are associated with changes in the salt bridge and metal coordination. The movement of Leu49/Val56 is also part of the structural adjustments that allow the guanidinium group of Arg73/Arg75 to stay in approximately the same place.

The next region of sequence similarity corresponds to the edge strand (β2, i.e., residues Gly51-Gly60 in Rv1284) that is part of the dimer interface. The conserved glycine that begins this strand adopts an αi conformation, whereas the highly conserved asparagine side chain near the end (Asn55) forms an Asn turn hydrogen bond. The arginine residue in the middle of the strand, which may be conservatively replaced by a lysine, is buried within the structure, without compensating interactions to negatively charged side chains. Instead, the guanidinium group of Arg77 donates four hydrogen bonds to main chain carbonyl groups in Rv1284; the equivalent Arg76 is solvated in Rv3588c.

The next area of conservation lies in β3, the central strand of each subunit. Branched side chains in the middle of this strand interact with α-helices above and below the plane of the sheet (α2 and α3 on one side, and the β1/β2 loop and α1’ on the other). His58 and Cys51 at the C-terminal end of the strand are active site ligands. The conserved glycine immediately following, residue 92, adopts an α-helical conformation; any side chain group at this position would impinge on the active site cavity. Helix α4 packs on the surface of the sheet, in particular on β4, to give a conserved pattern of branched hydrophobic side chains (Val127, Leu131, and Ile134 from the helix, and Leu146 and Gly148 from β4). The loop connecting the C terminus of α4 and β4 is variable in length and structure. The final highly conserved residue is Gly156, which lies in a variation of the classical three-residue reverse turn (29) between β4 and β5.

Metal Content—Rv1284 and Rv3588c protein samples were analyzed for a number of metals: Zn, Ni, Fe, Cu, Mn, Al, Cr, Co, Cd, Pb, As, Ba, and Hg. Rv3588c was found to contain 0.94 equivalent of Zn per protein molecule; in addition it contained 0.04 equivalent of each of Ni and Cu and 0.01 equivalent of Al. The Rv1284 sample contained only 0.30 equivalent of Zn, together with 0.18 equivalent of Ni, 0.05 equivalent of Al, and 0.03 equivalent of Fe. Only trace amounts (<0.01 equivalent) were found for the other metals assayed.

Activity Assays—The ability of Rv1284 and Rv3588c to catalyze the hydration of CO2 was tested using a colorimetric assay (9) at several different starting pH values. These measurements indicate that Rv3588c is an active carbonic anhydrase at pH 8.4 (Fig. 5). However, no activity was detectable at pH 7.5 under the conditions used in our assay (data not shown). The activity of Rv1284 was not detectable at either pH value; omitting sulfate or chloride from the solutions did not affect this result.

In common with the other β-CA enzymes that have been tested (e.g. Ref. 30), neither Rv1284 nor Rv3588c showed the esterase activity that is characteristic of the α-CA family.

DISCUSSION

Carbonic anhydrases catalyze the reversible hydration of carbon dioxide to form bicarbonate. This simple conversion of a membrane-permeable gas substrate into a membrane-impermeable ionic product is vital to many important biological functions; such enzymes are thus widely distributed in nature. At least three different CA families are known to exist, α-, β-, and γ-CAs, of which the α-class has been studied in the most detail (31). Despite the fact that the three families are unrelated in sequence and structure, all are comprised of zinc-containing enzymes. All CAs in the animal kingdom are of the α-class, and the human enzyme is a validated drug target. The topical CA inhibitor dorzolamide, for example, binds to the active site zinc and is used for the treatment of glaucoma (32, 33). The α-class CAs have also been found in algae and prokaryotes. β-Class enzymes have not been found in the animal kingdom but have been observed in a wide variety of other organisms.

In an α CA enzyme, a zinc-bound OH− ion attacks a CO2 molecule to form a metal-bound HCO3− ion that is subsequently displaced by a new water molecule (31). Proton transfer to the
external buffer then regenerates the zinc-bound OH− ion. A detailed molecular scheme has been outlined that includes the initial binding of carbon dioxide in a pocket close to the zinc-bound hydroxide (where Thr199 in the human isozyme CA II plays an important role as the “door-keeper”) and a proton transfer route from the metal center to His64 via a network of hydrogen-bonded water molecules (31).

