Carbonic Anhydrase in the Marine Diatom *Thalassiosira weissflogii*

by

Samantha B. Roberts

B.S. , Biological Sciences
Stanford University, 1992

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of

Master of Science in
Civil and Environmental Engineering

at the

Massachusetts Institute of Technology
August 1995

© 1995 Massachusetts Institute of Technology
All rights reserved

Signature of Author...
Department of Civil and Environmental Engineering

Certified by...
Professor François M. M. Morel
Thesis Advisor

Accepted by...
Professor Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 25 1995
Barker Eng
LIBRARIES

Carbonic Anhydrase in the Marine Diatom *Thalassiosira weissflogii*

by

Samantha B. Roberts

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of

Master of Science in
Civil and Environmental Engineering

ABSTRACT

Oceanic phytoplankton play an important role in the global carbon cycle. Some species of microalgae apparently use the enzyme carbonic anhydrase to accumulate inorganic carbon. This enzyme has been found in the marine diatom *Thalassiosira weissflogii*. Zinc, an essential cofactor of carbonic anhydrase, has been shown to limit *T. weissflogii* growth at low carbon dioxide concentrations. The regulation of carbonic anhydrase activity by CO_2 concentration implies that the enzyme is important for the acquisition of inorganic carbon (Morel et al. 1994). In order to gain insight into the mechanism of this carbon-zinc co-limitation, the carbonic anhydrase protein was examined in greater detail. The enzyme was purified and sequenced, and the corresponding gene cloned. The carbonic anhydrase gene sequence was different from other algal carbonic anhydrase genes, and encoded a protein of roughly 32 kilodaltons. The amino terminal amino acids sequenced from the purified *T. weissflogii* carbonic anhydrase are 72 residues downstream of the putative starting methionine predicted by the CA44 cDNA. This difference may be due to the presence of a short-lived signal sequence designed to guide the enzyme to the correct cellular location. This work opens the door for additional experiments to examine the mechanism of *T. weissflogii* inorganic carbon acquisition.

Thesis Supervisor: Dr. Francois M. M. Morel
Title: Professor of Civil and Environmental Engineering

Table of Contents

Acknowledgements 5

1. Introduction

1.1 The Zinc Hypothesis 6

1.2 The Biological Pump 6

1.3 The Redfield Ratio and HNLC Regions 8

1.4 The Rain Ratio 9

1.5 The Iron Hypothesis 9

1.6 The Zinc Hypothesis Extended 10

1.7 Carbonic Anhydrase

1.7 a Carbonic Anhydrase in Microalgae 10

1.7 b *T. weissflogii* Carbonic Anhydrase 12

2. Materials and Methods

2.1 *T. weissflogii* Growth Conditions 13

2.2 *T. weissflogii* Harvest and Protein Extraction 13

2.3 Bromcresol Purple Assay

2.3 a Materials 13

2.3 b The Assay 14

2.4 Carbonic Anhydrase/ CO_2 Experiment 14

2.5 Protein Purification

2.5 a Materials 14

2.5 b Affinity Resin Preparation 15

2.5 c Purification Protocol 15

2.6 Protein Purification

2.6 a Materials 16

2.6 b Concentration of Purified Protein 16

2.7 Protein Sequencing

2.7 a N-terminal Protein Sequencing 16

2.7 b Internal Protein Sequencing 17

2.8 Generation of *T. weissflogii* cDNA Library

2.8 a Total RNA Extraction 17

2.8 b mRNA Extraction 18

2.8 c cDNA Synthesis 18

2.8 d Ligation into the Lambda Zap II Vector 18

2.8 e Packaging and Library Titration 19

2.9 PCR on Library and Probe Generation
- 2.9 a Obtaining DNA from Amplified Library 19
- 2.9 b PCR on Library DNA 19
- 2.9 c Search for Sequence Upstream of the CAN2 Primer Region 20
- 2.9 d Generation of a Probe for Screening the Library 21

2.10 Probing the *T. weissflogii* cDNA library
- 2.10 a Solutions Used 21
- 2.10 b Probe Labeling 22
- 2.10 c Filter Lifts 22
- 2.10 d Probing the Filters 22
- 2.10 e PCR Screening to Identify Full Length Clones 23
- 2.10 f Isolation and Retesting of Full Length Clones 23
- 2.10 g Sequencing 24

3. Results
- 3.1 Carbonic Anhydrase/ CO_2 Experiment 25
- 3.2 Protein Purification 25
- 3.3 N-terminal Protein Sequence 26
- 3.4 Internal Protein Sequence 26
- 3.5 Total RNA Extraction 27
- 3.6 The Library 27
- 3.7 PCR Using Degenerate Primers from Protein Sequence 27
- 3.8 PCR Product from Nested Reaction 27
- 3.9 Library Screening 28
- 3.10 Clone CA44 Vital Statistics 28

4. Conclusions 32

Appendices
- One: Restriction Map of Clone CA44 34
- Two: Summary of Restriction Sites 39
- Three: Sequences of Primers and Locations in the Gene 41
- Four: Kits Used in Materials and Methods 42

References 43

Acknowledgements

I would like to thank François Morel and the Morel lab for all of their help and support. In particular, John Reinfelder provided valuable advice on aquatic chemistry, algal culture and experimental design. Jennifer Lee, Beth Ahner, Judy Sheldon and Jennifer Jay were cheerful and willing sources of information, encouragement, and friendship. Finally, I could not have even begun this project without the advice and encouragement of my husband, Radclyffe Roberts.

1 Introduction

1.1 The Zinc Hypothesis

Microalgae use a variety of mechanisms for the accumulation of inorganic carbon from the waters in which they live. Some species apparently use the enzyme carbonic anhydrase to concentrate dissolved inorganic carbon. Algal carbonic anhydrase has been most extensively studied in freshwater species, but has also been found in marine microalgae such as *Dunaliella tertiolecta* (Aizawa and Miyachi 1984) and the diatoms *Phaeodactylum tricornutum* (Patel and Merrett 1986) and *Thalassiosira weissflogii* (Morel at al. 1994). Since zinc is an essential cofactor of carbonic anhydrase, these diatoms, when growing under conditions of low CO_2, require more zinc than cells living in CO_2 sufficient conditions. However, many areas of the ocean have extremely low zinc concentrations. For example, in the north Pacific surface waters the zinc concentration is about 2 picomolar (Bruland 1989). Similar to Martin's (1988) theory regarding iron limitation of primary production in high nutrient-low chlorophyll regions of the ocean, the zinc hypothesis states that surface water zinc levels can, through modulation of carbonic anhydrase activity, affect phytoplankton growth, species composition, and CO_2 consumption in many areas of the ocean.

1.2 The Biological Pump

The carbon cycle in the ocean is largely driven by the biological pump, which involves the complementary reactions of photosynthesis and respiration. Photosynthesis is the creation of biomass from CO_2 and H_2O through the reaction $CO_2 + H_2O \rightarrow (CH_2O)_n + O_2$. Respiration is the subsequent breakdown of this biomass by the reaction $(CH_2O)_n + O_2 \rightarrow CO_2 + H_2O$. Algae are the primary producers, and the biomass they create by photosynthesis is utilized up the food chain by grazers and other non-photosynthetic organisms. Primary productivity in the ocean is thought to be limited by nitrate or phosphate. Once one of these nutrients is exhausted, no more biomass can be produced until it is recycled during respiration. Thus respiration and photosynthesis are interdependent.

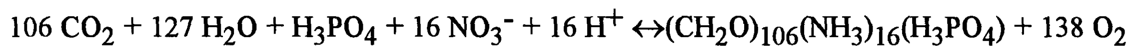
The biological pump has two loops. One is the alternating cycle of photosynthesis and respiration that occurs in the surface waters of the ocean. The second is the cycling of organic carbon between the surface waters and the deep ocean. Some of the biomass drops out of the surface water loop and enters the deep sea loop through particles sinking and the downwelling of dissolved organic matter. This amount is estimated to be about 20% of the surface biomass (Sarmiento 1993), although it can vary greatly. The organic matter that sinks to the deep ocean is eventually respired back into its original components (mineralization). This causes the deep waters to become enriched in CO_2. During upwelling, dissolved nutrients and CO_2 are brought back to the surface waters and re-enter the surface pool, completing the second loop of the biological pump.

Although upwelled seawater is supersaturated with CO_2 with respect to the atmosphere, the abundant nutrients that this water contains create a fertile area for the algae. If an algal bloom occurs in an area of upwelling, it can consume much or all of the CO_2 (until the nutrients are depleted) and therefore reduce the amount of CO_2 that is released to the atmosphere at that location. Thus, the biota can play a significant role in controlling whether a region of the ocean is a local source or a sink of atmospheric CO_2. Siegenthaler and Sarmiento (1993) estimated that if the biological pump were fully effective, and surface nitrate and phosphate levels were drawn down to zero everywhere, the atmospheric CO_2 level would drop from 280 ppm (pre-industrial value) to approximately 160 ppm, whereas if the biological pump were turned off the level would rise to 450 ppm.

The effectiveness of the biological pump is regulated to a large extent by phytoplankton physiology and ecology. In particular, it is controlled by the efficiency with which phytoplankton communities can fix carbon. This efficiency is manifested in the ratio of carbon to nitrogen or phosphorus, or the ratio of inorganic carbon to organic carbon in sinking particulate matter and may be controlled by the abundance of iron or zinc.

1.3 The Redfield Ratio and HNLC Regions

The Redfield Ratio describes the relationship between nitrogen, phosphorus, and carbon in organic matter. This ratio was defined by the observation that the ratio of carbon to nitrogen to phosphorus is the same in seawater as it is in living organisms: 106 C: 16 N: 1 P (Redfield 1963). The Redfeld Ratio can be combined with the photosynthesis equation to write the following equation for the formation of biomass:



$$106\ CO_2 + 127\ H_2O + H_3PO_4 + 16\ NO_3^- + 16\ H^+ \leftrightarrow (CH_2O)_{106}(NH_3)_{16}(H_3PO_4) + 138\ O_2$$

Thus, the nitrate and phosphate present in a parcel of water can be used to calculate the maximum amount of biomass that could be produced in that parcel. In most areas of the ocean, nitrate and phosphate are depleted in the surface waters, indicating that the biological pump is operating at its maximum level. There are, however, three well known exceptions. The Southern Ocean, the equatorial Pacific, and the northern Pacific have been classified as high nutrient-low chlorophyll (HNLC) regions (Murray et al. 1994). Although there is a large amount of biomass present in these regions, the nitrate and phosphate are high enough to support additional biomass. It is not known exactly why this occurs, although several theories have been proposed. One of these theories, the iron hypothesis, will be discussed in section 1.5.

The Redfield Ratio has been taken as constant in almost all ocean models. Slight variations in this ratio, however, could affect the biological pump. If some phytoplankton were able to fix carbon more efficiently than others and therefore use less nitrate per CO_2 fixed, then more CO_2 could be consumed before the limitation imposed by nitrate depletion was reached. Sambrotto et al. (1993) have provided some evidence for this in the ocean. They measured the carbon and nitrogen content of sinking particulate matter in the north Atlantic during an algal bloom, and estimated that using the Redfield Ratio underestimates the actual primary productivity by as much as fifty percent. This disparity could be explained by zinc-carbon colimitation of some species but not others.

1.4 The Rain Ratio

Coccolithophores are a large group of algae which produce a calcium carbonate shell by the following reaction: $Ca^{+2} + 2HCO_3^- \rightarrow CaCO_3 + H_2O + CO_2$ (Nimer and Merrett 1992). A portion of this calcium carbonate sinks to the deep ocean along with the sinking particulate organic matter in the deep loop of the biological pump. The Rain Ratio is defined as the ratio of calcium carbonate to organic carbon in particulate material falling from the surface waters to the deep ocean (Berger and Keir 1984). The ratio depends upon the relative abundance of calcifying phytoplankton in the surface waters, their growth rates, and the exact proportions of carbonate and organic carbon produced by each species present. Because calcium carbonate production decreases alkalinity, it actually causes a concurrent production of CO_2 to the water and potentially to the atmosphere (Sundquist 1993). Therefore, the percentage of a local phytoplankton population that precipitates calcium carbonate (the total biomass is still limited by nitrate and phosphate) can affect how much CO_2 is released or consumed in that particular area.

