Carbonic anhydrase is an ancient enzyme widespread in prokaryotes

Kerry S. Smith*, Claudia Jakubzick*, Thomas S. Whittam†, and James G. Ferry*‡

*Department of Biochemistry and Molecular Biology, and †Department of Biology, The Pennsylvania State University, University Park, PA 16802

Communicated by Arnold L. Demain, Massachusetts Institute of Technology, Cambridge, MA, October 18, 1999 (received for review March 26, 1999)

Carbonic anhydrases catalyze the reversible hydration of CO_2 and are ubiquitous in highly evolved eukaryotes. The recent identification of a third class of carbonic anhydrase (γ class) in a methanoarchaeon and our present finding that the β class also extends into thermophilic species from the Archaea domain led us to initiate a systematic search for these enzymes in metabolically and phylogenetically diverse prokaryotes. Here we show that carbonic anhydrase is widespread in the Archaea and Bacteria domains, and is an ancient enzyme. The occurrence in chemolithoautotrophic species occupying deep branches of the universal phylogenetic tree suggests a role for this enzyme in the proposed autotrophic origin of life. The presence of the β and γ classes in metabolically diverse species spanning the Archaea and Bacteria domains demonstrates that carbonic anhydrases have a far more extensive and fundamental role in prokaryotic biology than previously recognized.

Carbonic anhydrase is a zinc-containing enzyme catalyzing the reversible hydration of CO_2 [$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$]. Since the discovery of the enzyme in bovine erythrocytes in 1933 (1), isozymes have been found in virtually all mammalian tissues and cell types, where they function in CO_2 transport and other physiological processes (2). Carbonic anhydrases are also abundant in plants and unicellular green algae, where they are essential for photosynthetic CO_2 fixation (3). Although they are ubiquitous in highly evolved organisms from the Eukarya, the extent to which carbonic anhydrases occur in the Archaea and Bacteria domains is unknown; the enzyme has been purified from only five prokaryotic species (4–8) since 1963, when it was first identified in *Neisseria sicca* (9).

All carbonic anhydrases are divided into three distinct classes (α, β, and γ) that evolved independently and have no sequence homology (10). Carbonic anhydrases from mammals (including the 10 active human isoforms) (10, 11), together with the two periplasmic enzymes from the unicellular green alga *Chlamydomonas reinhardtii* (12, 13), belong to the α class. The β class is comprised of enzymes from the chloroplasts of both monocotyledonous and dicotyledonous plants (10). Within the Bacteria domain, the enzymes purified from *Neisseria gonorrhoeae* and *Escherichia coli* also belong to the α and β classes (14, 15), respectively. Recently, a gene encoding a putative β type carbonic anhydrase in the methanoarchaeon *Methanobacterium thermoautotrophicum* was expressed in *E. coli* and found to encode a thermostable carbonic anhydrase (16). A carbonic anhydrase purified from the archaeon *Methanosarcina thermophila* is decidedly distinct from the α and β classes, and is the prototype of a different class, the γ class (17).

Our detection of documented β and γ carbonic anhydrases in the methanoarchaea prompted us to investigate the distribution of carbonic anhydrases in the Archaea and Bacteria domains. The results indicate not only that carbonic anhydrases are widely distributed in prokaryotes, but that the β and γ classes are predominant. Perhaps most importantly, the results indicate the remarkable feature that both the β and γ classes of carbonic anhydrase are ancient enzymes. The presence of both the β and γ classes in thermophilic chemolithoautotrophs suggests that ancient CO_2-fixation pathways depended on carbonic anhydrase for efficient CO_2 fixation.

Materials and Methods

Western Blot Analysis. Cells were suspended in 50 mM potassium phosphate (pH 6.8), and passed twice through a chilled French pressure cell at 138 MPa; the cell extract was collected after centrifugation at 20,000 × *g* for 20 min to remove unbroken cells. Carbonic anhydrase activity was measured by a modification of the electrometric method of Wilbur and Anderson (18). Cell extract proteins were separated by SDS/PAGE on 12.0% gels and electrotransferred to a poly(vinylidene difluoride) (PVDF) membrane (Immobilon; Millipore) as recommended (Bio-Rad). Additional protein sites were blocked by incubating the membrane in 10 mM Tris·HCl (pH 8.0) containing 150 mM NaCl, 0.05% Tween 20, and 10% evaporated milk. Primary antisera raised to *Anabaena* PCC7120 carbonic anhydrase (α class) (19), the *M. thermoautotrophicum* ΔH Cab (β class) (16), and the *M. thermophila* Cam (γ class) (17) were used at dilutions of 1:10,000, 1:5,000, and 1:10,000, respectively. A 1:7,500 dilution of anti-rabbit IgG–alkaline phosphatase conjugate was used. The antibody–antigen complex was detected with 5-bromo-4-chloro-3-indolyl phosphate and 4-nitroblue tetrazolium chloride, as recommended (Boehringer Mannheim).

Molecular Clock Analysis. ORFs with deduced amino acid sequences showing identity with those of carbonic anhydrase were identified by either BLASTp or tBLASTn (20) searches of sequence databases.§ The amino acid sequences were aligned with CLUSTALX (21) by using a Gonnet PAM 250 weight matrix, with a gap-opening penalty of 10.0 and a gap-extension penalty of 0.05. A phylogenetic tree was constructed by the neighbor-joining algorithm of MEGA (22) by using a gamma distance estimation ($a = 2$), in which all gap sites were ignored in the distance estimation. Trees were constructed based on pairwise distance estimates of the expected number of amino acid replacements per site (0.25 in the scale bar). Percent identity and percent similarity were determined by the GAP program of the GCG Wisconsin package (Genetics Computer Group; Madison, WI), by using the algorithm of Needleman and Wunsch (23), with a gap-creation penalty of 12.0 and a gap-extension penalty of 4.0.