The first structural information on β-CAs revealed an entirely different protein architecture and supplied the atomic details of its distinct type of active site. Independent structure determinations of β-CAs from pea (P. sativum) (26) and red algae (Porphyridium purpureum) (24) revealed similar overall folds but different coordination of the metal at the active site. In the pea enzyme, the protein supplied three ligands, whereas in the red algae enzyme a nearby aspartic acid residue furnished a fourth protein side chain ligand; this aspartate was conserved in the pea enzyme but involved instead in a salt link with a nearby arginine. Subsequent structure determinations of the β-CAs from M. thermoautotrophicum (16) and E. coli (23) provided additional examples of the two alternate ligand coordinations in this class of enzyme. So far, there is no structural evidence that a particular β-CA can flip between alternate coordination geometries.

A reaction mechanism similar to that proposed for the α-CAs has been suggested for the β-class enzymes, including the attack of a zinc-bound hydroxide on the carbon dioxide substrate and subsequent proton transfer from a new water ligand (26). Mitsuhashi et al. (24) have described a variation on this scheme in which the observed zinc-bound aspartate (Asp37 in Rv1284) is more directly involved. The authors propose that this residue functions as a base that activates a water molecule and that the resulting hydroxide displaces the aspartyl group as the fourth zinc ligand in the next steps of the reaction. The

---

**FIG. 4. Structure-based sequence alignments.** Structures were aligned in O. Secondary structural elements of Rv1284 are indicated above the sequence, whereas segments that are structurally equivalent to those of Rv1284 are underlined. Shading of the sequences was performed with BOXSHADE (K. Hoffman, unpublished program) with a cutoff of 0.7. Only the first of the two duplicated CA domains of Protein Data Bank code 1DDZ is illustrated; the last 10 residues of Protein Data Bank code 1I6P are also omitted.

**FIG. 5. Enzymatic assays.** CA activity was measured in TAPS buffer, pH 8.4. The non-enzymatic hydration reaction (Blank) and the negative control approached equilibrium in <20 s, as did the reaction including Rv1284. The sample containing Rv3588c (2.5 μM) showed a much faster reaction. However, that catalyzed by bovine α-CA II (3.3 nM), the positive control, is significantly faster.
roles of the aspartyl side chain and the neighboring arginine residue (the equivalents of Asp\(^{37}\) and Arg\(^{39}\) in Rv1284, both of which are conserved in the whole family) have been analyzed by site-directed mutagenesis for the \textit{M. thermoaautotrophicum} enzyme (34). The Asp→Ala change reduced the \(k_{\text{cat}}/K_m\) value, but not to an extent that would suggest a vital role for the residue in the reaction mechanism. A much larger reduction in \(k_{\text{cat}}/K_m\) was observed for the Arg→Ala mutant, which was interpreted to mean that the aspartyl residue, once free of the salt link, coordinates the zinc. Consistent with this hypothesis, the double mutant had a higher activity than the Arg→Ala variant. Use of an imidazole buffer rescued activity of both the Asp→Ala and the double mutants, suggesting that this buffer could take on the role of the mutated groups in the proton transfer step. Site-directed mutagenesis of two side chains in the active site of the \textit{Arabidopsis thaliana} enzyme (corresponding to His\(^{209}\) and Tyr\(^{205}\) in Protein Data Bank code 1EKJ) suggested that these residues are important for efficient proton transfer in that enzyme (35). However, these residues are not conserved in the whole \(\beta\)-CA family.

Our structural studies of two \textit{M. tuberculosis} gene products, Rv1284 and Rv3588c, indicate that they are both \(\beta\)-CAs. Our functional assays confirm this assignment for Rv3588c, although not for Rv1284 (Fig. 5). Metal content analysis indicates that the lack of activity is correlated with a significant zinc depletion at the active site of Rv1284. This interpretation is further supported by the Rv1284 structure itself; the electron density of the metal is only half that observed in Rv3588c. However, simply adding more zinc causes the protein to precipitate. Because the His tag in the present genetic construct may act as an internal zinc-chelating agent and thus cause the observed precipitation, we are planning to modify the construct in the future to produce protein without the tag. Both tertiary structural (DALI) (36) and motif-based (SPASM) (37) searches of the structural data bases indicate that the structure of Rv1284 is only consistent with a CA activity. Our Rv3588c crystals are grown at a pH value at which the enzyme has no activity. The observed structure, therefore, lends further support to the hypothesis that the inactive enzyme has four side chains coordinating the active site zinc atom.