1.5 The Iron Hypothesis

One way to increase the net productivity of the biological pump is fertilization of the HNLC regions. Although nitrate and phosphate are traditionally considered to be limiting, in these regions they are still plentiful and therefore not limiting growth. Martin et al. (1988) suggested that iron was the element limiting productivity in the Southern Ocean. Large organisms, with a smaller surface area to volume ratio, would be expected to be more limited than smaller organisms (Morel et al. 1991). Although addition of iron to sea water samples caused an enhancement in biomass and growth rates (Martin et al. 1988), it is difficult to extrapolate these bottle experiments to what occurs *in situ*. The algae in this area are probably iron limited, but it is uncertain whether this has a large effect on primary productivity. Phenomena such as grazing and iron complexation might limit the total biomass level that could be reached. This hypothesis was tested in the equatorial Pacific Ocean, approximately 500 kilometers south of the Galapagos Islands. Nearly 7,800 moles of iron were added to an ocean patch approximately 8 km by 8 km. The patch was then monitored for increases in algae growth, iron uptake, and many

other factors. The results of this first experiment were not clear cut, however (Watson et al. 1994, Martin et al. 1994), so a second experiment was performed. It has not yet been analyzed, but appears to have resulted in a much larger algal bloom than was observed in the first experiment.

1.6 The Zinc Hypothesis Extended

Similar to the theory that iron limits primary production is the theory of zinc limitation. Experiments on the marine diatom *Thalassiosira weissflogii* show that if there is sufficient available zinc (15 picomolar) in the culture medium, CO_2 stress caused by diffusion limitation is relieved through the action of carbonic anhydrase (Morel et al. 1994). Coccolithophores, which either lack carbonic anhydrase, as in the case of *Emiliana huxleyi*, or have very little, as in the case of *Pleurochrysis carterae* (Sikes and Wheeler 1982) do not require carbonic anhydrase to acquire inorganic carbon. Therefore, in areas of the ocean with low zinc concentrations, at low CO_2, growth of a population of coccolithophores might be favored over growth of a population of diatoms or a mixture of the two. If there is abundant zinc, it is unclear which organisms will be favored. The consequence of changing this species composition is a change in the amount of CO_2 drawn down per nutrients used. If coccolithophores predominate, less CO_2 will be drawn into the biological pump than the amount a mixed population will draw down, which in turn will be less than the amount a population composed entirely of diatoms will sequester.

1.7 Carbonic Anhydrase

1.7 a Carbonic Anhydrase in Microalgae

Carbonic anhydrase, which requires one or more zinc ions for function, reversibly catalyzes the following reaction: $HCO_3^- + H^+ \leftrightarrow CO_2 + H_2O$. In this way, the cell can use the bicarbonate ion to obtain the required amount of CO_2. In plants, this enzyme functions to supply the enzyme ribulose bisphosphate carboxylase-oxygenase (RUBISCO) with CO_2 during photosynthesis. RUBISCO then fixes the CO_2 into 3-phosphoglyceric acid (PGA). In some algae, however, carbonic anhydrase has been

found to have an additional role in inorganic carbon acquisition. Two freshwater species, *Chlamydomonas reinhardtii* and *Chlorella saccharophila*, have been shown to have extracellular as well as intracellular carbonic anhydrase isoforms (Husic et al. 1989, Williams and Coleman 1993).

Among algal carbonic anhydrases, the most work has been done on the green alga *Chlamydomonas reinhardtii*. In addition to a chloroplastic carbonic anhydrase isoform, two periplasmic isoforms have been found (Husic et al. 1989). The first periplasmic form, the product of the gene CAH1, is regulated at the transcriptional level (Ishida et al. 1993). The large and small subunits are translated as a single protein which is then cleaved. The functional enzyme is a heterotetramer composed of two 37 kD subunits bound to two 4 kD subunits by disulfide bonds. Two zinc atoms are present per holoenzyme (Kamo et al. 1990). Activity of this isoform is induced by low CO_2 levels and repressed by high CO_2 levels.

The second *Chlamydomonas* periplasmic carbonic anhydrase is the product of the gene CAH2, which is 91.2 % identical to the CAH1 gene sequence (Rawat and Moroney 1991). The protein consists of two 39 kD and two 4.5 kD subunits bound together in the same manner as the other isoform. This isoform is transcribed at high CO_2 levels, but not at low CO_2 levels (Ishida et al. 1993).

Chlorella saccharophila also has two distinct isoforms of carbonic anhydrase (Williams and Coleman 1993). The periplasmic isoform cross-reacts with antibodies raised against the *Chlamydomonas* periplasmic carbonic anhydrases. This isoform is present at low CO_2 and absent at high CO_2. The chloroplastic isoform does not cross react with the *Chlamydomonas* periplasmic antibody, and is present at a roughly constant level regardless of CO_2 concentration.

In both cases, the multiple isoforms of carbonic anhydrase are proposed to function as parts of a carbon uptake and concentration system. This would involve carbon uptake at the external cell surface in the form of bicarbonate, transport across the cell membrane, and provision as CO_2 to RUBISCO in the chloroplast. The details of exactly where and how carbonic anhydrase is involved in this system have not been worked out yet. There are at least two possible scenarios. Carbonic anhydrase may function as a transporter, facilitating uptake of bicarbonate, and either concurrent

conversion to CO_2 or conversion to CO_2 at a later point by an additional carbonic anhydrase. Another possibility is that bicarbonate is converted to CO_2 at the cell surface by an extracellular or periplasmic carbonic anhydrase. This creates a local area of high CO_2 which then diffuses into the cell. The goal of the work in *T. weissflogii* is to determine which mechanism is operating and how it is regulated.

1.7 b *T. weissflogii* Carbonic Anhydrase

Preliminary experiments in *T. weissflogii* have shown that there is at least one isoform of carbonic anhydrase present, and that there may be more. Most of this diatom's cellular zinc appears to be present in carbonic anhydrase. Zinc limited growth curves show faster growth at high CO_2 than at low CO_2, and bromcresol purple assay experiments show a direct correlation at low CO_2 between the amount of zinc present and carbonic anhydrase activity (Morel et al. 1994). The experiments described in this thesis were designed to test the presence of carbonic anhydrase activity at different CO_2 concentrations. It was determined that more carbonic anhydrase was present in cells grown at low CO_2 than in cells grown at high CO_2. Additionally, carbonic anhydrase activity was activated or inactivated within eighteen hours of exposure to a new gas level.

In order to learn more about the properties of the *T. weissflogii* carbonic anhydrase, the protein was purified and the gene cloned. Sequence analysis revealed a predicted protein size of 32 kD, longer than expected from the size of the purified protein. This size disparity might be due to the presence of signal sequences designed to guide the protein to the correct location in the cell. The protein sequence did not resemble published sequences of other algal carbonic anhydrase genes. This work provides a foundation for many more experiments that will help elucidate the mechanism of inorganic carbon uptake in *T. weissflogii*.

2. Materials and Methods

2.1 *T. weissflogii* Growth Conditions

T. weissflogii (clone Actin) cultures were grown in modified (lacking cobalt) Aquil growth medium following the recipe in Price et al. (1988). The final metal concentrations in media with 100 μM EDTA were: Fe = 8.32×10^{-6} M, Cu = 1.96×10^{-8} M, Mn = 1.21×10^{-7} M, Zn = 7.97×10^{-8} M, and Co = 0. Cell concentrations were followed using a Coulter Channelyzer 256. Gas mixes were prepared in a standard air N_2/ O_2 background with the designated concentration of CO_2.

2.2 *T. weissflogii* Harvest and Protein Extraction

Cells were harvested during late stationary phase, at densities of 90,000 to 100,000 cells per ml. They were then filtered through a 0.45 micron polycarbonate filter, resuspended in approximately 10 ml filtered sea water, and pelleted by centrifugation for 10 minutes at 800 x g in a Beckman TJ-6 swinging bucket rotor. The resultant pellet was resuspended in an appropriate volume of sea water (500 μl to 1.5 ml) and sonicated using a Branson Sonifier 250 equipped with a microtip. Sonication was performed at 80% duty cycle, output five, for 45 seconds on ice. Unbroken cells and cellular debris were pelleted in an Eppendorf model 5415C microfuge at 13,800 x g for 10 minutes at 4 ° C. The supernatant was then removed to a fresh tube.

2.3 Bromcresol Purple Assay

2.3 a Materials

10 X Running Buffer (1L): 15.1 g tris base, 94 g glycine
Bromcresol purple solution: 0.1 % bromcresol purple in 1 X running buffer

2.3 b The Assay

Carbonic anhydrase was detected using a modification of the method of Patterson et al. (1971). Equal amounts of protein were run on a 10% or 12% non-dentauring polyacrylamide gel using the Laemmli method (Sambrook et al. 1989). The gel was soaked in bromcresol purple solution and blotted dry with kimwipes. When saturated CO_2 gas was blown over the gel, red or yellow carbonic anhydrase bands appeared against a purple gel background. The gel was then frozen on dry ice and photographed under longwave ultraviolet light using a Wratten 74 green filter and Polaroid 667 black and white film.

2.4 Carbonic Anhydrase/ CO_2 Experiment

1 L *T. weissflogii* cultures were grown in Aquil media lacking cobalt (see section 2.1). The cultures were bubbled with either 100, 300, or 1000 ppm CO_2, until 27920, 58192, and 34823 cells per ml respectively. For the final 18 hours, the cultures were split into two 500 ml cultures and 8 μCi ^{65}Zn was added to each. One culture remained at the same gas concentration, and one was switched in the following scheme: bottle a: 1000 ppm→1000 ppm, bottle b: 1000 pm→300 ppm, bottle c: 300 ppm→300 ppm, bottle d: 300 ppm→100 ppm, bottle e: 100 ppm→100 ppm, and bottle f: 100 ppm→300 ppm. Cells were harvested at bottle a = 141,000 cells per ml, bottle b = 96,300 cells per ml, bottle c = 168,700 cells per ml, bottle d = 164,000 cells per ml, bottle e = 61,800 cells per ml, and bottle f = 46,800 cells per ml. Harvested cells were resuspended in sterile sea water volumes which resulted in samples of equal concentrations cells per volume. The bromcresol purple assay was performed as already described, and the gel was dried and exposed to film for two weeks.

2.5 Protein Purification

2.5 a Materials

Carbonate Buffer: 0.1 M Na_2CO_3, 20% v/v dioxane, pH 11
Dialysis Buffer: 100 mM NaCl, 1 mM EDTA, 20 mM NaH_2PO_4, pH 6.8

Wash I: 25 mM Tris, 22 mM Na_2SO_4, pH 8.2
Wash II: 25 mM Tris, 300 mM $NaClO_4$, pH 8.7
Elution Buffer: 100 mM NaOAc, 500 mM $NaClO_4$, pH 5.6

2.5 b Affinity Resin Preparation

A carbonic anhydrase affinity resin was prepared according to the protocol of Yang et al. (1985) as follows:

6 g epoxy-activated Sepharose 6B (Pharmacia) was dissolved in milli-Q H_2O, then washed over a scintered glass filter with 1 L milli-Q H_2O followed by 800 ml carbonate buffer. 0.93 g p-aminomethylbenzene sulfonamide in carbonate buffer was added to the sepharose, the volume brought to 80 ml, and the pH was returned to 11. The solution was shaken gently for 40 hours at 50 ° C, then filtered on a scintered glass filter and washed with 300 ml carbonate buffer, 300 ml milli-Q H2O, 300 ml 0.1 M Na2CO3 pH8, and 300 ml 0.1 M NaOAc pH 4. This was procedure was repeated two times, and then the resin was resuspended in 200 ml 1M ethanolamine pH 8 and left at room temperature for 18 hours. Finally, the resin was filtered and washed with 1 L dialysis buffer then resuspended in 80 ml dialysis buffer and stored at 4° C until used.

2.5 c Purification Protocol

The purification protocol was a slight modification of the procedure described by Rawat and Moroney (1991). Four liters of *T. weissflogii* were grown and harvested as explained in sections 2.1 and 2.2. A 35% ammonium sulfate precipitation was performed on the supernatant after centrifugation according to the method of Sambrook et al. (1989). The supernatant was removed and precipitated by 70% ammonium sulfate. The pellet obtained at this stage was resuspended in 5 ml dialysis buffer and dialysed in 15,000 dalton cutoff dialysis tubing for 1 week at 4° C. The sample was spun for 15 minutes at 4° C and 13,800 x g, and the supernatant loaded onto a 1 ml p-aminomethylbenzene sulfonamide affinity resin column at a flow rate of approximately 12 to 15 drips per minute. The column was washed with 10 ml wash I followed by 8 ml wash II, and finally eluted with 2 x 600 μl elution buffer. To examine the results of the

procedure, 30 μl of each step was run on a 10 % denaturing gel according to the method of Sambrook et al. (1989).

