The Active Site Structure of *Thalassiosira weissflogii* Carbonic Anhydrase 1†

Elizabeth H. Cox,[‡,§] George L. McLendon,[§] François M. M. Morel,*[,‡] Todd W. Lane,[‡] Roger C. Prince,[||] Ingrid J. Pickering,[⊥] and Graham N. George*[,⊥]

Center for Environmental Bioinorganic Chemistry, Departments of Chemistry and Geosciences, Princeton University, Princeton, New Jersey 08544, ExxonMobil Research and Engineering Company, Annandale, New Jersey 08801, and Stanford Synchrotron Radiation Laboratory, SLAC, Stanford University, P.O. Box 20450, MS 69, Stanford, California 94309

Received June 21, 2000; Revised Manuscript Received August 3, 2000

ABSTRACT: X-ray absorption spectroscopy at the Zn K-edge indicates that the active site of the marine diatom *Thalassiosira weissflogii* carbonic anhydrase is strikingly similar to that of mammalian α-carbonic anhydrase enzymes. The zinc has three histidine ligands and a single water at 1.98 Å. This is quite different from the *β*-carbonic anhydrases of higher plants in which zinc is coordinated by two cysteine thiolates, one histidine, and a water molecule. The diatom carbonic anhydrase shows no significant sequence similarity with other carbonic anhydrases and may represent an example of convergent evolution at the molecular level.

Carbonic anhydrases (CA)[1] are ubiquitous enzymes that catalyze the interconversion of carbon dioxide and bicarbonate, an important reaction in a number of physiological processes including photosynthesis and respiration (*1*).

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$$

CA was first identified in red blood cells over 65 years ago (*2*, *3*) and has since been shown to exist in a wide range of isoforms in animals, plants, and bacteria, including cyanobacteria and a variety of eukaryotic algae from freshwater and marine environments. The most notable common property among all well-characterized CA is the presence of a Zn^{2+} ion, which is essential for catalytic activity. By far the most studied and best understood of the carbonic anhydrases are the mammalian or α-CA (*1*). In general, these are characterized by molecular masses ranging from 26 to 30 kDa, and they share several highly conserved sequence elements. X-ray crystallography has established that the zinc at the active site is coordinated by three histidine ligands and one catalytically relevant, ionizable water molecule (*1*) (Figure 1).

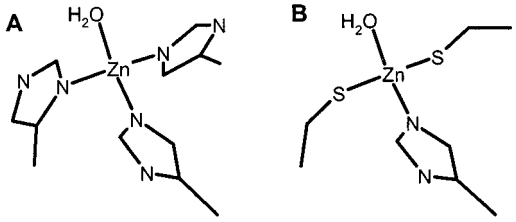

FIGURE 1: Schematic representations of the active site structures of α-CA (A) and *β*-CA (B).

The majority of the plant CA isoforms make up a second class, referred to as the *β*-CA. The *β*-CA exhibit no sequence homology to α-CA and tend to be high molecular mass proteins (140−250 kDa), often composed of multiple subunits of around 30 kDa. Recent extended X-ray absorption fine structure (EXAFS) analyses of the spinach *β*-CA (*4*, *5*) and X-ray crystallography of pea *β*-CA *(6)* indicate that the Zn^{2+} ion coordination includes two cysteine residues and one histidine residue, in addition to the active water molecule, as shown in Figure 1. A closely related coordination is observed in the *β*-CA of the red alga *Porphyridium purpureum* (*7*). The prototype for the third (*γ*) class of CA enzymes has been discovered in the methanoarchaeon *Methanosarcina thermophila* in which the zinc coordination resembles that of α-CA (*8*, *9*).

Algal CA has been most extensively studied in the freshwater species *Chlamydomonas reinhardtii*, where a multimeric CA composed of subunits more closely resembling α-CA has been identified (*10*, *11*). *C. reinhardtii* CA exhibits sequence homology to mammalian CA (*10*, *11*), in particular with respect to the putative Zn^{2+}-binding histidine residues as well as residues forming the H-bonding network to Zn^{2+}-bound solvent molecules in α-CA. Recently, a 27 kDa zinc metalloprotein exhibiting CA activity was identified in the marine diatom *Thalassiosira weissflogii* (TWCA1) (*12*). Its primary sequence reveals <10% homology to the previously

† E.H.C. was supported by a Camille and Henry Dreyfus Postdoctoral Fellowship in Environmental Sciences. Research at Princeton University is supported through funding from the Center for Environmental Bioinorganic Chemistry (CEBIC), which is supported by the NSF and DOE. SSRL is funded by the Department of Energy, Office of Basic Energy Sciences, under Contract DE-AC03-76SF00515. The SSRL Structural Molecular Biology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and by the Department of Energy, Office of Biological and Environmental Research.