Phylogenetic relative-rate tests (24) were used to address the hypothesis that γ carbonic anhydrases have evolved as a molecular clock. Analysis of 20 γ carbonic anhydrase sequences from the larger A clade, rooted with the *M. thermophila* sequence from the B clade, revealed that two γ carbonic anhydrases (*Coxiella burnetii* and *Deinococcus radiodurans*) deviated significantly from the prediction of a uniform evolutionary rate. Without these two sequences, the two-cluster test for the 19 sequences showed no deviation from the molecular clock hypothesis for any

‡To whom reprint requests should be addressed at: Center for Microbial Structural Biology, 205 South Frear Laboratory, The Pennsylvania State University, University Park, PA 16802. E-mail: jgf3@psu.edu.

§located at http://capsulapedia.uchicago.edu, http://www.genome.ou.edu/, http://www.genomecorp.com/, http://www.ncbi.nlm.nih.gov/BLAST/, http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html, http://www.pseudomonas.com/, http://www.sanger.ac.uk/, and http://www.tigr.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Detection of carbonic anhydrases in microbes

Species	Molecular mass in Western blotting,[a] kDa			Specific activity,[b] units/mg
	α-class	β-class	γ-class	
Bacteria domain[c]				
Acetobacterium woodii[d]	–	22	19	7.3 ± 0.6
Bacillus subtilis[d]	–	*	*	<0.01
Carboxydothermus hydrogenoformans[dk]	–	23	19	<0.01
Clostridium thermoaceticum[dk]	–	24	18	1.0 ± 0.1
Helicobacter pylori[e]	23[l]	25[l]	–	<0.01
Methylobacterium extorquens[f]	–	23	19	1.4 ± 0.2
Pseudomonas aeruginosa[g]	–	23[l], 25[l]*	20[l]*	<0.01
Rhizobium meliloti[f]	–	24	18	1.5 ± 0.1
Rhodobacter capsulatus[f]	–	23[l]	18[l]	1.5 ± 0.2
Rhodobacter sphaeroides[f]	–	23	18	1.6 ± 0.2
Rhodospirillum rubrum[f]	–	21	19	3.6 ± 0.2
Salmonella typhimurium[g]	22	26[l]	*	0.6 ± 0.2
Staphylococcus aureus[d]	–	23	–	1.2 ± 0.1
Thermotoga maritima[hk]	–	–	–	0.01
Vibrio fischeri[g]	–	–	19	1.5 ± 0.3
Archaea domain[c]				
Methanobacterium formicicum[i]	–	21	–	<0.01
Methanobacterium thermoautotrophicum ΔH[ik]		21[l]	17[l]	0.8 ± 0.2
Methanobacterium thermoautotrophicum Marburg[ik]	–	21	17	1.3 ± 0.1
Methanococcus jannaschii[jk]	–	–	*	0.3 ± 0.1
Methanococcus maripaludis[i]	–	–	–	<0.01
Methanosaeta concilii[i]	–	–	21	3.0 ± 0.4
Methanosarcina barkeri[i]	–	–	37	1.4 ± 0.2
Methanosarcina thermophila[jk]	–	–	37[l]	1.8 ± 0.2
Methanospirillum hungateii[i]	–	24	–	7.1 ± 0.5
Natronococcus occultus[i]	–	–	–	0.01
Pyrococcus furiosus[ik]	–	–	20[l]	<0.01
Sulfolobus solfataricus[jk]	–	23	19	1.6 ± 0.2

[a]Approximate molecular masses of cross-reactive proteins. A minus sign (–) indicates no crossreactive proteins were detected. An asterisk (*) indicates that crossreactive protein expected from sequence data was not detected.

[b]Specific activity was measured at 23°C for mesophiles and 55°C for thermophiles. One unit = $(t_0 - t)/t$, where t_0 (uncatalyzed reaction) and t (catalyzed reaction) are the times required for the pH to decrease from 7.8 to 7.0.

[c]Species from the Bacteria domain belong to the following kingdoms: Gram-positive,[d] proteobacteria (ε class),[e] proteobacteria (α class),[f] proteobacteria (γ class),[g] or thermotoga.[h] Species from the Archaea domain belong to the following kingdoms: Euryarchaeota[i] and Crenarchaeota.[j]

[k]Thermophilic species.

[l]The molecular mass detected by Western blotting is similar to the molecular mass calculated from the deduced amino acid sequence.

internal node or, overall, for the entire tree ($Q = 21.7$, df = 18, $P > 0.05$). To estimate the timing of evolutionary divergence in the γ carbonic anhydrase tree, we extrapolated divergence times from the distances between the deduced amino acid sequences. For the nodes of interest, the average proportion of identical residues between pairs of sequences was used to estimate evolutionary distance (d), which was measured in terms of the number of amino acid substitutions per site (25). Distance was converted to time, based on a slope of 0.088 $d/10^6$ yr ago, calibrated from the fossil record of the vertebrate groups (26).

Results and Discussion

Genome sequencing (27, 28) of the thermophilic, obligately chemolithoautotrophic methanoarchaeon *M. thermoautotrophicum* ΔH revealed an ORF with a deduced sequence 34.3% identical to CynT, the β carbonic anhydrase of *E. coli* (29). The ORF-encoded protein, Cab (*c*arbonic *a*nhydrase *b*eta), was expressed in *E. coli* and found to have carbonic anhydrase activity (16). These results show that β carbonic anhydrases occur not only in species of the Archaea domain, but also in thermophilic prokaryotes with an obligate chemolithoautotrophic mode of growth, species that represent some of the deepest branches of the universal tree of life (30).