2.6 Protein Sequencing

2.6 a Materials

Transfer buffer I: 10 mM CAPS, 20% methanol, pH 11
Transfer buffer II: (4L) 12.5% methanol, 25 mM tris, 0.19 M glycine, pH 8.3
Coomassie stain: 50% methanol, 10% acetic acid, 0.5% coomassie brilliant blue G250
Destain: 10% acetic acid, 10% isopropanol

2.6 b Concentration of Purified Protein

A 14 L culture was purified using the method described in section 2.5. The eluate was concentrated using a Centricon 10,000 mw cutoff concentration column. To block the column, it was filled with a solution of 1 mg per ml bovine serum albumin (Sigma A0281) in elution buffer and spun for 10 minutes at 4,000 x g in a JA-17 rotor. The column was then rinsed by filling the loading reservoir with elution buffer then flicking out the liquid. This was repeated 20 times, then the column was filled with elution buffer and spun for 5 minutes at 4,000 x g in a JA-17 rotor. Next, all remaining elution buffer was flicked out of the loading reservoir and the sample applied. The column was not allowed to dry before the sample was loaded. After the sample was applied, the column was spun for 2 hours at 3,400 x g in the JA-17 rotor. The final volume was approximately 60 μl.

2.7 Protein Sequencing

2.7 a Preparation of Sample for N-Terminal Protein Sequencing

The sample was run on a 12% denaturing polyacrylamide gel, then transferred to PVDF membrane in transfer buffer I for 1 hour at 0.47 amps. The membrane was

stained with fresh coomassie stain and the protein band cut out. N-Terminal protein sequencing was performed by Dr. Paul Matsudaira at the Whitehead Institute for Biomedical Research, Cambridge, MA.

2.7 b Internal Protein Sequencing

Concentrated protein was run on a 12% denaturing polyacrylamide gel and transferred onto a nitrocellulose membrane in transfer buffer II for 1 hour at 0.47 amps. The sample was sequenced by Dr. Richard Cook, MIT Biopolymers Laboratory, Cambridge, MA. Protein was digested by the endoproteinase Lys-C and the resultant fragments analysed using an Applied Biosystems model 477A.

2.8 Generation of *T. weissflogii* cDNA library

Two liters of *T. weissflogii* were grown to a density of 150,000 cells per ml, then bubbled with 300 ppm CO_2 for 1 hour prior to harvesting. Cells were harvested as previously described, total RNA was extracted, mRNA was isolated, and then the mRNA was reverse transcribed into cDNA. Next, EcoRI linkers were ligated onto the cDNA and then cloned into Stratagene's Lambda Zap® II vector. Finally, the library was amplified.

2.8 a Total RNA Extraction

All equipment and solutions used in this section and sections 2.8 b,c were treated with diethyl pyrocarbonate (DEPC) to remove any RNAses that might otherwise degrade the RNA. Two aliquots of 667 ml each were harvested (10^8 cells) and resuspended in 25 ml ice cold milli-Q H_2O. Total RNA was extracted by the guanidine thiocyanate method according to the instructions in Promega's RNAgents™ Total RNA Isolation Kit.

2.8 b mRNA Isolation

Messenger RNA was separated from total RNA using Promega's PolyATract® mRNA Isolation System. This technique utilizes a biotinylated poly-T oligonucleotide probe, which hybridizes to the 3' poly-A tail and then, when reacted with streptavidin coupled to paramagnetic particles, is separated magnetically from the rest of the RNA. The pellet obtained after following the instructions in the kit was resuspended in 100 μl RNAse free water and stored at -70° C.

2.8 c cDNA Synthesis

cDNA was created from the mRNA using Promega's Riboclone® cDNA Synthesis System. 0.5 μg Xba I primer was annealed to 1 μg mRNA and the first strand synthesized using AMV reverse transcriptase according to the kit instructions. The second strand was synthesized using DNA Polymerase I, and the product resuspended in 30 μl TE (also provided in the kit.)

2.8 d Ligation into the Lambda Zap II Vector

First, the cDNA was fractionated by separation on a 1.2 % TAE agarose gel according to Sambrook et al. (1989). All cDNA larger than 450 base pairs was cut from the gel and extracted using the Geneclean®II Kit. This size-fractionated cDNA was resuspended in 5 μl sterile milli-Q H_2O. EcoRI linkers were ligated to 0.2 μg size-fractionated cDNA using Promega's Riboclone® EcoRI Linker Ligation System. This product was next digested with EcoRI (New England Biolabs), and then ethanol precipitated and resuspended in 2.5 μl sterile TE from the Linker Ligation Kit. The following reaction was performed to ligate the cDNA into the Lambda Zap II vector: 2 μl digested cDNA was mixed with 1 μl (1 μg) Lambda Zap II arms, 0.5 μl 10 X Ligase Buffer and 0.5 μl T4 DNA Ligase (New England Biolabs). This reaction was incubated at 16° C for 14 hours.

2.8 e Packaging and Library Titration

The Lambda Zap II vector with the cDNA inserts was packaged into phage particles using Stratagene's Gigapack®II Packaging Extract Kit. Three μl of the ligation reaction described in the previous section were packaged. Next, the library was titered on NZY plates following the instructions in this kit. Finally, the entire packaging reaction was amplified according to directions in this kit and the amplified library re-titered.

2.9 PCR on Library and Probe Generation

2.9 a Obtaining DNA from Amplified Library

Stratagene's Mass Excision Protocol was used to generate library plasmid DNA in the vector pBluescript II sk-. 10^8 plaque forming units (pfu) were excised using this technique. Finally, the DNA was extracted using a Qiagen maxi prep.

2.9 b PCR on Library DNA

PCR primers were designed to the N-terminal and internal carbonic anhydase protein sequence. Degenerate symbols used are according to the IUPAC/IUB nomenclature system.

CANterm: 5' ggI ttY MgI MgI caY catY taY gaY 3'
CAN2: 5' gaR gtI caR gaY ggI ttY 3'
RevMQR: 5' cat Rtc Rtc IcK Ytg cat 3'
ForMQR: 5' atg caR MgI gaY gaY atg 3'
RevDLY: 5' ccN ggI gcR taI aRR tc 3'

PCR was performed using the 1 ng each of the CAN2 and RevDLY primers in a 50 μl reaction volume with 25 ng cDNA, 0.2 mM dNTP's, 1 X Taq Polymerase buffer, 2.5 mM $MgCl_2$, and 5 units Taq DNA Polymerase (Promega). Thirty three cycles were

performed with a denaturing step of 94° C for 30 seconds, annealing at 48° C for 30 seconds, and extension at 72 °C for 45 seconds. The PCR reaction was run on a 2% TBE agarose gel and the band at approximately 500 bp cut out. The DNA was then extracted from agarose using the Geneclean® II Kit, ligated into the TA cloning vector using Invitrogen's TA Cloning® Kit, and plated on LB/ ampicillin/ X-Gal plates according to instructions in the kit. White colonies were picked into 3 ml LB/ ampicillin and grown overnight. Minipreps were performed according to Sambrook et al. (1989), and sequenced to check for homology to the protein sequence. DNA sequencing was performed using Amersham's Sequenase® 2.0 DNA Sequencing Kit with ^{35}S dATP.

2.9 c Searching for Sequence Upstream of the CAN2 Primer Region

To determine if there was more gene sequence upstream of the 5' end of the PCR product, nested PCR was performed using primers specific to the pBluescript vector in conjunction with primers made from DNA sequence obtained in part 2.9 b. The Bluescript-specific primers used were T3 and SK (available from Stratagene). The DNA sequence obtained in section 2.9 b was used to design the following primers:

BestHind: 5' gtg cgt aca agt cat cac tca tgt c 3'
CA4r 5' ctt gcg atc aac ttg cca gg 3'

First, PCR was performed using 1 ng each of T3 and BestHind in a 50 μl reaction. Reaction components were otherwise identical to those described in section 2.9 b. The reaction conditions were: 23 cycles of denaturing at 95° C for 30 seconds, annealing at 55° C for 30 seconds, followed by extension at 72° C for 39 seconds. This reaction was then diluted 1:100, and PCR performed on 1 μl of this dilution using the primers SK and CA4r, with the rest of the components and conditions identical to those of the T3/BestHind reaction.

This PCR reaction was then run on a TBE agarose gel, and a band of approximately 1.1 Kb cut out of the gel. DNA was extracted from the agarose using the Geneclean® II Kit. This was then cloned into the TA cloning vector as described in section 2.9 b. Clones TA2 and TA3 were partially sequenced and found to match the sequence of the PCR product generated in the previous section.

2.9 d Generation of a Probe for Screening Library

Clone TA3 was digested with EcoRI and HindIII (New England Biolabs) to generate a DNA fragment approximately 900 bp long. The fragment was run on an agarose gel, cut and genecleaned as previously described. The fragment was resuspended in sterile milli-Q H_2O at a concentration of 2.1 ng per μl.

2.10 Probing the *T.weissflogii* cDNA library

2.10 a Solutions Used

Church Buffer (500 ml)	Denaturant Buffer	Neutralization Buffer
5 g BSA	1.5 M NaCl	1.5 M NaCl
1 ml 0.5 M EDTA	0.5 N NaOH	0.5 M Tris
9.66 g NaH_2PO_4		pH 8.0
25.55 g Na_2HPO_4		
35 g SDS		

Rinse Buffer	Wash I	Wash II	SM Buffer
0.2 M Tris	2 X SSC	0.2X SSC	100 mM NaCl
2 X SSC	0.5% SDS	0.5% SDS	50 mM tris pH 7.5
pH 7.5			10 mM $MgSO_4$

Southern Base (500 ml)	Southern Neutral (500 ml)	20 X SSC (1L)
NaCl 17.5 g	NaCl 43.85 g	175.3 g NaCl
NaOH 4 g	Tris HCl 66.1 g	88.2 g NaCitrate
	Trizma base 9.7 g	pH 7.0

NZY Media (1L)

5 g NaCl

$MgSO_4 \cdot 7 H_2O$

5 g yeast extract

10 g NZ amine

pH 7.5

(Note: for NZY agar, add 15 g/L bacto-Agar, and for 2 g NZY agarose, add 7 g agarose per liter.)

2.10 b Probe Labeling

The probe was created using Stratagene's Prime-it II® Random Primer Labeling Kit. In each labeling reaction, 25 ng DNA was labeled with $\alpha^{32}P$ dCTP according to kit instructions.

2.10 c Filter Lifts

Eighty 150 mm plates were plated at approximately 40,000 pfu each, for a total of 3.2 million pfu screened. The library was plated on NZY plates according to the instructions in Stratagene's Lambda Zap® II vector kit. Plates were then chilled for 2 hours at 4° C, and overlayed with Hybond N filters (Amersham) for 2 minutes. Filters were immersed for 2 minutes in denaturant buffer, then 5 minutes in neutralization buffer, followed by 30 seconds in rinse buffer. Finally, they were baked at 80° C for 2 hours. The plates were stored at 4° C until they were needed again.

2.10 d Probing the Filters

Probing was carried out in Stratagene's Personalhyb™ hybridization oven in 240 x 80 mm bottles. Ten 150 mm filters were screened per bottle. They were prehybridized in 30 ml Church buffer at 65 ° C for 30 minutes to 2 hours. The probe was denatured by boiling for 5 minutes, then added directly to the prehybridization reaction and hybridized at 65° C overnight. The filters were then washed by 30 ml wash

I at 65° C for 15 minutes, followed by 30 ml wash I at 65° C for 30 minutes, and finally 30 ml wash II at 65° C for 30 minutes. Filters were wrapped in Saran Wrap and exposed to film with an intensifying screen at -70° C overnight.

Cores in the area of potential positives were picked into 1 ml SM buffer + 20 μl chloroform, vortexed, and stored at 4° C. Each contained a mixture of plaques, including the potential positive.

2.10 e PCR Screening to Identify Full Length Clones

Stratagene's Rapid Excision Kit protocol was used to generate plasmid DNA from each of the picks. This DNA was used as template in a PCR reaction using primers to the 5' end of the gene. The primers used were NtForw and BestHind. The sequence of primer NtForw is: 5' gaa aca ctg cat tgg gat gg 3'.

PCR was performed using 1 μl DNA and component concentrations as previously described. The reaction consisted of 23 cycles, denaturing at 95° C for 30 seconds, annealing at 55° C for 30 seconds, and extension at 72° C for 39 seconds. The PCR reactions were run on an agarose gel, and clones with a band at approximately 500 bp were selected for further analysis.