* Corresponding authors. F.M.M.M.: phone, (609) 258-2416; fax, (609) 258-5242; e-mail, morel@geo.princeton.edu. G.N.G.: phone, (650) 926-4604; fax, (650) 926-4100; e-mail, george@slac.stanford.edu.

‡ Department of Geosciences, Princeton University.

§ Department of Chemistry, Princeton University.

|| ExxonMobil Research and Engineering Co.

⊥ Stanford Synchrotron Radiation Laboratory.

[1] Abbreviations: CA, carbonic anhydrase; EXAFS, extended X-ray absorption fine structure; TWCA1, Zn-containing carbonic anhydrase from *Thalassiosira weissflogii*; SSRL, Stanford Synchrotron Radiation Laboratory.

10.1021/bi001416s CCC: $19.00
Published on Web 09/13/2000

identified classes of CA and also indicates that the conserved regions defining the α, *β*, and *γ* forms of CA are not found in TWCA1. Specifically, the HxH motif common to the α-CA and the HxxC motif common to the *β*-CA, which are highly conserved and shown through crystal structure determination to be involved in zinc coordination, are not found in TWCA1. The lack of these highly conserved zinc-binding motifs and the presence of several potential metal-binding amino acid residues, including a number of His and Cys residues, lead to uncertainty in predictions of the zinc coordination environment. We describe herein an X-ray absorption spectroscopic investigation of the zinc site of TWCA1.

MATERIALS AND METHODS

T. weissflogii carbonic anhydrase 1 was purified as previously described (*12*). Carbonic anhydrase activity was detected using a variation of the bromcresol purple assay described by Patterson et al. (*13*) utilizing bromothymol blue (*12*). In brief, solutions containing CA were run on 12% nondenaturing polyacrylamide gel. The gel was stained with a solution of running buffer containing 0.1% bromothymol blue, a pH indicator. Bands of CA activity were visualized by blowing a stream of saturated CO_2 gas over the gel. Conversion of CO_2 to HCO_3^- produces H^+, resulting in a localized drop in pH in the area of the gel around CA and producing a yellow band against a blue background. Bovine CA was run as a control.

Samples for X-ray absorption spectroscopy were prepared at a final concentration of approximately 0.2 mM Zn in 20 mM phosphate buffer at pH 6.8. Bovine carbonic anhydrase II was obtained from the Sigma Chemical Co. and used without further purification. Data acquisition was carried out on SSRL beamline 7-3 as previously described (*14*), with samples at 10 K. The energy was calibrated with reference to the lowest energy inflection point of zinc foil, which was assumed to be 9660.7 eV. Sixteen 30-min scans were accumulated for each sample. The extended X-ray absorption fine structure (EXAFS) oscillations $\chi(k)$ were quantitatively analyzed using the EXAFSPAK suite of computer programs,[2] employing ab initio theoretical phase and amplitude functions generated with the program FEFF version 8.03 (*15*, *16*). No smoothing or related manipulation was performed upon any of the data.

RESULTS AND DISCUSSION

The EXAFS Fourier transforms of TWCA1, bovine CA, and spinach CA (*4*) are compared in Figure 2. The bovine and *T. weissflogii* enzymes show considerable similarity, with the EXAFS being dominated by the first shell Zn−N interaction from three histidine ligands, plus an oxygen ligand from $-OH_2$ or $-OH$. Outer shell features are also visible and are attributable to the EXAFS (including multiple scattering) from the distant carbon and nitrogen atoms of the histidines. The spinach enzyme is distinctly different from the other two, with contributions from both nitrogen and sulfur donors. A detailed EXAFS curve-fitting analysis of TWCA1, including full multiple-scattering treatment of the outer shells, is shown in Figure 3. Our analysis employed a

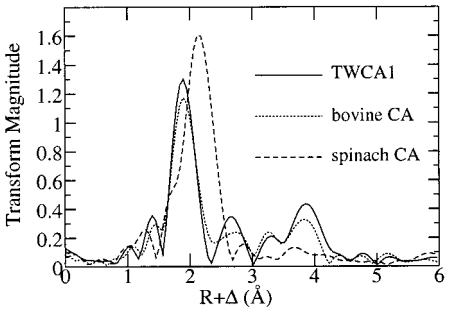

FIGURE 2: Comparison of the EXAFS Fourier transforms of TWCA1, bovine CA, and spinach CA. The data were transformed between $k = 1.0$ and 12.0 $Å^{-1}$ and phase-corrected for Zn−N backscattering. The EXAFS data for spinach CA are those reported by Bracey et al. (*4*).