The recent discovery of the γ class of carbonic anhydrase in Archaea (7) and the finding that an enzyme from the β class occurs in deeply branching thermophilic chemolithotrophic methanoarchaea (16) suggest that these enzymes are more ancient in origin and more central to prokaryotic metabolism than previously recognized. This idea was further tested by examining metabolically diverse species from both the Archaea and Bacteria domains for carbonic anhydrase in cell extracts and using both activity assays and Western blotting with antisera raised against the α-class carbonic anhydrase from *Anabaena* PCC7120 (19), the β-class enzyme from *M. thermoautotrophicum* ΔH (16), or the γ-class enzyme from *M. thermophila* (17). Activity and (or) crossreacting proteins were present in the extracts of all but 4 of the 24 species for which carbonic anhydrase had not previously been documented (Table 1). No crossreactive proteins were detected with the preimmune antisera. In addition, the antisera derived against the α-, β-, and γ-carbonic anhydrases do not crossreact. Western blotting and activity assays are two independent methods to detect carbonic anhydrase, either of which could fail for one reason or another, but a positive result from either method is strong evidence for the presence of carbonic anhydrase.

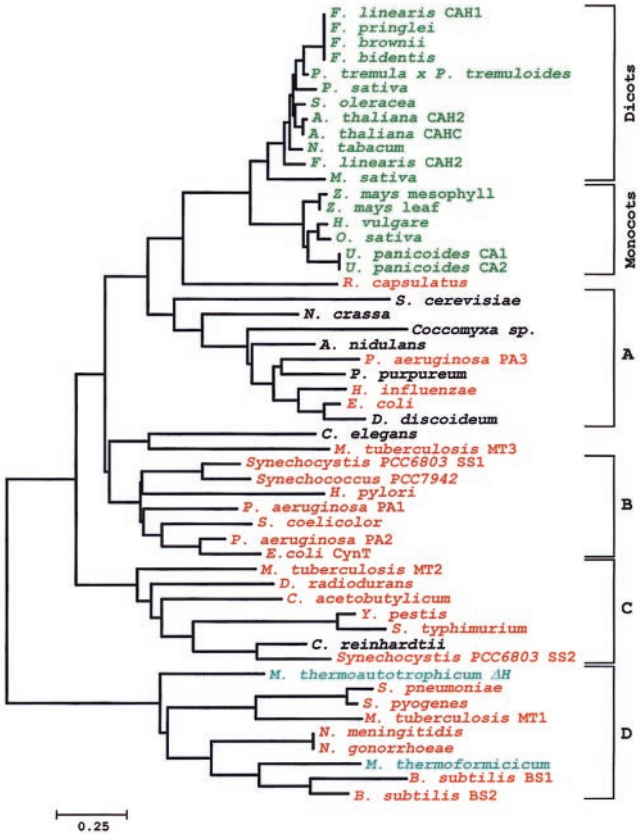

Fig. 1. Phylogeny of the β class of carbonic anhydrase. A neighbor-joining tree of proteins related to β carbonic anhydrases was constructed. Putative β carbonic anhydrases were identified in the Eukarya [plant sequences (green) and lower eukaryotes (black)], Bacteria (red), and Archaea (blue) domains. GenBank accession numbers or sequence source, their percent identity, and similarity to the *M. thermoautotrophicum* ΔH Cab (accession number 1272331) appear in parentheses, as follows, in alphabetical order: *Arabidopsis thaliana* CAH2 (1168739, 28.4, 58.0), *Arabidopsis thaliana* CAHC (115470, 26.9, 55.7), *Aspergillus nidulans* (AA786939, 18.0, 45.3), *Bacillus subtilis* BS1 (1945660, 40.4, 60.8), *Bacillus subtilis* BS2 (2293156, 39.0, 59.0), *Caenorhabditis elegans* (1049368, 32.5, 61.4), *Chlamydomonas reinhardtii* (1323549, 25.0, 53.8), *Clostridium acetobutylicum* (Genome Therapeutics, 27.2, 53.3), *Coccomyxa* sp. (1663720, 24.1, 50.6), *Deinococcus radiodurans* (The Institute for Genomic Research, 28.0, 53.6), *Dictyostelium discoideum* (1513236, 31.1, 56.3), *Escherichia coli* CynT (1657535, 34.2, 54.4), *Escherichia coli* (1723190, 29.0, 54.0), *Flaveria bidentis* (1168745, 27.5, 53.3), *Flaveria brownii* (1168746, 27.5, 53.3), *Flaveria linearis* CAH1 (1168738, 27.5, 56.3), *Flaveria linearis* CAH2 (882243, 27.8, 56.3), *Flaveria pringlei* (1168747, 27.5, 53.3), *Haemophilus influenzae* (1175500, 28.7, 57.3), *Helicobacter pylori* (2313081, 31.9, 53.4), *Hordeum vulgare* (729003, 25.3, 52.5), *Medicago sativa* (1938227, 26.0, 52.1), *Methanobacterium thermoautotrophicum* ΔH (1272331), *Methanobacterium thermoformicicum* (1279772, 51.8, 73.8), *Mycobacterium tuberculosis* MT1 (1722951, 34.4, 60.0), *Mycobacterium tuberculosis* MT2 (2950411, 26.6, 49.1), *Mycobacterium tuberculosis* MT3 (1877328, 28.4, 50.3), *Neisseria gonorrhoeae* (Univ. of Oklahoma Genome Center, 39.4, 64.7), *Neisseria meningitidis* (Sanger Centre, 39.4, 64.7), *Neurospora crassa* (AA901623, 29.8, 49.6), *Nicotiana tabacum* (115473, 25.0, 51.2), *Oryza sativa* (606817, 24.8, 53.3), *Pisum sativum* (115471, 22.6, 49.4), *Populus tremula* × *Populus tremuloides* (1354517, 23.3, 51.5), *Porphyridium purpureum* (1395172, 26.7, 55.8), *Pseudomonas aeruginosa* PA1 (Univ. of Washington/PathoGenesis, 31.5, 55.2), *Pseudomonas aeruginosa* PA2 (Univ. of Washington/PathoGenesis, 37.0, 59.3), *Pseudomonas aeruginosa* PA3 (Univ. of Washington/PathoGenesis, 29.6, 54.3), *Rhodobacter capsulatus* (Univ. of Chicago/Institute of Molecular Genetics, 27.4, 54.2), *Saccharomyces cerevisiae* Nce3p (1277232, 29.8, 48.2), *Salmonella typhimurium* (2460259, 28.7, 54.8), *Spinacia oleracea* (115472, 23.3, 52.7), *Streptococcus pneumoniae* (The Institute for Genomic Research, 34.8, 62.6), *Streptococcus pyogenes* (Univ. of Oklahoma Genome Center, 31.7, 62.7), *Streptomyces coelicolor* (Sanger Centre, 28.7, 52.1), *Synechococcus* PCC7942 (118069, 28.7, 56.9), *Synechocystis* PCC6803 SS1 (1653251, 30.9, 59.4), *Synechocystis* PCC6803 SS2 (1001130, 23.3, 51.5), *Urochloa panicoides* CA1 (882248, 20.7, 52.7), *Urochloa panicoides* CA2 (882246, 20.7, 52.7), *Yersinia pestis* (Sanger Centre, 28.1, 53.8), *Zea mays* mesophyll (606815, 23.0, 55.2), and *Zea mays* leaf (606811, 23.0, 55.6).