2.10 f Isolation and Retesting of Full Length Clones

Bacteria containing the excised picks from step 2.10 e were plated onto 100 mm LB/ampicillin plates and grown at 37° C overnight. Colony lifts were then performed. Plates were chilled at 4° C for 30 minutes, and then overlayed with Hybond N filters for 2 minutes. Next, the filters were soaked in southern base for 2 minutes, southern neutral for 2 minutes, and finally 2 X SSC for 2 minutes. Filters were then baked and probed exactly as described previously. Positive colonies were picked into 3 ml LB/ ampicillin for minipreps and sequencing.

2.10 g Sequencing

Clones were sequenced using Amersham's Sequenase ® 2.0 DNA Sequencing Kit with α thio ^{35} S dATP at 12.5 μCi per μl. The primer Revcheck was used to confirm that the clones matched carbonic anhydrase sequence. Revcheck sequence: 5' tgt acg ctt cgt cgt cag g 3'. Two clones were selected for further analysis. They were grown up in 100 ml LB/ampicillin and prepped using the Qiagen Plasmid Midi Kit. The DNA was resuspended in 330 μl sterile milli-Q H_2O, extracted with phenol-chloroform and precipitated with ethanol (Sambrook et al. 1989), and finally resuspended in 340 μl sterile milli-Q H_2O. Clone CA44 was sent to the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University for complete sequencing. These sequencing reactions were performed on an Applied Biosystems 373A DNA Sequencer using fluorescently-labeled dideoxynucleotides and Taq DNA polymerase.

3. Results

3.1 Carbonic Anhydrase/ CO_2 Experiment

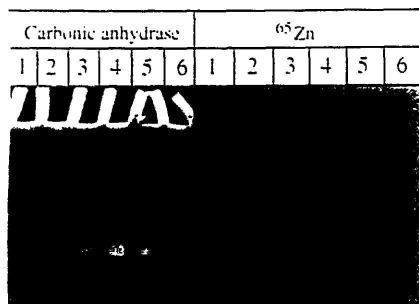


Figure 1. Bromcresol Purple Assay and Autorad. The left panel is a photo of the bromcresol purple assay, the right is an autorad of the same gel. The CO_2 concentrations were as follows: Lane 1: 1000 ppm, Lane 2: 1000 ppm switched to 300 ppm, Lane 3: 300 ppm, Lane 4: 300 ppm switched to 100 ppm, Lane 5: 100 ppm, and Lane 6: 100 ppm switched to 300 ppm.

Figure 1 is a photo of the bromcresol purple assay and corresponding autoradiograph of the gel. Almost all of the cellular zinc was present in the same region of the gel as the carbonic anhydrase activity, indicating that cellular zinc levels may be involved in regulating the amount of carbonic anhydrase and therefore CO_2 uptake capacity. Cells grown at low CO_2 levels had more carbonic anhydrase activity than cells grown at higher CO_2, and adaptation to new CO_2 levels occurred, at least partially, within 18 hours of the switch to a different gas concentration.

3.2 Protein Purification

The gel of the protein purification shows two proteins of very similar size at approximately 27 kilodaltons as well as several fainter bands of smaller sizes (Figure 2). The lower bands may be isoforms or degradation products of the 27 kD proteins. The doublet at 27 kD was confirmed to have carbonic anhydrase activity using the bromcresol purple assay. The smaller bands did not appear to have carbonic anhydrase

activity, although there might not have been enough protein present for activity visible using this assay. Note that the bands present on the gel at approximately 66 kD are artifacts from the sample loading buffer.

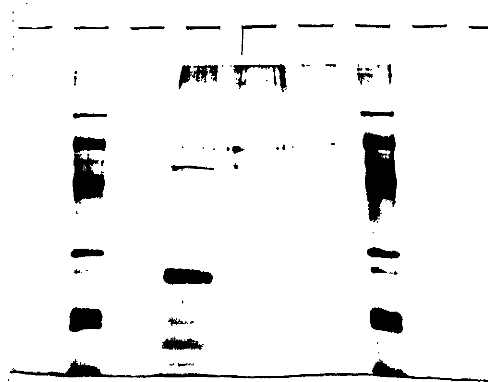


Figure 2. Denaturing Polyacrylamide Gel of Protein Purification. Bio-Rad low range SDS-PAGE standards, found in the far left and far right lanes, are (from the top): 97.4 kD, 66 kD, 45 kD, 31 kD, 21.5 kD, and 14.5 kD. The high molecular weight bands (seen in empty lanes as well as sample lane) are artifacts from the sample loading buffer.

3.3 N-terminal Protein Sequence

N-terminal protein sequencing identified the first 15 amino acids of the protein as: 1. ala, 2. unknown, 3. glu or gly, 4. glu, 5. val, 6. gln, 7. asp, 8. gly, 9. phe, 10. arg, 11. arg or tyr, 12. his, 13. his, 14. tyr, 15. asp.

3.4 Internal Protein Sequences

Two fragments of the protein were sequenced. Eight amino acids of fragment 1 were identified as: 1. ile, 2. ser, 3. ala, 4. ser, 5. ser, 6. phe, 7. asp, 8. lys.

Eighteen amino acids of fragment 2 were identified as: 1.lys, 2. met, 3. gln, 4. arg, 5. asp, 6. asp, 7. met, 8. ser, 9. asp, 10. asp, 11. leu, 12. tyr, 13. ala, 14. pro, 15. gly, 16. unknown, 17. arg, 18. gln.

3.5 Total RNA Extraction

An estimate of cellular RNA concentration as approximately 8 pg per cell was obtained based on very rough calculations. The total yield of mRNA was 1.6 X 10^{-6} g mRNA obtained from 2 X 10^{8} cells, or 8 X 10^{-15} g mRNA per cell. Assuming that mRNA is about 1% of total RNA, and 10% efficiency in the harvest protocol, the cellular RNA concentration is therefore approximately 8 pg per cell.

3.6 The Library

The titer of the *T. weissflogii* Lambda Zap II library on January 9, 1994 was determined to be 10^{10} pfu per ml. The ratio of empty vector to vector with insert was approximately 4 to 1.

3.7 PCR Using Degenerate Primers from Protein Sequence

From the PCR reaction using primers CAN2 and RevDLY, a PCR fragment clone 11.3 was isolated. Clone 11.3 contained sequence coding for both the N-terminal protein fragment as well as both internal protein fragments. The location of this clone on the carbonic anhydrase gene map in figure 3 is from base 543 through base 1072.

3.8 PCR product from Nested PCR Reaction

A larger segment of the carbonic anhydrase gene sequence was obtained by performing PCR with first the T3 and BestHind primers, followed by PCR with the SK and CA4r primers. Two clones, TA2 and TA3 were sequenced and found to partially match the sequence of clone 11.3. The position of the TA2 and TA3 sequence on the

carbonic anhydrase map in figure 3 is from base 01 to base 1076. This sequence also indicated that the carbonic anhydrase gene had more 5' terminal sequence located upstream of the N-terminal protein sequence described in section 3.3. Therefore, this carbonic anhydrase might have a cleaved signal sequence used for cellular localization, or alternatively the protein used for N-terminal sequencing had been partially degraded by proteases prior to the sequencing reaction.

3.9 Library Screening

Of the 3.2 million pfu screened with the probe created from clone TA3, 70 clones were isolated that hybridized to the probe during both the primary and secondary screening. PCR on these 70 clones using the NtForw and BestHind primers identified 11 clones that were long enough that they might contain the 5' end of the probe. Clones CA1, 2, 3, 6, 9, 13, 27, 44, 56, 58, and CA60 were partially sequenced and found to be identical. Clone CA44 was selected for complete sequencing.

3.10 Clone CA44 Vital Statistics

Clone CA44 was 1.4 kb long and encoded a 32 kD protein (Figure 3). The coding region began with a methionine (Met 106) at position 315 and ended with a stop codon at position 1210. The sequence encoding the amino acids predicted to be at the N-terminus was present, but it was 72 amino acids into the coding region. A comprehensive restriction map of the carbonic anhydrase gene is provided in appendices one and two.

This protein had some slightly hydrophobic regions (figure 4), including one of roughly 20 amino acids at the very beginning of the translated protein sequence (in the putative signal sequence). The remainder of the protein was hydrophilic, with the exception of a slightly hydrophobic stretch near the middle.

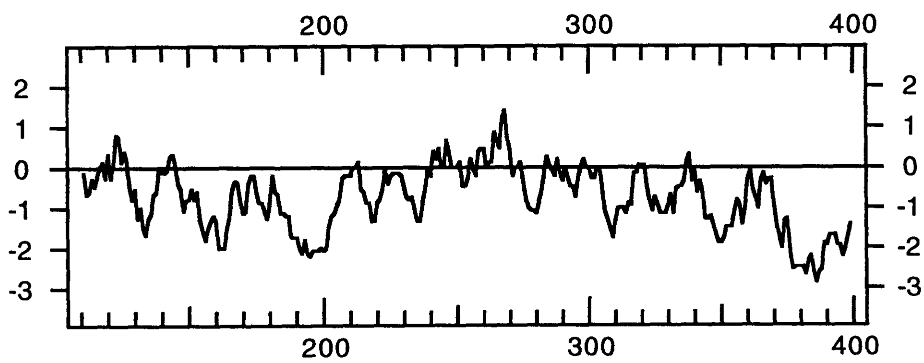
Figure 3. CA44 DNA Sequence and Corresponding Amino Acid Sequence

```
1/1                                     31/11
CAA CAA CAC GCC AAA GAC GAA AAC TCC AAG TTG AAA GCC GCT GTG TGC GTG CTT GGA CTC
gln gln his ala lys asp glu asn ser lys leu lys ala ala val cys val leu gly leu
61/21                                   91/31
TCC ACT GTC GGA CTT CTT GCT TCC ACC ATC GCG CTT GCC GTT CAG AAC AAC AGT TCC AGC
ser thr val gly leu leu ala ser thr ile ala leu ala val gln asn asn ser ser ser
121/41                                  151/51
AAC GCA GCC GCT GCA CCG GCT ACT GCC GAG GAT AAG ACC GTT GCG ACC CTC GAA GCC AGT
asn ala ala ala ala pro ala thr ala glu asp lys thr val ala thr leu glu ala ser
181/61                                  211/71
GGA TCT GAC AAC TTC CTT GTT CCC ATC GAC ATC GTC CCC GAA AGG GCC ACT GCT GAG ATC
gly ser asp asn phe leu val pro ile asp ile val pro glu arg ala thr ala glu ile
241/81                                  271/91
GCC ACC ATC TTC GAC GAG GGC ACC AAC GTA TGT GCC GAG AAG GCG ATC AAG CTG GAC AAC
ala thr ile phe asp glu gly thr asn val cys ala glu lys ala ile lys leu asp asn
301/101                                 331/111
GTT GAC TGC ATC CAC ATG CCT GGT CCT CAA GCT GGC GCT AAC GTC ACC AAG GGA TTC AAG
val asp cys ile his met pro gly pro gln ala gly ala asn val thr lys gly phe lys
361/121                                 391/131
GGA TTG ATG GAA GTC GAC GTA GTC CCC AAC ACC AAG AAT TAC TGG CAA AGC TCC ATG TGC
gly leu met glu val asp val val pro asn thr lys asn tyr trp gln ser ser met cys
421/141                                 451/151
CCC GTC AAT GTT CAC TGG CAT CTT GGA ACC GAA CAC TAC TCT GTC GGC GAG TAT GAC GAA
pro val asn val his trp his leu gly thr glu his tyr ser val gly glu tyr asp glu
481/161                                 511/171
AAT GGC AGT GGT CCG AAC GGA AAC GTT GGC GTT CCT TAC CGC CGT ACC CTT GCC GAG GGA
asn gly ser gly pro asn gly asn val gly val pro tyr arg arg thr leu ala glu gly
541/181                                 571/191
GAA GTG CAG GAT GGA TTT CGC TGC CAT CAC TAC GAC CCT GAC GAC GAA GCG TAC ACC AGG
glu val gln asp gly phe arg cys his his tyr asp pro asp asp glu ala tyr thr arg
601/201                                 631/211
CCC TAT GAA TGG AAA CAC TGC ATT GGG ATG GAA GTT GGA GAG ACA TAT GAA GTT CAT TGG
pro tyr glu trp lys his cys ile gly met glu val gly glu thr tyr glu val his trp
661/221                                 691/231
CCT CAC TCT GGA GCC GGC GCA TGT GGA ACC ACC TAT CAG TAC CAA ACA CCT TTC TAC GAT
pro his ser gly ala gly ala cys gly thr thr tyr gln tyr gln thr pro phe tyr asp
721/241                                 751/251
GGT GTA TTC TGC AAC CTC GAT ATG GAG ACT CTT CAA ACT CTT GCG CCC CAG GAC ATT GCG
gly val phe cys asn leu asp met glu thr leu gln thr leu ala pro gln asp ile ala
781/261                                 811/271
AAC GCA GTT GGA GTT CAA GGA CAA ATC TTC ACC ATT GTC AAT GAC GAC ACA TAC TAC TAC
asn ala val gly val gln gly gln ile phe thr ile val asn asp asp thr tyr tyr tyr
841/281                                 871/291
CCT GAT TTG ATT CGA GGA TGG ATT GTC GAT GAA GAA ATG GGA ATG GGT CAA GAC ATC GCC
pro asp leu ile arg gly trp ile val asp glu glu met gly met gly gln asp ile ala
901/301                                 931/311
ATG TAC ACC GGA TCC ACC ACT GGG GAG AGT CGC AGC AAT GAA ATT TGT TCA TCC TAC TCC
met tyr thr gly ser thr thr gly glu ser arg ser asn glu ile cys ser ser tyr ser
961/321                                 991/331
CCC ATT ACC TGG CAA GTT GAT CGC AAG TGC CAC AAG ATC AGT GCT TCC TCC TTC GAT AAG
pro ile thr trp gln val asp arg lys cys his lys ile ser ala ser ser phe asp lys
1021/341                                1051/351
CTT TGC TAC GAT ATG AAG ATG CAA CGT GAT GAC ATG AGT GAT GAC TTG TAC GCA CAT GGA
leu cys tyr asp met lys met gln arg asp asp met ser asp asp leu tyr ala his gly
1081/361                                1111/371
TCC AGG GAA CTG GTT ACA CCC GAG TAT GTA GCT AAC AAC CAG CAA ACC CGT CGT CTC ACT
ser arg glu leu val thr pro glu tyr val ala asn asn gln gln thr arg arg leu thr
```

```
1141/381                                          1171/391
GAG AAG CAT GAA CAC AAT CAC AGC CAT GGT CAC AGC CAT GTA CGT GGT CAC CAG CAC CAC
glu lys his glu his asn his ser his gly his ser his val arg gly his gln his his
1201/401                                          1231/411
CAA TGG TTT TAG GTT GTC GAT GAG TGT ATG GAT GAT GCT CTT TAG TTT TGT ACG TCT CAC
gln trp phe AMB val val asp glu cys met asp asp ala leu AMB phe cys thr ser his

1261/421                                          1291/431
GAA TAT GTT TAT TAC AGA TTT CCG GAG CCA ATA TTA ATT TCA ATT AGT TAA TTC TAA ACA
glu tyr val tyr tyr arg phe pro glu pro ile leu ile ser ile ser OCH phe OCH thr
1321/441                                          1351/451
AAT TAT TAG TTT CCT TAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA
asn tyr AMB phe pro OCH lys lys lys lys lys lys lys lys lys lys lys lys lys lys
1381/461
AAA AAA AAA AAA AAA AAA AAA AAT
lys lys lys lys lys lys lys asn
```