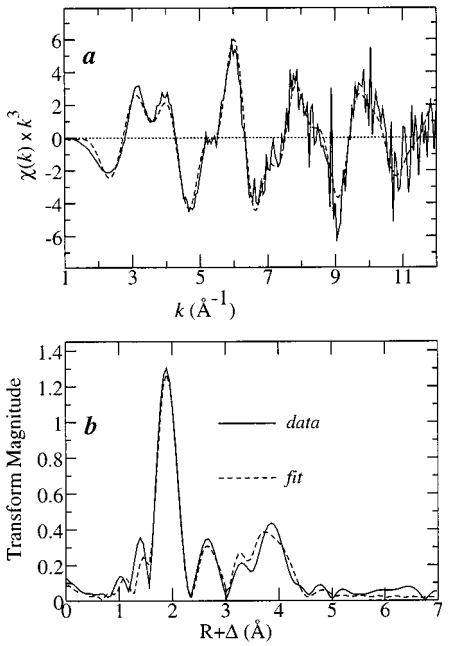

FIGURE 3: EXAFS curve-fitting analysis of TWCA1, showing EXAFS oscillations (a) and Zn−N phase-corrected EXAFS Fourier transforms (b). In both cases the solid line indicates the experimental data and the broken line represents the best fit, which was obtained as described in the text.

model based on the mammalian active site structure, with histidine nitrogen and water oxygen donors. It assumes that the imidazole of the histidine can be approximated by a regular five-membered ring with coordinates derived by averaging those of zinc tetraimidazole (*17*) using methods similar to those employed by Poiarkova and Rehr (*18*). The overall best fit for TWCA1 was obtained with zinc coordinated by three histidines plus a single oxygen, which is an identical coordination environment to that of mammalian CA. When the total coordination number was constrained to be four, and the imidazole and oxygen coordination numbers were refined, the best fit was obtained with 1.1 ± 0.2 oxygen and 2.9 ± 0.2 imidazole ligands.

We note that EXAFS cannot readily discriminate between backscatterers of similar atomic number. Thus, oxygen and nitrogen are difficult to discriminate, but sulfur and nitrogen, for example, are easily distinguished. In the present case, the outer shell EXAFS from Zn(His), visible in the Fourier transform of Figure 3 between 2.7 and 2.9 Å, allow us to determine the coordination numbers of these ligands. Curve-

[2] http://www-ssrl.slac.stanford.edu/exafspak.html.

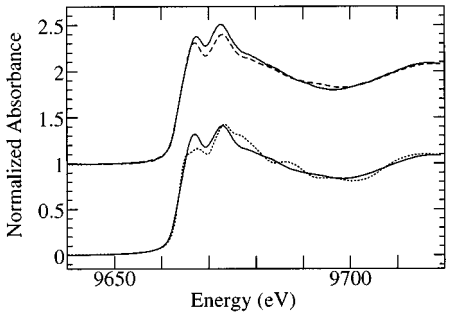

FIGURE 4: Comparison of the Zn K near-edge spectra of TWCA1 (solid lines), bovine CA (broken line), and $[Zn(imidazole)_4](NO_3)_2$ (dotted line). The upper pair of traces compares TWCA1 and bovine CA, while the lower pair compares bovine CA and $[Zn(imidazole)_4](NO_3)_2$, showing the sensitivity of the near-edge spectrum to the structure.

fitting analysis of the data of Figure 3 indicated that Zn−O and Zn−N distances were within the resolution of the data, which is about 0.13 Å ($\sim\pi/2k$). For the final fit these bond lengths were therefore not refined independently, and the best fit was obtained with one Zn−O at 1.975(4) Å, $\sigma^2 = 0.0045(7)$ Å^2, and three Zn−N(His) at the same distance and σ^2. When the Zn−O and Zn−N(His) distances were refined individually, the fit was improved by less than 1%, which was judged to be insignificant (*19*), and the Zn−O and Zn−N bond lengths differed by only 0.04 Å. The σ^2 values are the mean-square displacements, and the values in parentheses are the estimated standard deviations (precisions) obtained from the diagonal elements of the covariance matrix. We note that the precisions will always be somewhat smaller than the accuracies, which are expected to be in the range ±0.02 Å for bond lengths and ±25% for Debye−Waller factors (*19*).

Although difficult to interpret quantitatively, near-edge spectra are very sensitive to the geometry and the nature of the ligands. Thus, they can be very useful as "fingerprints" of particular coordination environments. For example, the Zn K near-edge spectra of mammalian (bovine) CA and $[Zn(imidazole)_4]^{2-}$ are very different (Figure 4), even though the sites differ only in orientation of imidazole ligands and in substitution of a single water molecule for an imidazole in the enzyme. The near identity of the Zn K near-edge spectra of bovine α-CA and TWCA1 (Figure 4) argues for a very similar active site structure in the two enzymes, despite their very different sources.