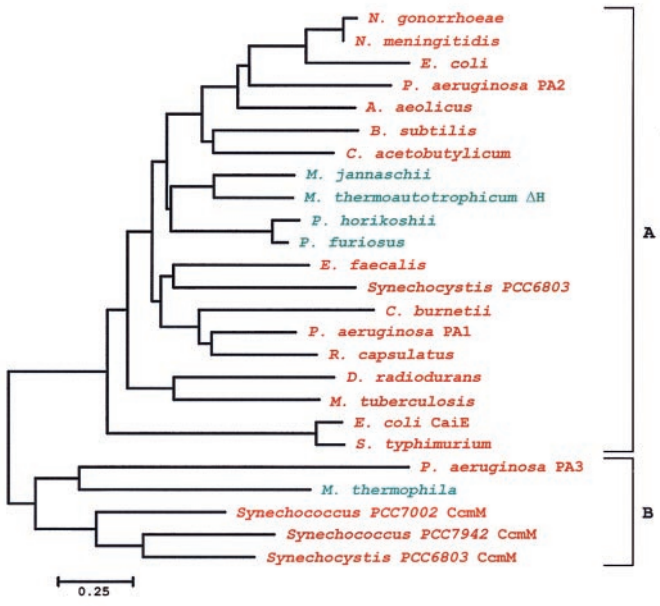

Fig. 2. Phylogeny of the γ class of carbonic anhydrase. A neighbor-joining tree of proteins related to γ carbonic anhydrases was constructed. Putative γ carbonic anhydrases were identified in the Bacteria (red) and the Archaea (blue) domains. GenBank accession nos. or sequence source and their percent identity and similarity to the *M. thermophila* Cam (accession no. 729004) appear in parentheses, as follows: *Aquifex aeolicus* (2983997, 26.9, 48.5), *Bacillus subtilis* (2293300, 24.9, 45.5); *Clostridium thermoaceticum* (Genome Therapeutics, 28.7, 51.5), *Coxiella burnetii* (1075473, 28.7, 54.3, *Deinococcus radiodurans* (The Institute for Genomic Research, 28.0, 51.7), *Enterococcus faecalis* (The Institute for Genomic Research, 26.8, 61.8), *Escherichia coli* (1176852, 22.6, 49.6), *Escherichia coli* CaiE (729006, 24.2, 49.5), *Methanobacterium thermoautotrophicum* ΔH (2622711, 30.7, 60.7), *Methanococcus jannaschii* (2493491, 31.8, 61.8), *Methanosarcina thermophila* (729004), *Mycobacterium tuberculosis* (1666135, 37.2, 59.1), *Pseudomonas aeruginosa* PA1 (729466, 28.0, 52.6), *Pseudomonas aeruginosa* PA2 (Univ. of Washington/PathoGenesis, 24.2, 52.4), *Pseudomonas aeruginosa* PA3 (Univ. of Washington/PathoGenesis, 29.4, 49.2), *Pyrococcus furiosus* (Center of Marine Biotechnology/Univ. of Utah, 31.4, 56.8), *Pyrococcus horikoshii* (3258020, 30.8, 58.0), *Rhodobacter capsulatus* (Univ. of Chicago/Institute of Molecular Genetics, 25.3, 52.9), *Salmonella typhimurium* (U81260, 23.1, 51.1), *Synechocystis* PCC6803 (1652147, 23.4, 53.2), *Synechocystis* PCC6803 CcmM (1651846, 36.3, 58.5), *Synechococcus* PCC7002 CcmM (2331052, 37.5, 60.4), and *Synechococcus* PCC7942 CcmM (416775, 37.2, 60.2).