Figure 4. Kyte Doolittle Hydrophobicity Plot of the CA44 Protein Sequence. Translated protein sequence begining at Met 108. Amino acid position is on the horizontal axis and hydrophobicity on the vertical axis.



4. Conclusions

The *T. weissflogii* carbonic anhydrase gene described in this thesis does not appear to have homology to any published algal chloroplastic or periplasmic carbonic anhydrase isoforms. The *Chlamydomonas reinhardtii* periplasmic genes exhibit some homology to mammalian carbonic anhydrases (Fukuzawa et al. 1991), and many of the algal chloroplastic genes are homologous to the chloroplastic carbonic anhydrases of higher plants such as pea (Fukuzawa et al. 1992). Since the *T. weissflogii* protein is in the same size range as the other carbonic anhydrases and has the carbonic anhydrase enzyme activity, the lack of homology might simply reflect the fact that *T. weissflogii* is evolutionarily quite distant from these other organisms. Alternatively, the sequence differences might suggest slightly different properties of the *T. weissflogii* enzyme. For example, it might be associated with a transporter, a proton pump, or some other protein utilized in the carbon accumulation mechanism.

The protein predicted to be translated from clone CA44 was 32 kD, approximately 8 kD longer than expected. The amino terminal amino acids sequenced from the purified carbonic anhydrase are 72 residues downstream of the putative start methionine predicted by the CA44 cDNA. It has not been determined whether the proteolytic events that generated the sequenced polypeptide were a result of proteolysis during protein purification or a physiological protein processing event. If, however, the N-terminus of the sequenced protein was not degraded, then the roughly 8 kD protein encoded upstream of it is probably a signal sequence responsible for sending the protein to the correct cellular location. Many of the amino acids at the beginning of this stretch of 72 residues are slightly hydrophobic, so it is possible that this portion of the protein might serve as a signal sequence and be cleaved after insertion into the endoplasmic reticulum.

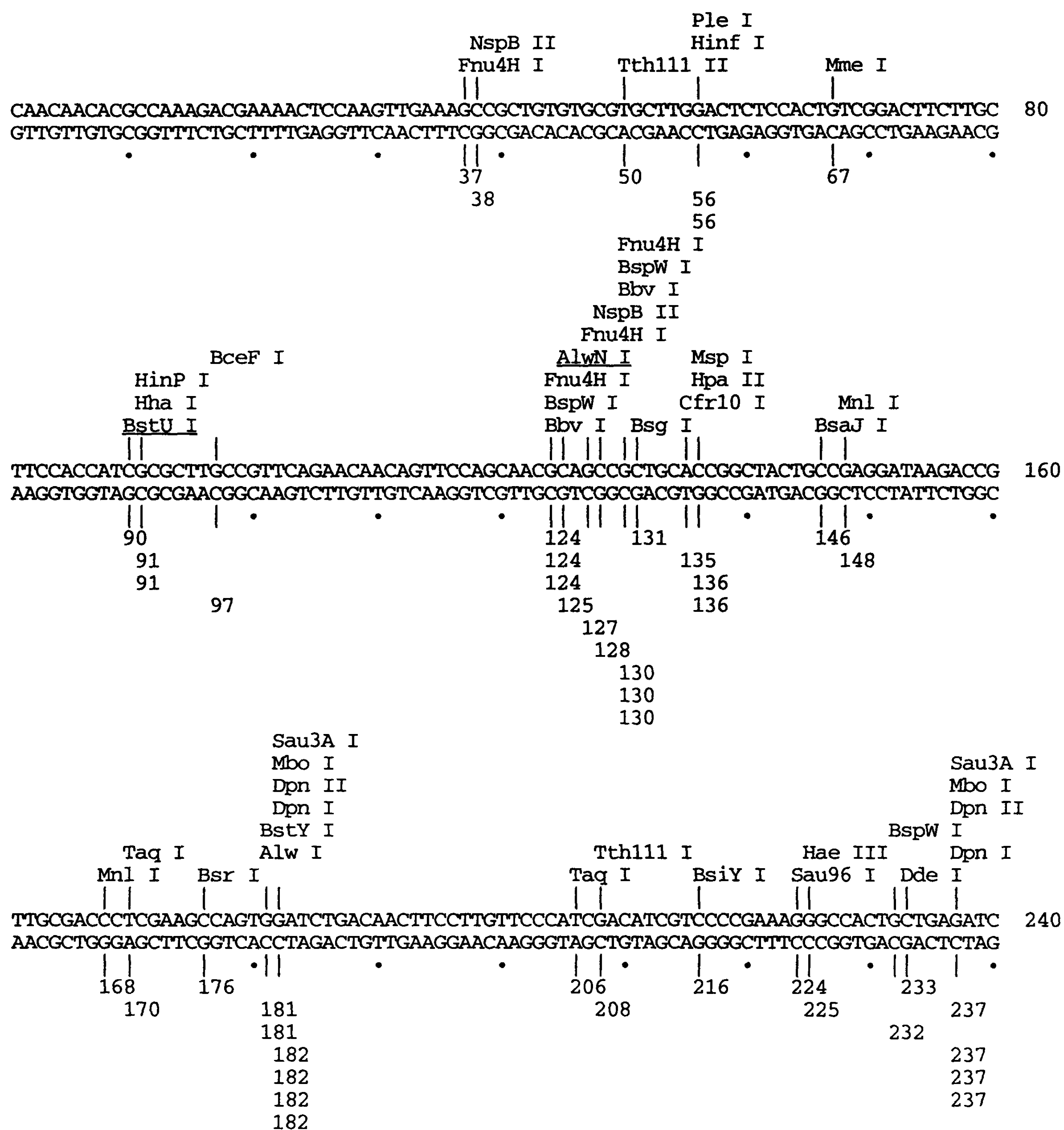
The apparent size of the purified carbonic anhydrase on the denaturing polyacrylamide gel was 27 kD. This is different from the size of the predicted protein beginning at Met 106 (~32 kD) as well as the predicted protein beginning at Ala 178 (~24 kD). There are at least two possible explanations for these discrepancies. The simplest explanation is that salts from the purification protocol might have caused the sample to run differently from the standards in the gel. Additionally, one or more post-translational modifications such as glycosylation and signal sequence cleavage might

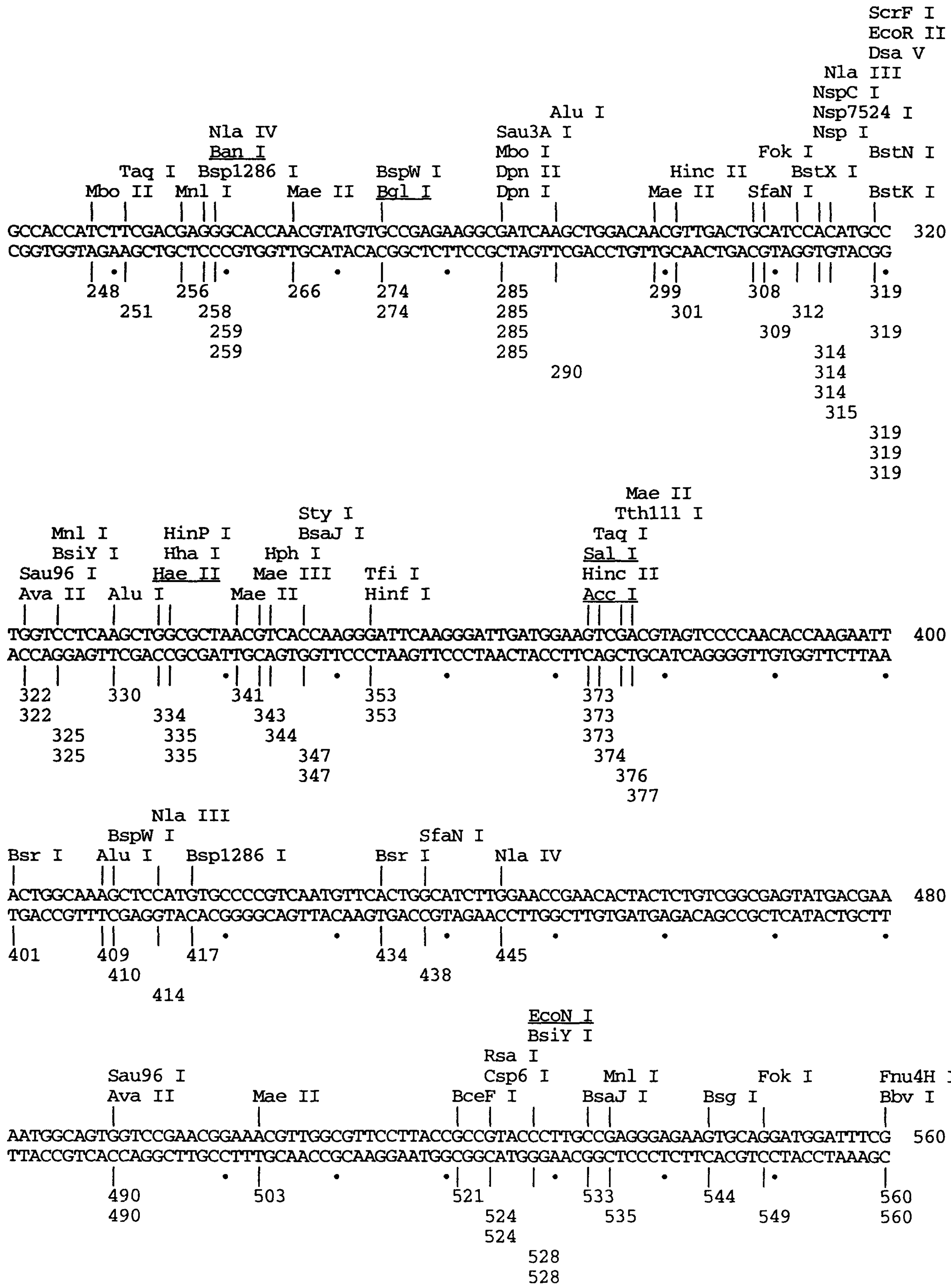
cause the final protein size to be larger or smaller than that predicted by the gene sequence. There are three potential N-linked glycosylation sites present in the protein which would act to make the protein larger than predicted. The consensus sequence for this glycosylation event, Asn-X-Ser/Thr, is present at Asn 114, Asn 161, and Asn 386. Therefore, if the preprotein that is initially translated is glycosylated at one or more sites, it would be larger than 32 kD, and if a signal sequence were then cleaved, the protein would be smaller again. This could result in a final protein of 32 kD.