TWCA1 is the only diatom CA that has been sequenced, and its relationship to other known CA enzymes is unclear. A sequence comparison between TWCA1 and the large number of known CA enzymes, from diverse taxa, indicates that the conserved regions defining the α, β, and γ forms of CA are not found in the same relative positions in TWCA1 (*12*). However, PROPSEARCH, a protein database search based on protein properties (*20*), using the full 34 kDa-derived peptide sequence of TWCA1 suggests some similarities with CAH1 of *Chlamydomonas* (in particular, a HxH motif). However, the similarity is in the amino-terminal region which is cleaved in vivo to form the active 27 kDa enzyme (*12*). Since there is no evidence that this fragment reassociates with the larger fragment, final conclusions as to the relationship of the diatom CA enzyme to other known forms must await sequence data from more than one example. In any case, the relationship to all other CA enzymes is clearly a distant one. Nevertheless, the similarity of the TWCA1 and mammalian CA spectra (Figure 4) supports our conclusion from the EXAFS that the ligands are the same and, furthermore, indicates that the organization and geometry of the active sites are almost identical. The carbonic anhydrases thus seem to present another striking example of convergent evolution at the molecular level (e.g., ref *21*).

ACKNOWLEDGMENT

We are indebted to Martin J. George of SSRL for use of his data collection software. We thank Drs. Jason Christiansen and Stephen P. Cramer for generously allowing us access to the raw EXAFS data of spinach carbonic anhydrase.

REFERENCES

1. Lindskog, S. (1997) *Pharmacol. Ther. 74*, 1−20.
2. Meldrum, N. U., and Roughton, F. J. W. (1933) *J. Physiol. 80*, 113−142.
3. Stadie, W. C., and O'Brien, H. (1933) *J. Biol. Chem. 103*, 521−529.
4. Bracey, M. H., Christiansen, J., Tovar, P., Cramer, S. P., and Bartlett, S. G. (1994) *Biochemistry 33*, 13126−13131.
5. Rowlett, R. S., Chance, M. R., Wirt, M. D., Sidelinger, D. E., Royal, J. R., Woodroffe, M., Wang, Y.-F. A., Saha, R. P., and Lam, M. G. (1994) *Biochemistry 33*, 13967−13976.
6. Kimber, M. S., and Pai, E. F. (2000) *EMBO J. 19*, 1407−1418.
7. Mitsuhashi, S., Mizushima, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Moriyama, H., Ueki, T., Miyachi, S., and Tsukihara, T. (2000) *J. Biol. Chem. 275*, 5521−5526.
8. Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996) *EMBO J. 15*, 2323−2330.
9. Iverson, T. M., Alber, B. E., Kisker, C., Ferry, J. G., and Rees, D. C. (2000) *Biochemistry* (in press).
10. Fukuzawa, H., Fujiwara, S., Yamamoto, Y., Dionisio-Sese, M. L., and Miyachi, S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4383−4387.
11. Fukuzawa, H., Fujiwara, S., Tachiki, A., and Miyachi, S. (1990) *Nucleic Acids Res. 18*, 6441−6442.
12. Roberts, S. B., Lane, T. W., and Morel, F. M. M. (1997) *J. Phycol. 33*, 845−850.
13. Patterson, B. D., Atkins, C. A., Graham, D., and Wills, R. B. H. (1971) *Anal. Biochem. 44*, 388−391.
14. George, G. N., Garrett, R. M., Graf, T., Prince, R. C., and Rajagopalan, K. V. (1998) *J. Am. Chem. Soc. 120*, 4522−4523.
15. Rehr, J. J., Mustre de Leon, J., Zabinsky, S. I., and Albers, R. C. (1991) *J. Am. Chem. Soc. 113*, 5135−5140.
16. Mustre de Leon, J., Rehr, J. J., Zabinsky, S. I., and Albers, R. C. (1991) *Phys. Rev. B44*, 4146−4156.
17. Bear, C. A., Duggan, K. A., and Freeman, H. C. (1975) *Acta Crystallogr., Sect. B 31*, 2713−2715.
18. Poiarkova, A. V., and Rehr, J. J. (1999) *Phys. Rev. B 59*, 948−957.
19. George, G. N. (1997) *J. Biol. Inorg. Chem. 2*, 790−796.
20. Hobohm, U., and Sander, C. (1995) *J. Mol. Biol. 251*, 390−399.
21. Makarova, K. S., and Grishin, N. V. (1999) *Protein Sci. 8*, 2537−2540.

BI001416S