A simple explanation for undetectable activity in the thermophiles *Carboxydothermus hydrogenoformans* and *Pyrococcus furiosus* is that the upper temperature limit of the assay (55°C) is well below the expected activity optimum (85°C) for the enzymes. The decreased solubility of CO_2 at higher temperatures under atmospheric pressure precludes the determination of activity. The absence of activity in mesophilic or thermophilic species containing crossreacting proteins could also be explained by either inactivation or nonspecific crossreactivity with another protein. The absence of a crossreactive protein in cases where sequence data indicate a putative carbonic anhydrase gene is not unexpected. It is likely that expression of these genes is strictly regulated; each carbonic anhydrase may not be expressed under the growth condition tested for each species. Thus, the inability

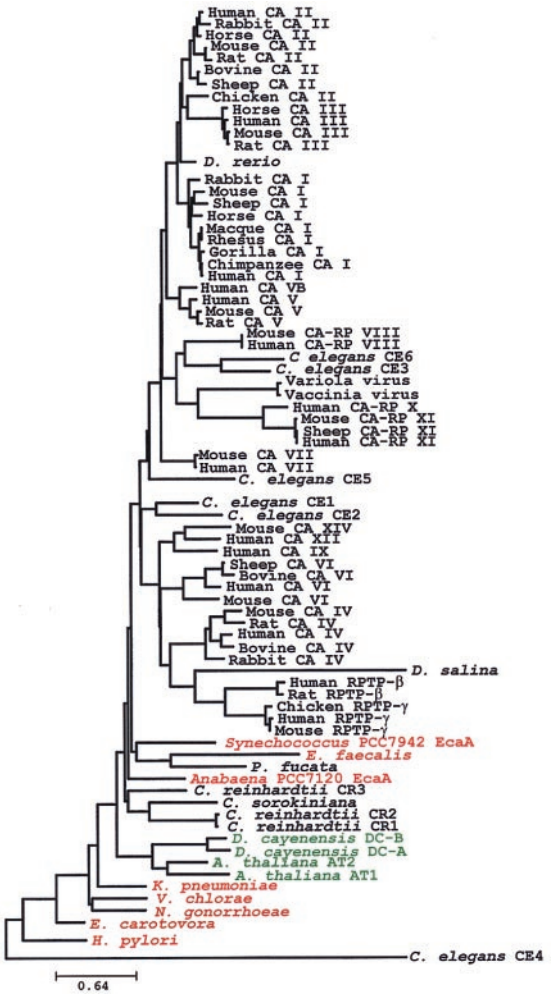

Fig. 3. Phylogeny of the α class of carbonic anhydrase. A neighbor-joining tree of proteins related to α carbonic anhydrases was constructed. Putative α carbonic anhydrase were identified in the Eukarya [plant sequences (green) and other eukaryotes (black)] and the Bacteria (denoted in red) domain. GenBank accession nos. or sequence source and their percent identity and similarity to the *Anabaena* PCC7120 α carbonic anhydrase (accession no. 1825485) appear in parentheses, as follows: *Arabidopsis thaliana* AT1 (1916014, 29.8, 39.3), *Arabidopsis thaliana* AT2 (AAD29832, 28.7, 38.0), bovine CA II (115453, 37.8, 48.3), bovine CA IV (1575298, 26.1, 35.3), bovine CA VI (1526572, 32.5, 38.3), *Caenorhabditis elegans* CE1 (1109858, 34.5, 42.3), *Caenorhabditis elegans* CE2 (2493489, 31.6, 39.1), *Caenorhabditis elegans* CE3 (746517, 27.2,35.3), *Caenorhabditis elegans* CE4 (CAA92190, 31.7, 36.5), *Caenorhabditis elegans* CE5 (AAA20616, 23.1, 41.1), *Caenorhabditis elegans* CE6 (3217404, 25.2, 34.1), chicken CA II (115454, 34.5, 41.6), chicken receptor-type tyrosine phosphatase γ (AAB16910, 27.3, 35.1), chimpanzee CA I (478302, 40.7, 47.3), *Chlamydomonas reinhardtii* CR1 (115447, 24.4, 32.4), *Chlamydomonas reinhardtii* CR2 (115455, 23.1, 30.7), *Chlamydomonas reinhardtii* CR3 (2301259, 37.0, 44.4), *Chlorella sorokiniana* (3133261, 17.4, 26.1), *Danio rerio* (3123190, 38.2, 45.3), *Dioscorea cayenensis* DCA (1364059, 30.6, 38.3), *Dioscorea cayenensis* DCB (2147328, 32.2, 40.8), *Dunaliella salina* (1431878, 26.7, 34.0), *Enterococcus faecalis* (The Institute for Genomic Research, 30.2, 35.7), *Erwinia carotovora* (2773324, 39.5, 50.6), gorilla CA I (478303, 41.2, 46.9), *Helicobacter pylori* (2314346, 23.4, 32.3), horse CA I (115448, 40.7, 46.9), horse CA II (223999, 37.5, 45.5), horse CA III (115461, 32.3, 45.1), human CA I (115449, 40.7, 46.9), human CA II (115456, 37.5, 45.5), human CA III (1070516, 36.2, 45.1), human CA IV (115465, 25.9, 35.7), human CA V (461680, 31.0, 37.5), human CA VB (6005723, 34.1, 41.0), human CA VI (1070519, 31.6, 40.2), human CA VII (1168744, 34.2, 42.5), human CA-RP VIII (461681, 28.2, 35.5), human CA IX (2135876, 29.2, 38.5), human CA-RP X ((Ref. 39), 20.2, 28.2), human CA-RP XI (3283386, 26.5, 33.0), human CA XII (2708639, 28.9, 37.2), human receptor-type tyrosine phosphatase β (190744, 23.7, 37.2), human receptor-type tyrosine phosphatase γ (292411, 27.3, 35.0), *Klebsiella pneumoniae* (2773319, 39.1, 46.9), macaque CA I (461679, 38.5, 46.9), mouse CA I (1345656, 36.3, 44.5), mouse CA II (115457, 38.1, 45.6), mouse CA III (115463, 37.3, 44.9), mouse CA IV (2493487, 27.8, 36.7), mouse CA V (1168742, 28.2, 36.7), mouse CA VI (18761, 29.8, 37.2), mouse CA VII ((Ref. 40), 37.4, 44.9), mouse CA-RP VIII (115476, 30.5, 37.4), mouse CA-RP XI (AF050105, 28.1, 34.2), mouse CA XIV (BAA78709, 29.1, 36.5), mouse receptor-type tyrosine phosphatase γ (293774, 29.2, 36.5), *Neisseria gonorrhoeae* (1841441, 36.4, 45.6), *Pinctada fucata* (1480031, 20.0, 30.0), rabbit CA I (115452, 36.8, 46.8), rabbit CA II (1168741, 39.1, 46.2), rabbit CA IV (1345657, 26.3, 33.2), rat CA II (115459, 37.2, 44.7), rat CA III (2708636, 38.2, 45.3), rat CA IV (1345658, 26.5, 33.2), rat CA V (1168743, 34.6, 41.5), rat receptor-type tyrosine phosphatase β (487781, 25.5, 34.4), rhesus CA I (115450, 38.7, 47.1), sheep CA I (1345651, 38.3, 46.1), sheep CA II (115460, 36.9, 44.0), sheep CA VI (115469, 34.8, 40.0), Sheep CA-RP XI (Y07785, 28.8, 35.0), *Synechococcus* PCC7942 (18254581, 37.6, 43.9), vaccinia virus (335595, 20.0, 35.0), variola virus (416745, 25.0, 41.7), and *Vibrio cholerae* (The Institute for Genomic Research, 35.1, 42.3).