This carbonic anhydrase gene is a valuable tool that can be used to characterize the inorganic carbon acquisition mechanism utilized by *T. weissflogii*. The cloned gene can now be manipulated to create recombinant protein which can be synthesized and purified in large quantities. This recombinant protein can be used to perform *in vitro* experiments designed to ascertain the structure and function of the enzyme. For example, the enzyme kinetics can be examined at different CO_2 concentrations. Also, the recombinant protein can be used to raise antibodies. These antibodies can then be used to visualize carbonic anhydrase cellular localization, as well as in field studies to examine the occurrence of this enzyme in the environment. Also, the gene sequence can be used to generate probes to examine regulation of the *T. weissflogii* carbonic anhydrase in response to different environmental stimuli. Specifically, the molecular mechanisms by which zinc and CO_2 modulate carbonic anhydrase enzyme levels can be studied. The probes can be used to examine mRNA levels due to changes in transcription induced by different zinc and CO_2 concentrations. Finally, the purification technique can be modified to try to detect additional carbonic anhydrase isoforms. All of the information obtained by these experiments should lead to a better understanding of exactly how *T. weissflogii* acquires inorganic carbon.

Appendix One

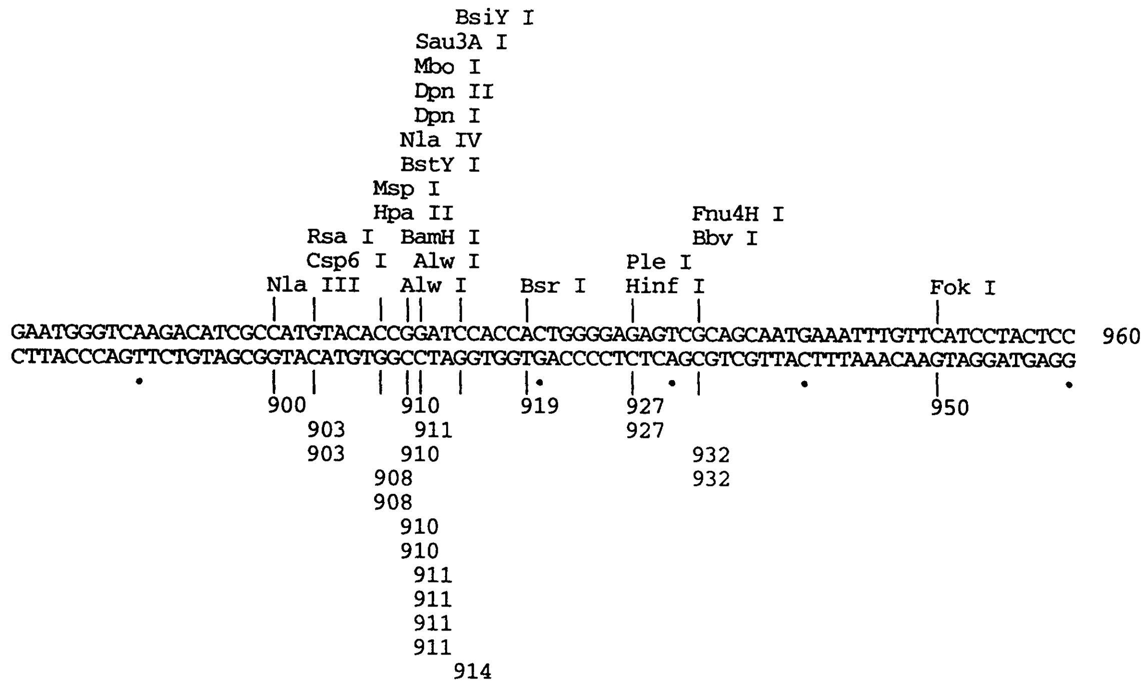
Restriction Map of Carbonic Anhydrase clone CA44



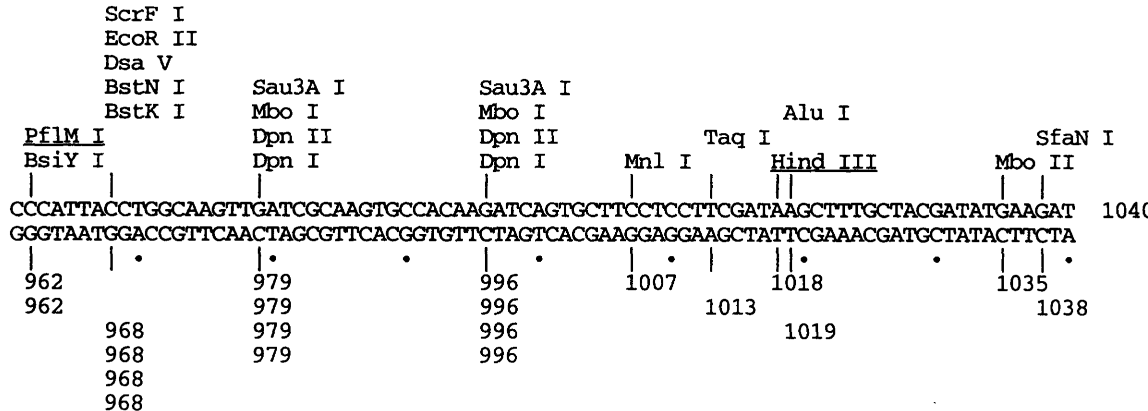


ScrF I
EcoR II
Dsa V
Nla III
NspC I
Alu I
Nsp7524 I
Nla IV
Sau3A I
Nsp I
Ban I
Mbo I
Fok I
BstN I
Taq I
Bsp1286 I
BspW I
Dpn II
Hinc II
BstX I
Mbo II
Mnl I
Mae II
Bgl I
Dpn I
Mae II
SfaN I
BstK I
GCCACCATCTTCGACGAGGGCACCAACGTATGTGCCGAGAAGGCGATCAAGCTGGACAACGTTGACTGCATCCACATGCC 320
CGGTGGTAGAAGCTGCTCCCGTGGTTGCATACACGGCTCTTCCGCTAGTTCGACCTGTTGCAACTGACGTAGGTGTACGG
248 256 266 274 285 299 308 319
251 258 274 285 301 312
259 285 309 319
259 285 314
290 314
314
315
319
319
319
Mae II
Sty I
Tth111 I
Mnl I
HinP I
BsaJ I
Taq I
BsiY I
Hha I
Hph I
Sal I
Sau96 I
Hae II
Mae III
Tfi I
Hinc II
Ava II
Alu I
Mae II
Hinf I
Acc I
TGGTCCTCAAGCTGGCGCTAACGTCACCAAGGGATTCAAGGGATTGATGGAAGTCGACGTAGTCCCCAACACCAAGAATT 400
ACCAGGAGTTCGACCGCGATTGCAGTGGTTCCCTAAGTTCCCTAACTACCTTCAGCTGCATCAGGGGTTGTGGTTCTTAA
322 330 341 353 373
322 334 343 353 373
325 335 344 373
325 335 347 374
347 376
377
Nla III
BspW I
SfaN I
Bsr I
Alu I
Bsp1286 I
Bsr I
Nla IV
ACTGGCAAAGCTCCATGTGCCCCGTCAATGTTCACTGGCATCTTGGAACCGAACACTACTCTGTCGGCGAGTATGACGAA 480
TGACCGTTTCGAGGTACACGGGGCAGTTACAAGTGACCGTAGAACCTTGGCTTGTGATGAGACAGCCGCTCATACTGCTT
401 409 417 434 445
410 438
414
EcoN I
BsiY I
Rsa I
Sau96 I
Csp6 I
Mnl I
Fok I
Fnu4H I
Ava II
Mae II
BceF I
BsaJ I
Bsg I
Bbv I
AATGGCAGTGGTCCGAACGGAAACGTTGGCGTTCCTTACCGCCGTACCCTTGCCGAGGGAGAAGTGCAGGATGGATTTCG 560
TTACCGTCACCAGGCTTGCCTTTGCAACCGCAAGGAATGGCGGCATGGGAACGGCTCCCTCTTCACGTCCTACCTAAAGC
490 503 521 533 544 560
490 524 535 549 560
524
528
528

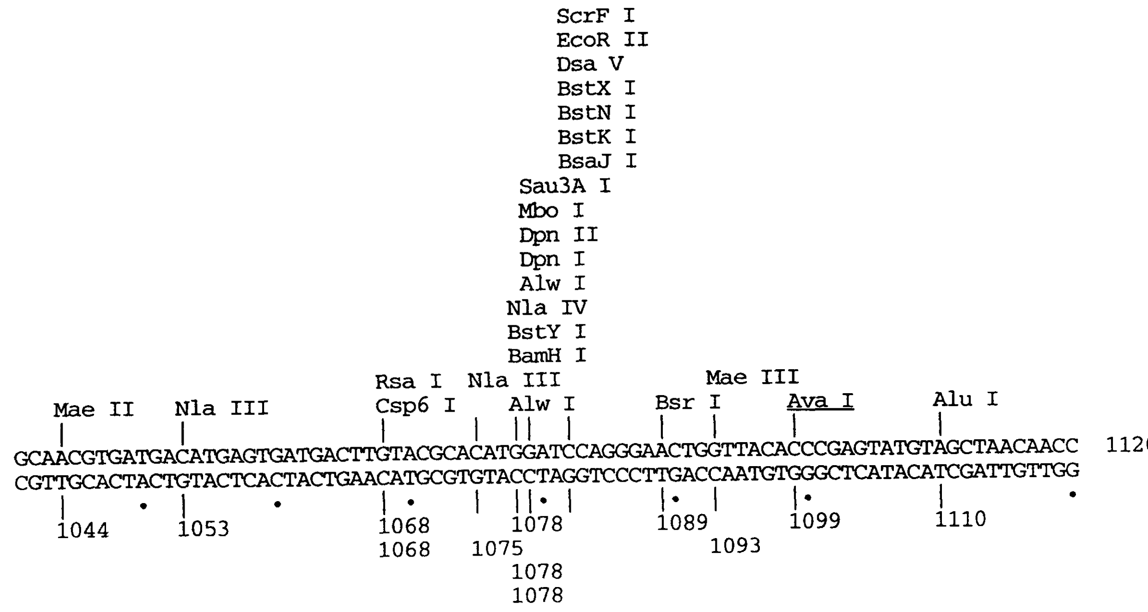
```
                                         BsiY I
                                      Sau96 I
                                   ScrF I
                                   EcoR II
                                   Dsa V
                                   BstN I
                                   BstK I
                              Rsa I  EcoO109 I                                  BsmA I
                              Csp6 I  Hae III                     Fok I   Mme I
CTGCCATCACTACGACCCTGACGACGAAGCGTACACCAGGCCCTATGAATGGAAACACTGCATTGGGATGGAAGTTGGAG  640
GACGGTAGTGATGCTGGGACTGCTGCTTCGCATGTGGTCCGGGATACTTACCTTTGTGACGTAACCCTACCTTCAACCTC
                              591     599                         626     634
                              591    598                                        640
                                   596
                                   596
                                   596
                                   596
                                   596
                                      599
                                         602
                                      NspC I
                                      Nsp7524 I
                                    HinP I
                                 Msp I
                                 Hpa II
                    Mnl I           Nae I  Nla III
                    BsiY I          Cfr10 I                       Tth111 II
                  Hae III    Nla IV   Nsp I                   Rsa I
   Nde I         Hae I      Gsu I   Hha I   Nla IV            Csp6 I
AGACATATGAAGTTCATTGGCCTCACTCTGGAGCCGGCGCATGTGGAACCACCTATCAGTACCAAACACCTTTCTACGAT  720
TCTGTATACTTCAAGTAACCGGAGTGAGACCTCGGCCGCGTACACCTTGGTGGATAGTCATGGTTTGTGGAAAGATGCTA
   644           658        668     677     685               699
                  659        670      679                     699
                    661         673                               703
                    661         673    680
                                 674
                                 674
                                    677
                                      679
                                      679
                                               ScrF I
                                               EcoR II
                                               Dsa V
                             Mbo II            BstN I
                            Ear I              BstK I
                Taq I     Ple I               BsaJ I
              Mnl I       Hinf I          HinP I
              BsiY I    BsmA I            Hha I                 Mme I
GGTGTATTCTGCAACCTCGATATGGAGACTCTTCAAACTCTTGCGCCCCAGGACATTGCGAACGCAGTTGGAGTTCAAGG  800
CCACATAAGACGTTGGAGCTATACCTCTGAGAAGTTTGAGAACGCGGGGTCCTGTAACGCTTGCGTCAACCTCAAGTTCC
              735       745               763                   787
              735         747             763
                737       747                 767
                            749                768
                             750               768
                                               768
                                               768
                                               768
                                                     Mnl I
                                                   Taq I
        Hph I                                   Tfi I  Fok I          Mbo II
     Mbo II                                     Hinf I           Taq I
ACAAATCTTCACCATTGTCAATGACGACACATACTACTACCCTGATTTGATTCGAGGATGGATTGTCGATGAAGAAATGG  880
TGTTTAGAAGTGGTAACAGTTACTGCTGTGTATGATGATGGGACTAAACTAAGCTCCTACCTAACAGCTACTTCTTTACC
     806                                        849              866
        809                                     849    856            871
                                                   852
                                                     854
```