The six \(\beta\)-CA structures now known enable us to define the structural core of this class of enzyme and to recognize the differences associated with a change from three to four protein side chain coordination at the zinc. The structures indicate that the active site metal of the \(\beta\)-CA enzymes is located in a shallower pocket than is seen in the \(\alpha\)-CA family. However, the accessibility of the zinc ion varies between members of the \(\beta\)-CA family, ranging from the almost fully closed binding site cavity in Rv1284 to the \textit{M. thermoaautotrophicum} \(\beta\)-CA, which is open to solvent and indeed binds to a molecule of HEPES buffer (16). The accessibility is primarily determined by the loop connecting the first \(\alpha\)-helix in the structure to the first \(\beta\)-strand. In Rv1284 and the pea enzyme (26) this loop crosses over the active site, and a side chain (Met\(^{74}\) and Gln\(^{151}\), respectively) points toward the zinc. The glutamine residue of the pea enzyme has been suggested to interact with the gaseous carbon dioxide substrate as well as the zinc-bound bicarbonate. In Rv1284, the methionine side chain closes off the active site cavity, helping to form a mostly hydrophilic internal lining. The only polar residue, His\(^{84}\), is interacting with the hydroxyl group of Ser\(^{74}\) at N\(_1\) and a water molecule at N\(_2\). This water molecule is in turn positioned by hydrogen bonds with main chain carbonyl oxygen (from residue 24′) and nitrogen (from residue 38). A fourth (somewhat longer) hydrogen bond from this water links it to the nitrogen atom of the thiocyanate ligand. Because this nitrogen is positioned only 2.9 Å from the zinc hydroxide, we suggest that thiocyanate is a good starting model for the docking of carbon dioxide to the active site of this \(\beta\)-CA (see Fig. 2). Rv3588c lacks the occluding loop of Rv1284 because of local disorder and therefore belongs to the category of \(\beta\)-CA enzymes with more open active sites. It lacks the histidine residue that lines the cavity in Rv1284 but retains the tyrosine residue (Tyr\(^{39}\) in Rv3588c) implicated in proton transfer for the \textit{A. thalina} enzyme (35).

Bicarbonate/carbon dioxide is important in fatty acid biosynthesis as well as in the synthesis of various small molecules. Merlin et al. (38) have shown that the \(\beta\)-CA encoded by the \textit{can} (previously designated \textit{yadF}) gene is required for the growth of \textit{E. coli} in air, where it is estimated that the demand for bicarbonate is \(10^3\) to \(10^4\)-fold greater than can be produced by an uncatalyzed hydration mechanism. Whereas multiple CAs in a single organism are not uncommon (humans have at least 14 \(\alpha\)-class enzymes) (7), our studies of two of the three CAs that have been identified in \textit{M. tuberculosis} demonstrate that these enzymes have a number of distinct properties, and so they presumably have different functional niches. Several studies have suggested that Rv1284 has a particularly vital role. The closer relationship of this protein to the enzyme from the thermostophilic methanoarchaeon \textit{M. thermoaautotrophicum} seems interesting in this context, although it is not yet clear which aspects of this unusual organism are most relevant. The second mycobacterial \(\beta\)-CA, Rv3588c, appears to be very important in mycobacterial survival, although its closer relationships to the plant and bacterial enzymes may make it a less desirable target for drug design. We anticipate that such information will assist in the search for new inhibitors that can make effective use of the differences between these and the \(\alpha\)-class CAs found in humans.

Acknowledgment—We thank Stewart Cole of the Institut Pasteur for \textit{M. tuberculosis} DNA.

REFERENCES

1. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., III, Tekala, P.,Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Pelwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., James, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squires, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrett, B. G. (1998) Nature 393, 537–544.