to detect either activity or crossreacting proteins does not rule out the existence of carbonic anhydrases.

By searching both DNA and protein databases, we identified 51 ORFs (listed in the legend of Fig. 1) with deduced sequences and having significant identity and similarity to the β carbonic anhydrase from *M. thermoautotrophicum* ΔH. Of the 26 ORFs from species of the Bacteria and Archaea domains, only two are known to encode documented carbonic anhydrases. Distinct from all other β carbonic anhydrase sequences, the plant sequences form two monophyletic groups representing monocotyledons and dicotyledons. All but one (*Rhodobacter capsulatus*) of the prokaryotic sequences separate into four clades (designated *A–D* in Fig. 1), three of which (*A*, *C*, and *D*) are strongly supported by the bootstrap values (data not shown). The clades denoted *A* and *C* both contain a mixture of sequences from the Eukarya and Bacteria domains. The first of two exclusively prokaryotic clades (*B*) consists primarily of sequences from Gram-negative species in the Bacteria domain, whereas the second exclusively prokaryotic clade (*D*) is distantly related to the other clades and consists primarily of sequences from Gram-positive species. Each of the these two clades contains one of two documented prokaryotic β carbonic anhydrases, CynT from *E. coli* and Cab from *M. thermoautotrophicum* ΔH. The results are consistent with the remaining putative sequences representing valid β carbonic anhydrases in these two clades. Indeed, all of the protein sequences identified in this search contain essential residues, signature residues for β-class carbonic anhydrases, providing strong evidence that they are β-class carbonic anhydrases. These signature residues are Asp-152, which is essential for activity of the pea β carbonic anhydrase, and a unique motif (Cys-Xaa$_n$-His-Xaa$_2$-Cys), identified as essential for ligation of the active-site zinc (31, 32). Therefore, the phylogenetic analysis in Fig. 1 bolsters the results in Table 1, and provides further support for the conclusion that the β class is widespread in metabolically and phylogenetically diverse prokaryotes.

We identified 25 ORFs (listed in the legend of Fig. 2) from the Bacteria and Archaea domains that have deduced sequences with significant identity and similarity to the γ carbonic anhydrase (Cam) from *M. thermophila*. A clade (designated *B* in Fig. 2) is formed by Cam and the sequences deduced from an ORF from *Pseudomonas aeruginosa* (designated PA3) and the *ccmM* genes of *Synechocystis* PCC6803, *Synechococcus* PCC7002, and *Synechococcus* PCC7942. The cyanobacterial genes are essential for growth at low pCO_2, consistent with roles for carbonic anhydrases in the transport and concentration of CO_2 for fixation into cell material (33). The clade denoted *B* in Fig. 2, containing the only documented γ carbonic anhydrase (Cam), is distinct from the remaining sequences, which form a second larger clade (*A*). Although there are no sequences of documented enzymes in the larger clade, Western blotting results (Table 1) suggest that three members (*P. furiosus*, *P. aeruginosa*, and *M. thermoautotro-*

phicum ΔH) have a γ carbonic anhydrase. The γ-class carbonic anhydrase from *M. thermophila* is unique in that the active-site zinc is ligated with a histidine residue from one subunit and by two histidine residues from an adjacent subunit (34). A unique motif (His–Xaa_n-His-XAA_4-His) has been identified in the γ-class carbonic anhydrase from *M. thermophila* for ligation of the active-site zinc. This motif is present in all of the putative γ-class carbonic anhydrases and strongly supports their validity as γ-class carbonic anhydrases. In addition, Arg-59 and Gln-75, which are essential for activity, are also present in all of the putative γ carbonic anhydrase sequences (B. C. Tripp and J. G. F., personal communication).