BsiY I
Sau3A I
Mbo I
Dpn II
Dpn I
Nla IV
BstY I
Msp I
Hpa II
Fnu4H I
Rsa I
BamH I
Bbv I
Csp6 I
Alw I
Ple I
Nla III
Alw I
Bsr I
Hinf I
Fok I
GAATGGGTCAAGACATCGCCATGTACACCGGATCCACCACTGGGGAGAGTCGCAGCAATGAAATTTGTTCATCCTACTCC 960
CTTACCCAGTTCTGTAGCGGTACATGTGGCCTAGGTGGTGACCCCTCTCAGCGTCGTTACTTTAAACAAGTAGGATGAGG
900 910 919 927 950
903 911 927
903 910 932
908 932
908
910
910
911
911
911
911
914



ScrF I
EcoR II
Dsa V
BstN I
Sau3A I
Sau3A I
BstK I
Mbo I
Mbo I
Alu I
PflM I
Dpn II
Dpn II
Taq I
SfaN I
BsiY I
Dpn I
Dpn I
Mnl I
Hind III
Mbo II
CCCATTACCTGGCAAGTTGATCGCAAGTGCCACAAGATCAGTGCTTCCTCCTTCGATAAGCTTTGCTACGATATGAAGAT 1040
GGGTAATGGACCGTTCAACTAGCGTTCACGGTGTTCTAGTCACGAAGGAGGAAGCTATTCGAAACGATGCTATACTTCTA
962 979 996 1007 1018 1035
962 979 996 1013 1038
968 979 996 1019
968 979 996
968
968



ScrF I
EcoR II
Dsa V
BstX I
BstN I
BstK I
BsaJ I
Sau3A I
Mbo I
Dpn II
Dpn I
Alw I
Nla IV
BstY I
BamH I
Mae III
Rsa I
Nla III
Mae II
Nla III
Csp6 I
Alw I
Bsr I
Ava I
Alu I
GCAACGTGATGACATGAGTGATGACTTGTACGCACATGGATCCAGGGAACTGGTTACACCCGAGTATGTAGCTAACAACC 1120
CGTTGCACTACTGTACTCACTACTGAACATGCGTGTACCTAGGTCCCTTGACCAATGTGGGCTCATACATCGATTGTTGG
1044 1053 1068 1078 1089 1099 1110
1068 1075 1093
1078
1078

```
                                                        Mae III       Mae II
                                                    Nla III          BsaA I
                                                   Sty I            Rsa I
                                                   Nco I            Csp6 I  Hph I
                    Dde I                          Dsa I          Nla III  Mae III
              BsmA I      Nla III                  BsaJ I        BstX I   BstE II
              |     |           |                  ||   |         || |||   |||
AGCAAACCCGTCGTCTCACTGAGAAGCATGAACACAATCACAGCCATGGTCACAGCCATGTACGTGGTCACCAGCACCAC  1200
TCGTTTGGGCAGCAGAGTGACTCTTCGTACTTGTGTTAGTGTCGGTACCAGTGTCGGTACATGCACCAGTGGTCGTGGTG
         .    |     |.       |  .         .         ||   |.         || |||   ||| .         .
              1133          1147                   1164           1176     1186
                    1139                           1164            1177     1187
                                                   1164               1180    1188
                                                   1164               1180
                                                    1165               1181
                                                        1169            1182
                                                                 BsmA I
                                                               Mae II
                                             SfaN I          Rsa I
                          Taq I         Fok I               Csp6 I
                          |             |   |               | | |
CAATGGTTTTAGGTTGTCGATGAGTGTATGGATGATGCTCTTTAGTTTTGTACGTCTCACGAATATGTTTATTACAGATT  1280
GTTACCAAAATCCAACAGCTACTCACATACCTACTACGAGAAATCAAAACATGCAGAGTGCTTATACAAATAATGTCTAA
         .                |  .          |   |    .          | | |       .         .         .
                          1217          1230                1250
                                            1234            1250
                                                              1252
                                                                1254
   Nla IV
 Msp I        Mse I
 Hpa II      Ase I
BspE I    Ssp I               Mse I                      Mse I
|| |      |  ||               |                          |
TCCGGAGCCAATATTAATTTCAATTAGTTAATTCTAAACAAATTATTAGTTTCCTTAAAAAAAAAAAAAAAAAAAAAAAA  1360
AGGCCTCGGTTATAATTAAAGTTAATCAATTAAGATTTGTTTAATAATCAAAGGAATTTTTTTTTTTTTTTTTTTTTTTT
|| |      |  ||     .         | .         .         .    |    .         .         .
1281      1290                1308                       1335
 1282        1293
 1282         1294
   1284


AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT  1404
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA
         .         .         .         .
```

Appendix Two

Summary of Restriction Sites in Clone CA44

Acc I	gt/mkac		1	1(372) 2	373(840) 1			
AlwN I	cagnnn/ctg		1	1(124) 2	125(1088) 1			
Ava I	c/ycgrg		1	1(1098) 1	1099(114) 2			
Ban I	g/gyrcc		1	1(258) 2	259(954) 1			
Bgl I	gccnnnn/nggc		1	1(273) 2	274(939) 1			
BsaA I	yac/gtr		1	1(1180) 1	1181(32) 2			
BstE II	g/gtnacc		1	1(1185) 1	1186(27) 2			
BstU I	cg/cg		1	1(89) 2	90(1123) 1			
Dsa I	c/crygg		1	1(1163) 1	1164(49) 2			
Ear I	ctcttc	1/4	1	1(748) 1	749(464) 2			
EcoN I	cctnn/nnnagg		1	1(527) 2	528(685) 1			
EcoO109 I	rg/gnccy		1	1(597) 2	598(615) 1			
Gsu I	ctggag	16/14	1	1(667) 1	668(545) 2			
Hae I	wgg/ccw		1	1(657) 1	658(555) 2			
Hae II	rgcgc/y		1	1(333) 2	334(879) 1			
Hind III	a/agctt		1	1(1017) 1	1018(195) 2			
Nae I	gcc/ggc		1	1(672) 1	673(540) 2			
Nco I	c/catgg		1	1(1163) 1	1164(49) 2			
Nde I	ca/tatg		1	1(643) 1	644(569) 2			
PflM I	ccannnn/ntgg		1	1(961) 1	962(251) 2			
Sal I	g/tcgac		1	1(372) 2	373(840) 1			
Ava II	g/gwcc		2	1(321) 2	322(168) 3	490(723) 1		
BamH I	g/gatcc		2	1(909) 1	910(168) 2	1078(135) 3		
BceF I	acggc	12/13	2	1(96) 3	97(424) 2	521(692) 1		
Bsg I	gtgcag	16/14	2	1(130) 3	131(413) 2	544(669) 1		
Bsp1286 I	gdgch/c		2	1(257) 2	258(159) 3	417(796) 1		
Cfr10 I	r/ccggy		2	1(134) 3	135(538) 2	673(540) 1		
Dde I	c/tnag		2	1(232) 2	233(906) 1	1139(74) 3		
Hinc II	gty/rac		2	1(300) 2	301(72) 3	373(840) 1		
Nsp I	rcatg/y		2	1(313) 3	314(365) 2	679(534) 1		
Nsp7524 I	r/catgy		2	1(313) 3	314(365) 2	679(534) 1		
NspB II	cmg/ckg		2	1(37) 3	38(90) 2	128(1085) 1		
NspC I	rcatg/y		2	1(313) 3	314(365) 2	679(534) 1		
Sty I	c/cwwgg		2	1(346) 2	347(817) 1	1164(49) 3		
Tfi I	g/awtc		2	1(352) 3	353(496) 1	849(364) 2		
Tth111 I	gacn/nngtc		2	1(207) 2	208(168) 3	376(837) 1		
Tth111 II	caarca	11/9	2	1(49) 3	50(653) 1	703(510) 2		
BsmA I	gtctc	1/5	3	1(639) 1	640(105) 3	745(388) 2	1133(80) 4	
BstX I	ccannnnn/ntgg		3	1(311) 2	312(770) 1	1082(94) 3	1176(37) 4	
BstY I	r/gatcy		3	1(180) 2	181(729) 1	910(168) 3	1078(135) 4	
Hae III	gg/cc		3	1(224) 3	225(374) 2	599(60) 4	659(554) 1	
Hpa II	c/cgg		3	1(135) 4	136(538) 1	674(234) 3	908(305) 2	
Hph I	ggtga	8/7	3	1(343) 3	344(465) 1	809(379) 2	1188(25) 4	
Mme I	tccrac	20/18	3	1(66) 4	67(567) 1	634(153) 3	787(426) 2	
Msp I	c/cgg		3	1(135) 4	136(538) 1	674(234) 3	908(305) 2	
Ple I	gagtc	4/5	3	1(55) 4	56(691) 1	747(180) 3	927(286) 2	
SfaN I	gcatc	5/9	3	1(307) 2	308(130) 4	438(600) 1	1038(175) 3	
Bbv I	gcagc	8/12	4	1(123) 4	124(6) 5	130(430) 1	560(372) 2	932(281) 3
Hha I	gcg/c		4	1(90) 4	91(244) 3	335(342) 2	677(86) 5	763(450) 1
HinP I	g/cgc		4	1(90) 4	91(244) 3	335(342) 2	677(86) 5	763(450) 1
Mae III	/gtnac		4	1(342) 2	343(750) 1	1093(76) 3	1169(18) 5	1187(26) 4
Sau96 I	g/gncc		4	1(223) 2	224(98) 5	322(168) 3	490(109) 4	599(614) 1