2. Sassetti, C. M., and Rubin, E. J. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 12989–12994.

3. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Mol. Microbiol. 48, 77–84.

4. Betts, J. C., Lukey, P. T., Robb, L. C., Mchdm, R. A., and Duncn, K. (2002) Mol. Microbiol. 43, 717–731.

5. Smith, K. S., Jakubczik, C., Whittam, T. S., and Ferry, J. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 15184–15189.

6. Lindskog, S., and Liljas, A. (1993) Curr. Opin. Struct. Biol. 3, 915–920.

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

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

9. Khalilaf, R. G. (1971) J. Biol. Chem. 246, 2561–2573.

10. McPherson, A. J. (1982) Preparation and Analysis of Protein Crystals, John Wiley and Sons, New York.

11. Leslie, A. G. (1999) Acta Crystallogr. Sect. D55, 1696–1702.

12. Evans, P. R. (1993) Proceedings of CCP4 Study Weekend on Data Collection & Processing, pp. 114–122, Daresbury Laboratory, Warrington, England, United Kingdom.

13. Collaborative Computational Project Number 4. (1994) Acta Crystallogr. D50, 769–763.

14. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326.

15. Navaza, J. (1994) Acta Crystallogr. A50, 157–163.

16. Shrop, P., Smith, K. S., Iverson, T. M., Ferry, J. G., and Rees, D. C. (2001) J. Biol. Chem. 276, 10995–10999.

17. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M., and Jones, T. A. (2001) in International Tables for Crystallography: Crystallography of Biological Macromolecules (Bosmann, M. G., and Arnold, E., eds) Vol. F, pp. 353–356, Kluwer Academic Publishers, Dordrecht, The Netherlands.

18. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D53, 240–255.

19. Poteettter, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) Acta Crystallogr. D59, 1131–1137.

20. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–119.
21. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Acta Crystallogr. D57, 122–133
22. Brunger, 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., and Warren, G. L. (1998) Acta Crystallogr. D54, 905–923
23. Cronk, J. D., Endrizzi, J. A., Cronk, M. R., O’Neill J, W., and Zhang, K. Y. (2001) Protein Sci. 10, 911–922
24. Mitsuhashi, S., Mizushima, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Moriya, H., Ueki, T., Miyachi, S., and Tsukihara, T. (2002) J. Biol. Chem. 275, 5521–5526
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) Nucleic Acids Res. 28, 235–242
26. Kimber, M. S., and Pai, E. F. (2000) EMBO J. 19, 1407–1418
27. Kleywegt, G. J., and Jones, T. A. (1997) Methods Enzymol. 277, 525–545
28. Branden, C. I., and Tooze, J. (1999) Introduction to Protein Structure, Garland Publishing, New York
29. Jones, T. A., and Thirup, S. (1986) EMBO J. 5, 819–822
30. Smith, K. S., and Ferry, J. G. (1999) J. Bacteriol. 181, 6247–6253
31. Lindskog, S. (1997) Pharmacol. Ther. 74, 1–20
32. Palmberg, P. (1995) Arch. Ophthalmol. 113, 985–986
33. Maren, T. H. (1995) J. Glaucoma 4, 49–62
34. Smith, K. S., Ingram-Smith, C., and Ferry, J. G. (2002) J. Bacteriol. 184, 4240–4245
35. Rowlett, R. S., Tu, C., McKay, M. M., Preiss, J. R., Loomis, R. J., Hicks, K. A., Marchione, R. J., Strong, J. A., Donovan, G. S., Jr., and Chamberlin, J. E. (2002) Arch. Biochem. Biophys. 404, 197–209
36. Holm, L., and Sander, C. (1997) Nucleic Acids Res. 25, 231–234
37. Kleywegt, G. J. (1999) J. Mol. Biol. 285, 1887–1897
38. Merlin, C., Masters, M., McAteer, S., and Coulson, A. (2003) J. Bacteriol. 185, 6415–6424
39. Engh, R. A., and Huber, R. (1991) Acta Crystallogr. A47, 392–400
40. Kleywegt, G. J., and Jones, T. A. (1996) Structure (Lond.) 4, 1395–1400