The database searches buttress the results in Table 1, and further support the hypothesis that carbonic anhydrases of the β and γ classes are broadly dispersed in the Bacteria domain and in both kingdoms in the Archaea domain. We found only eight ORFs in the Bacteria domain and none in the Archaea domain with amino acid sequences similar to the α-class carbonic anhydrase from *Anabaena* PCC7120 (Fig. 3). In addition to the prokaryotic sequences, the α class is composed of 10 mammalian isozymes (CA I–VII, IX, XII, and XIV), several related mammalian and viral sequences that encode proteins lacking carbonic anhydrase activity, and several sequences from vertebrates, plants, nematodes, and algae that are expected to encode either carbonic anhydrases or carbonic anhydrase-related proteins. The mammalian carbonic anhydrase-related proteins (CA-RP VIII, X, and XI), two subtypes of receptor-type protein tyrosine phosphatases (RPTPβ and RPTPγ), and two pox virus (vaccinia and variola) transmembrane proteins do not catalyze the reversible hydration of CO_2 because of the absence of one or more of the three histidine residues that coordinate the active-site zinc (10). The three histidine residues coordinating the active-site zinc in the human isozymes are completely conserved in all of the prokaryotic α-class carbonic anhydrase sequences as are other residues that have been shown to be important for catalysis in the human isozymes, such as Gln-92, Glu-117, Ala/Val-121, and Val-143 of human CA II. Furthermore, crossreacting proteins to the α antisera were detected only in *Helicobacter pylori* and *Salmonella typhimurium* (Table 1), further suggesting that the β and γ classes predominate in prokaryotes. The evolution of the α carbonic anhydrases from a common ancestral gene 0.5–0.6 billion years ago (10) is consistent with the evidence that few prokaryotic genomes encode α carbonic anhydrases.

Analysis of the sequences in Fig. 2 gives divergence times that place the root of the γ class approximately 4.2 billion years ago, with the major cluster of bacterial and archaeal carbonic anhydrases having a common ancestor approximately 2.2 billion years ago. The carbonic anhydrases must then be extremely ancient enzymes, existing before the divergence of the Archaea and Bacteria domains (25). Analysis of the sequences in Fig. 1 indicates extensive divergence in the β class and a strong departure from the molecular clock, suggesting that there has been accelerated divergence in the primary structures of the different clades. The extensive sequence diversity suggests that the β class also has ancient origins, although deviations from the molecular clock preclude an estimate of divergence times.

Additional support for an ancient origin of carbonic anhydrase is the presence of the β and γ classes in thermophiles (Table 1 and Figs. 1 and 2) representing deep branches of the universal tree of life (*C. hydrogenoformans*, *M. thermoautotrophicum*, *Methanococcus jannaschii*, *Sulfolobus solfataricus, P. furiosus*, *Pyrococcus horikoshii*, and) (30). Some of these species (the first four just mentioned) are capable of a chemolithoautotrophic mode of growth and use ancient CO_2-fixation pathways. We propose that these pathways depend on carbonic anhydrase for efficient CO_2 fixation, analogous to the more evolutionarily recent photosynthetic organisms, in which the enzyme is essential for efficient CO_2 transport into the cell and elevation of the CO_2 concentration near the active site of the CO_2-fixing enzyme ribulose bisphosphate carboxylase (3, 35, 36). Some of the "CO_2"-fixation enzymes, such as pyruvate carboxylase and phospho*enol*pyruvate (PEP) carboxylase utilize HCO_3^- (37); thus, interconversion between CO_2 and HCO_3^- is another potential role for carbonic anhydrase. The deep branching of obligate chemolithoautotrophs in the universal tree of life is essential to the theory of an autotrophic origin of life (24). This theory is now further supported by the observation that carbonic anhydrase, which plays critical roles in CO_2-fixation pathways, is an ancient enzyme and is found in deeply branching chemolithoautotrophic species.

The results presented here demonstrate that carbonic anhydrases are far more prevalent in prokaryotes and distributed among far more metabolically diverse species than previously recognized, reflecting the importance of this enzyme in all three domains of life. Evidence for carbonic anhydrase was obtained for freshwater, marine, mesophilic, thermophilic, aerobic, anaerobic, pathogenic, symbiotic, methylotrophic, acetotrophic, methanogenic, acetogenic, autotrophic, heterotrophic, and photosynthetic species. Analysis of the genome sequences of *Borrelia burgdorferi*, *Chlamydia trachomatis*, *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Treponema pallidum* from the Bacteria domain and *Archaeoglobus fulgidus* from the Archaea domain does not indicate the presence of ORFs with deduced sequences that have significant identity to α, β, or γ carbonic anhydrases, suggesting either that there are undiscovered classes of these enzymes or that these prokaryotes do not require carbonic anhydrase activity.

Prokaryotic carbonic anhydrases may also have medical importance. The results reported in Table 1 and in Figs. 1, 2, and 3 indicate that carbonic anhydrases are present in several pathogenic species, suggesting the possibility that this enzyme may be required for microbial virulence. Evidence supporting this possibility was recently reported (38) for the Gram-negative, facultative intracellular pathogen *S. typhimurium*. Expression of *mig-5*, a gene encoding a putative β-class carbonic anhydrase (Fig. 1 and Table 1), is induced 24-fold upon infection in host macrophages, and it is necessary for bacterial survival within the host. Investigations into the ways that metabolically diverse groups in the Archaea and Bacteria domains use carbonic anhydrase will undoubtedly reveal novel aspects of cell physiology.

We thank Rob Barber, Jean Brenchley, Cheryl Ingram-Smith, Norman Pace, Allen Phillips, and William Whitman for their suggestions on the manuscript. Cells were provided by M. Adams, T. Anquish, P. Babitzke, P. Berg, L. Chistoserdova, C. Cox, F. Daldal, T. Donohue, H. Drake, P. Dunlap, S. Falkow, J. Gonzalez, S. Harris, C. Ingram-Smith, P. Jablonski, P. Kennelly, Y. Qi, M. Lidstrom, W. Lin, L. Ljundahl, P. Ludden, K. Miller, B. Mukhopadhyay, S. Ragsdale, F. Robb, R. Roy, J. Suflita, R. Valdivia, W. Whitman, V. Witthuhn, R. Wolfe, and S. Zinder. Antisera raised to α and γ carbonic anhydrases were kindly provided by J. Coleman and B. Alber. This work was supported by a National Institutes of Health grant (GM44661) and a National Aeronautics and Space Administration–Ames Research Laboratory Cooperative Agreement (NCC21057). C.J. was supported by the National Science Foundation (RTG-DBI-9602232), and T.S.W. was supported by a National Institutes of Health grant (AI/GM42391).