Alu I	ag/ct		5	1(289) 2	290(40) 6	330(79) 5	409(610) 1
				1019(91) 4	1110(103) 3		
Alw I	ggatc	4/5	5	1(180) 2	181(729) 1	910(1) 5	911(167) 3
				1078(1) 6	1079(134) 4		
BspW I	gcnnnnn/nngc		5	1(123) 3	124(6) 6	130(102) 4	232(42) 5
				274(136) 2	410(803) 1		
Bsr I	actgg	1/-1	5	1(175) 3	176(225) 2	401(33) 6	434(485) 1
				919(170) 4	1089(124) 5		
BstK I	c/cngg		5	1(318) 1	319(277) 2	596(172) 4	768(200) 3
				968(114) 6	1082(131) 5		
BstN I	cc/wgg		5	1(318) 1	319(277) 2	596(172) 4	768(200) 3
				968(114) 6	1082(131) 5		
Dsa V	/ccngg		5	1(318) 1	319(277) 2	596(172) 4	768(200) 3
				968(114) 6	1082(131) 5		
EcoR II	/ccwgg		5	1(318) 1	319(277) 2	596(172) 4	768(200) 3
				968(114) 6	1082(131) 5		
Fok I	ggatg	9/13	5	1(308) 1	309(240) 3	549(77) 6	626(230) 4
				856(94) 5	950(263) 2		
Hinf I	g/antc		5	1(55) 6	56(297) 2	353(394) 1	747(102) 4
				849(78) 5	927(286) 3		
Mbo II	gaaga	8/7	5	1(247) 2	248(502) 1	750(56) 6	806(65) 5
				871(164) 4	1035(178) 3		
ScrF I	cc/ngg		5	1(318) 1	319(277) 2	596(172) 4	768(200) 3
				968(114) 6	1082(131) 5		
BsaJ I	c/cnngg		6	1(145) 5	146(201) 3	347(186) 4	533(234) 2
				767(315) 1	1082(82) 6	1164(49) 7	
Csp6 I	g/tac		6	1(523) 1	524(67) 6	591(108) 5	699(204) 2
				903(165) 3	1068(112) 4	1180(33) 7	
Fnu4H I	gc/ngc		6	1(36) 5	37(87) 4	124(3) 6	127(3) 7
				130(430) 1	560(372) 2	932(281) 3	
Nla IV	ggn/ncc		6	1(258) 1	259(186) 4	445(225) 2	670(15) 7
				685(225) 3	910(168) 5	1078(135) 6	
Rsa I	gt/ac		6	1(523) 1	524(67) 6	591(108) 5	699(204) 2
				903(165) 3	1068(112) 4	1180(33) 7	
Dpn I	ga/tc		7	1(181) 2	182(55) 6	237(48) 7	285(626) 1
				911(68) 5	979(17) 8	996(83) 4	1079(134) 3
Dpn II	/gatc		7	1(181) 2	182(55) 6	237(48) 7	285(626) 1
				911(68) 5	979(17) 8	996(83) 4	1079(134) 3
Mae II	a/cgt		7	1(265) 2	266(33) 7	299(42) 5	341(36) 6
				377(126) 4	503(541) 1	1044(138) 3	1182(31) 8
Mbo I	/gatc		7	1(181) 2	182(55) 6	237(48) 7	285(626) 1
				911(68) 5	979(17) 8	996(83) 4	1079(134) 3
Sau3A I	/gatc		7	1(181) 2	182(55) 6	237(48) 7	285(626) 1
				911(68) 5	979(17) 8	996(83) 4	1079(134) 3
BsiY I	ccnnnnn/nngg		8	1(215) 2	216(109) 5	325(203) 3	528(74) 6
				602(59) 8	661(74) 7	735(179) 4	914(48) 9
				962(251) 1			
Taq I	t/cga		8	1(169) 3	170(36) 8	206(45) 7	251(123) 5
				374(363) 1	737(115) 6	852(14) 9	866(147) 4
				1013(200) 2			
Mnl I	cctc	7/7	9	1(147) 4	148(20)10	168(88) 7	256(69) 9
				325(210) 1	535(126) 5	661(74) 8	735(119) 6
				854(153) 3	1007(206) 2		
Nla III	catg/		9	1(314) 1	315(99) 5	414(266) 2	680(220) 3
				900(153) 4	1053(22) 8	1075(72) 6	1147(18) 9
				1165(12)10	1177(36) 7		

262 sites found

Appendix Three

Sequences of Primers and Their Locations on the Map in the Gene[a]

Primer Name	Sequence[b]	Position of 5' End
CANterm	5' ggI ttY MgI MgI caY caY taY gaY 3'	553
CAN2	5' gaR gtI caR gaY ggI ttY 3'	541
RevMQR	5' cat Rtc Rtc IcK Ytg cat 3'	1056
ForMQR	5' atg caR MgI gaY gaY atg 3'	1039
RevDLY[1]	5' ccN ggI gcR taI aRR tc 3'	1078
BestHind	5' gtg cgt aca agt cat cac tca tgt c 3'	1051
CA4r	5' ctt gcg atc aac ttg cca gg 3'	987
CA3f	5' tta cct ggc aag ttg atc gc 3'	965
NTForw	5' gaa aca ctg cat tgg gat gg 3'	512
Revcheck	5' tgt acg ctt cgt cgt cag g 3'	595
NewHindIII[2]	5' cct tcg ata agc ttt gct acg ata tg 3'	1010
NewNterm[3]	5' act cgg tct ccg aag tgc agg atg ggt ttc 3'	541

[b] Degenerate base symbols are according to the IUPAC nomenclature system.

[a] Base locations correspond to the map in figure three and the map in appendix one.

[1] This primer may have 2° structure problems.

[2] A mistake was made designing this primer. The fourth base should have been t.

[3] This adds a restriction site (BsaI) onto the 5' end.

Appendix Four

Kits Used in Materials and Methods

Kit Name	Company	Location
Geneclean® II	Bio 101, Inc.	Vista, CA
TA Cloning® Kit	Invitrogen Corp.	San Diego,CA
RNAgents® Total RNA Isolation Kit	Promega Corp.	Madison, WI
PolyATract® mRNA Isolation System	Promega Corp.	Madison, WI
Riboclone® cDNA Synthesis System	Promega Corp.	Madison, WI
Riboclone® EcoRI Linker Ligation System	Promega Corp.	Madison, WI
Lambda Zap®II Vector Kit	Stratagene	La Jolla, CA
Rapid Excision Kit	Stratagene	La Jolla, CA
Qiagen Plasmid Kits	Qiagen, Inc.	Chatsworth,CA
Prime-it® II Random Primer Labeling Kit	Stratagene	La Jolla, CA
Sequenase® 2.0 DNA Sequencing Kit	Amersham	Arlington Hghts, IL

References

Aizawa, K. and Miyachi, S. (1984) Carbonic Anhydrase Located on Cell Surface Increases the Affinity for Inorganic Carbon in Photosynthesis of *Dunaliella tertiolecta*. *FEBS Lett.*, 173: 41-44.

Berger, W.H. and Keir, R.S. (1984) Glacial-Holocene Changes in Atmospheric CO_2 and the Deep-Sea Record. in Climate Processes and Climate Sensitivity, J. Hanson and T. Takahashi, eds. AGU Monograph Ser. 29.

Bruland, K.W. (1989) Complexation of Zinc by Natural Organic Ligands in the Central North Pacific. *Limnol. Oceanogr.*, 34: 269-285.

Bruland, K.W. (1992) Complexation of Cadmium by Natural Organic Ligands in the Central North Pacific. *Limnol. Oceanogr.*, 37(5): 1008-1017.

Coleman, J.R. and Grossman, A.R. (1984) Biosynthesis of Carbonic Anhydrase in *Chlamydomonas reinhardtii* During Adaptation to Low CO_2. *Proc. Natl. Acad. Sci.*, 81: 6049-6053.

Husic, H.D., Kitayama, M., Togasaki, R.K., Moroney, J.V., Morris, K.L., and Tolbert, N.E. (1989) Identification of Intracellular Carbonic Anhydrase in *Chlamydomonas reinhardtii* Which is Distinct from the Periplasmic Form of the Enzyme. *Plant Physiol.*, 89: 904-909.

Ishida, S., Muto, S., and Miyachi, S. (1993) Structural Analysis of Periplasmic Carbonic Anhydrase 1 of *Chlamydomonas reinhardtii*. *Eur. J. Biochem.*, 214: 9-16.

Fukuzawa, H., Ishida, S., and Miyachi, S. (1991) cDNA Cloning and Gene Expression of Carbonic Anhydrase in *Chlamydomonas reinhardtii*. *Can. J. Bot.*, 69: 1088-1096.

Fukuzawa, H., Suzuki, E., Komukai, Y. and Miyachi, S. (1992) A Gene Homologous to Chloroplast Carbonic Anhydrase (icfA) is Essential to Photosynthetic Carbon Dioxide Fixation by *Synechococcus* PCC7942. *Proc. Natl. Acad. Sci.*, 89: 4437-4441.

Kamo, T., Shimogawara, K., Fukuzawa, H., Muto, S. and Miyachi, S. (1990) Subunit Composition of Carbonic Anhydrase from *Chlamydomonas reinhardtii*. *Eur. J. Biochem.*, 192: 557-562.

Martin, J.H., Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S.J., Hunter, C.N., Elrod, V. A., Nowicki, J.L., Coley, T.L., Barber, R.T., Lindley, S., Watson, A.J., Van Scoy, K., Law, C.S., Liddicoat, M.I., Ling, R., Stanton, T., Stockel, J., Collins C., Anderson, A., Bidigare, R., Ondrusek, M., Latasa, M., Millero, F.J., Lee, K., Yao, W., Zhang, J.Z., Friederich, G., Sakamoto, C., Chavez, F., Buck, K., Kolber, Z., Greene, R., Falkowski, P., Chisholm, S.W., Hoge, F., Swift, R., Yungel, J., Turner, S., Nightingale, P., Hatton, A., Liss, P. and Tindale, N.W. (1994) Testing the Iron Hypothesis in Ecosystems of the Equatorial Pacific Ocean. *Nature*, 371: 123-143.

Martin, J. H. and Fitzwater, S. E. (1988) Iron Deficiency Limits Phytoplankton Growth in the North-East Pacific Subarctic. *Nature*, 331: 341-343.

Morel, F.M.M., Hudson, R.J.M., and Price, N.M. (1991) Limitation of Productivity by Trace Metals in the Sea. *Limnol. Oceanogr.*, 36(8): 1742-1755.

Morel, F.M.M., Reinfelder, J.R., Roberts, S.B., Chamberlain, C.P., Lee, J.G., and Yee, D. (1994) Zinc and Carbon Co-limitation of Marine Phytoplankton. *Nature*, 369: 740-742.

Murray, J.W., Barber, R.T., Roman, M.R., Bacon, M.P., and Feely, R.A. (1994) Physical and Biological Controls on Carbon Cycling in the Equatorial Pacific. *Science*, 266: 58-64.

Nimer, N.A. and Merrett, M.J. (1992) Calcification and Utilization of Inorganic Carbon by the Coccolithophorid *Emiliana huxleyi* Lohmann. *New Phytol.*, 121 (2): 173-178.

Patel, B.N. and Merrett, M.J. (1986) Inorganic Carbon Uptake by the Marine Diatom *Phaeodactylum tricornutum. Planta (Berl)*, 169: 222-227.

Patterson, B.D., Atkins, C.A., Graham, D. and Wills, R.B.H. (1971) Carbonic Anhydrase: A New Method of Detection on Polyacrylamide Gels Using Low-Temperature Fluorescence. *Anal. Biochem.*, 44: 388-391.

Price, N.M., Harrison, G.I., Hering, J., Hudson, R.J., Nirel, P.M.V., Palenik, B., and Morel, F.M.M. (1989) Preparation and Chemistry of the Artificial Algal Culture Medium Aquil. *Biolog. Oceanog.*, 6: 443-461.

Rawat, M. and Moroney, J.V. (1991) Partial Characterization of a New Isoenzyme of Carbonic Anhydrase Isolated from *Chlamydomonas reinhardtii*. *J. Biol. Chem.*, 266(15): 9719-9723.

Redfield, A.C., Ketchum, G.H. and Richards, F.A. (1963) The Influence of Organisms on the Composition of Sewwater. In *The Sea*, ed. M.N. Hill, vol 2, 26-77. Wiley Press, New York.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, New York.

Sambrotto, R.N., Savidge, G., Robinson, C, Boyd, P., Takahashi, T., Karl, D.M., Langdon, C., Marra, J., and Codispoti, L. (1993) Elevated Consumption of Carbon Relative to Nitrogen in the Surface Ocean. *Nature*, 363: 248-250.

Sarmiento, J.L. (1993) Ocean Carbon Cycle. *Chem. and Eng. News*, May 31, 1993, 30-43.

Siegenthaler, U. and Sarmiento, J.L. (1993) Atmospheric Carbon Dioxide and the Ocean. *Nature*, 365: 119-125.

Sikes, C.S. and Wheeler, A.P. (1980) Carbonic Anhydrase and Carbon Fixation in Coccolithophorids. *J. Phycol.*, 18: 423-426.

Sundquist, E.T. (1993) The Global Carbon Dioxide Budget. *Science*, 259: 934-941.

Takahashi, T., Broeker, W.S., and Langer, S. (1985) Redfield Ratio Based on Data from Isopycnal Surfaces. *J. Geophys. Res.*, 90: 6907-6924.

Watson, A.J., Law, C.S., Van Scoy, K.A., Millero, F.J., Yao, W., Friederich, G.E., Liddicoat, M.I., Wanninkhof, R.H., Barber, R.T., and Coale, K.H. (1994) Minimal Effects of Iron Fertilization on Sea-surface Carbon Dioxide Concentrations. *Nature*, 371: 143-145.

Williams, T. G. and Coleman, B.C. (1993) Identification of Distinct Internal and External Isozymes of Carbonic Anhydrase in *Chlorella saccharophila*. *Plant Physiol.*, 103: 943-946.

Yang, S.-Y., Tsuzuki, M., and Miyachi, S. 1985. Carbonic Anhydrase of *Chlamydomonas*: Purification and Studies on its Induction Using Antiserum against *Chlamydomonas* Carbonic Anhydrase. *Plant Cell Physiol.*, 26(1): 25-34.