1. Meldrum, N. N. & Roughton, F. J. W. (1933) *Nature (London)* **80,** 113–142.
2. Henry, R. P. (1996) *Annu. Rev. Physiol.* **58,** 523–538.
3. Badger, M. R. & Price, G. D. (1994) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45,** 369–392.
4. Adler, L., Brundell, J., Falkbring, S. O. & Nyman, P. O. (1972) *Biochim. Biophys. Acta* **284,** 298–310.
5. Gill, S. R., Fedorka-Cray, P. J., Tweten, R. K. & Sleeper, B. P. (1984) *Arch. Microbiol.* **138,** 113–118.
6. Yagawa, Y., Shiraiwa, Y. & Miyachi, S. (1984) *Plant Cell Physiol.* **25,** 775–783.
7. Alber, B. E. & Ferry, J. G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6909–6913.

8. Braus-Stromeyer, S. A., Schnappauf, G., Braus, G. H., Gossner, A. S. & Drake, H. L. (1997) *J. Bacteriol.* **179,** 7197–7200.
9. Veitch, F. P. & Blankenship, L. C. (1963) *Nature (London)* **197,** 76–77.
10. Hewett-Emmett, D. & Tashian, R. E. (1996) *Mol. Phylogenet. Evol.* **5,** 50–77.
11. Mori, K., Ogawa, Y., Ebihara, K., Tamura, N., Tashiro, K., Kuwahara, T., Mukoyama, M., Sugawara, A., Ozaki, S., Tanaka, I. *et al.* (1999) *J. Biol. Chem.* **274,** 15701–15705.
12. Fukuzawa, H., Fujiwara, S., Yamamoto, Y., Dionisio-Sese, M. L. & Miyachi, S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4383–4387.
13. Fukuzawa, H., Fujiwara, S., Tachiki, A. & Miyachi, S. (1990) *Nucleic Acids Res.* **18,** 6441–6442.
14. Chirica, L. C., Elleby, B., Jonsson, B. H. & Lindskog, S. (1997) *Eur. J. Biochem.* **244,** 755–760.
15. Guilloton, M. B., Korte, J. J., Lamblin, A. F., Fuchs, J. A. & Anderson, P. M. (1992) *J. Biol. Chem.* **267,** 3731–3734.
16. Smith, K. S. & Ferry, J. G. (1999) *J. Bacteriol.* **181,** 6247–6253.
17. Alber, B. E. & Ferry, J. G. (1996) *J. Bacteriol.* **178,** 3270–3274.
18. Wilbur, K. M. & Anderson, N. G. (1948) *J. Biol. Chem.* **176,** 147–154.
19. Soltes-Rak, E., Mulligan, M. E. & Coleman, J. R. (1997) *J. Bacteriol.* **179,** 769–774.
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
21. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. J. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
22. Kumar, S., Tamura, K. & Nei, M. (1994) *Comput. Appl. Biosci.* **10,** 189–191.
23. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48,** 443–453.
24. Takezaki, N., Rzhetsky, A. & Nei, M. (1995) *Mol. Biol. Evol.* **12,** 823–833.
25. Grishin, N. V. (1995) *J. Mol. Evol.* **41,** 675–679.
26. Feng, D. F., Cho, G. & Doolittle, R. F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 13028–13033.
27. Nolling, J., Pihl, T. D., Vriesema, A. & Reeve, J. N. (1995) *J. Bacteriol.* **177,** 2460–2468.
28. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) *J. Bacteriol.* **179,** 7135–7155.
29. Guilloton, M. B., Lamblin, A. F., Kozliak, E. I., Gerami-Nejad, M., Tu, C., Silverman, D., Anderson, P. M. & Fuchs, J. A. (1993) *J. Bacteriol.* **175,** 1443–1451.
30. Stetter, K. O. (1996) in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)*, eds. Bock, G. R. & Goode, J. A. (Wiley, Chichester, U.K.), pp. 1–18.
31. Rowlett, R. S., Chance, M. R., Wirt, M. D., Sidelinger, D. E., Royal, J. R., Woodroffe, M., Wang, Y. F., Saha, R. P. & Lam, M. G. (1994) *Biochemistry* **33,** 13967–13976.
32. Bracey, M. H., Christiansen, J., Tovar, P., Cramer, S. P. & Bartlett, S. G. (1994) *Biochemistry* **33,** 13126–13131.
33. Price, G. D., Howitt, S. M., Harrison, K. & Badger, M. R. (1993) *J. Bacteriol.* **175,** 2871–2879.
34. Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G. & Rees, D. C. (1996) *EMBO J.* **15,** 2323–2330.
35. Price, G. D., Sultemeyer, D., Klughammer, B., Ludwig, M. & Badger, M. R. (1998) *Can. J. Bot.* **76,** 973–1002.
36. Kaplan, A. & Reinhold, L. (1999) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 539–570.
37. Jitrapakdee, S. & Wallace, J. C. (1999) *Biochem. J.* **340,** 1–16.
38. Valdivia, R. H. & Falkow, S. (1997) *Science* **277,** 2007–2011.
39. Lovejoy, D. A., Hewett-Emmett, D., Porter, C. A., Cepoi, D., Sheffield, A., Vale, W. W. & Tashian, R. E. (1998) *Genomics* **54,** 484–493.
40. Earnhardt, J. N., Qian, M., Tu, C., Lakkis, M. M., Bergenhem, N. C., Laipis, P. J., Tashian, R. E. & Silverman, D. N. (1998) *Biochemistry* **37,** 10837–10845.